



Cite this: *Nat. Prod. Rep.*, 2024, 41, 1020

## Plant peptides – redefining an area of ribosomally synthesized and post-translationally modified peptides

Jonathan R. Chekan, <sup>\*,a</sup> Lisa S. Mydy, <sup>b</sup> Michael A. Pasquale <sup>a</sup> and Roland D. Kersten <sup>\*,b</sup>

Covering 1965 to February 2024

Plants are prolific peptide chemists and are known to make thousands of different peptidic molecules. These peptides vary dramatically in their size, chemistry, and bioactivity. Despite their differences, all plant peptides to date are biosynthesized as ribosomally synthesized and post-translationally modified peptides (RiPPs). Decades of research in plant RiPP biosynthesis have extended the definition and scope of RiPPs from microbial sources, establishing paradigms and discovering new families of biosynthetic enzymes. The discovery and elucidation of plant peptide pathways is challenging due to repurposing and evolution of housekeeping genes as both precursor peptides and biosynthetic enzymes and due to the low rates of gene clustering in plants. In this review, we highlight the chemistry, biosynthesis, and function of the known RiPP classes from plants and recommend a nomenclature for the recent addition of BURP-domain-derived RiPPs termed burpitides. Burpitides are an emerging family of cyclic plant RiPPs characterized by macrocyclic crosslinks between tyrosine or tryptophan side chains and other amino acid side chains or their peptide backbone that are formed by copper-dependent BURP-domain-containing proteins termed burpitide cyclases. Finally, we review the discovery of plant RiPPs through bioactivity-guided, structure-guided, and gene-guided approaches.

Received 19th September 2023

DOI: 10.1039/d3np00042g

rsc.li/npr

1.	<b>Introduction</b>	3.3.3	<b>Bioactivity</b>
2.	<b>Plant peptide chemistry and biosynthesis</b>	3.4.	<b>Cysteine-rich peptides</b>
2.1.	<b>Precursor peptides</b>	3.4.1	<b>Structure</b>
2.2.	<b>Post-translational modification</b>	3.4.2	<b>Biosynthesis</b>
2.3.	<b>Proteolysis</b>	3.4.3	<b>Bioactivity</b>
3.	<b>Plant RiPP classes</b>	3.5.	<b>Linear plant RiPPs</b>
3.1.	<b>Cyclotides</b>	3.5.1	<b>Structure</b>
3.1.1	<b>Structure</b>	3.5.2	<b>Biosynthesis</b>
3.1.2	<b>Biosynthesis</b>	3.5.3	<b>Bioactivity</b>
3.1.3	<b>Bioactivity</b>	3.6.	<b>BURP-domain-derived RiPPs (burpitides)</b>
3.2.	<b>Orbitides</b>	3.6.1	<b>General recommendations for burpitide nomenclature</b>
3.2.1	<b>Structure</b>	3.6.1.1	<b>Macrocyclic bond</b>
3.2.2	<b>Biosynthesis</b>	3.6.1.2	<b>Ring size</b>
3.2.2.1	<b>Caryophyllaceae-type</b>	3.6.1.3	<b>Ring number and mixed modifications</b>
3.2.2.2	<b>PawL-type</b>	3.6.1.4	<b>Terminal modification</b>
3.2.2.3	<b>Others</b>	3.6.1.5	<b>Biosynthesis</b>
3.2.3	<b>Bioactivity</b>	3.6.2	<b>Lyciumin-type peptides (Trp-indole-N-to-C)</b>
3.3.	<b>PawS-derived peptides</b>	3.6.2.1	<b>Macrocyclic bond</b>
3.3.1	<b>Structure</b>	3.6.2.2	<b>Structure</b>
3.3.2	<b>Biosynthesis</b>	3.6.2.3	<b>Biosynthesis</b>
		3.6.2.4	<b>Bioactivity</b>
		3.6.3	<b>Cyclopeptide alkaloids (Tyr-phenol-O-to-C)</b>
		3.6.3.1	<b>Macrocyclic bond</b>

<sup>a</sup>Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, NC, USA. E-mail: jrchekan@uncg.edu

<sup>b</sup>Department of Medicinal Chemistry, University of Michigan, Ann Arbor, MI, USA. E-mail: rkersten@umich.edu



- 3.6.3.2 Ring size
- 3.6.3.3 Ring number
- 3.6.3.4 Structure
- 3.6.3.5 Biosynthesis
- 3.6.3.6 Bioactivity
- 3.6.4 Stephanotic acid-type peptides (Trp-C-to-C)
  - 3.6.4.1 Macrocyclic bond
  - 3.6.4.2 Structure
  - 3.6.4.3 Biosynthesis
  - 3.6.4.4 Bioactivity
- 3.6.5 Hibispeptin-type peptides (Tyr-C-to-C)
  - 3.6.5.1 Macrocyclic bond
  - 3.6.5.2 Structure
  - 3.6.5.3 Biosynthesis
  - 3.6.5.4 Bioactivity
- 4. Undefined plant-derived peptides
  - 4.1. Bouvardins

- 4.1.1 Structure
- 4.1.2 Biosynthesis
- 4.1.3 Bioactivity
- 5. Chemotaxonomy of plant peptides
- 6. Discovery of plant RiPPs
  - 6.1. Bioactivity-guided discovery
  - 6.2. Structure-guided discovery
  - 6.3. Gene-guided discovery
- 7. Future directions
- 8. Conflicts of interest
- 9. Acknowledgments
- 10. References

## 1. Introduction

RiPPs are a large, diverse group of natural products which are produced by bacteria, archaea, fungi, animals, and plants. The



Jonathan R. Chekan

*Jonathan R. Chekan is an Assistant Professor of Chemistry and Biochemistry at the University of North Carolina at Greensboro. He received his PhD in 2016 from the University of Illinois Urbana-Champaign under the supervision of Satish K. Nair. As the Simons Foundation Post-Doctoral Fellow of the Life Science Research Foundation in the laboratory of Bradley S. Moore, he investigated the biosynthesis of neurotoxic*

*marine natural products. In 2020, he joined University of North Carolina at Greensboro where his research group focuses on the bioinformatic guided discovery of new natural product biosynthetic pathways from across the tree of life.*



Michael A. Pasquale

*Michael A. Pasquale is pursuing his PhD in chemistry and biochemistry at the University of North Carolina at Greensboro. He is in the lab of Jonathan R. Chekan and his research focus is the biosynthesis of RiPP natural products in plants. Prior to his graduate studies, in 2021 Michael received his B.S. in Biochemistry from the University of North Carolina at Greensboro, where he studied natural product chemistry with Dr Nadja*

*Cech.*



Lisa S. Mydy

*Lisa Mydy is fascinated by protein structure and function in natural product biosynthesis. She earned her PhD in Chemistry with Dr Nicholas Silvaggi from the University of Wisconsin-Milwaukee in 2016. She was a postdoc with Dr Andrew Gulick at the University at Buffalo, spent time in industry with Abbott, and is currently a postdoc in Dr Roland Kersten's laboratory at the University of Michigan. She*

*determined the first crystal structure of a BURP domain protein with the co-mentorship of Dr Janet Smith and plans to continue exploring the BURP superfamily.*



Roland D. Kersten

*Roland Kersten is an Assistant Professor of Medicinal Chemistry at the University of Michigan, Ann Arbor. He pursued graduate studies in Dr Pieter Dorrestein's and Dr Brad Moore's labs at University of California, San Diego on microbial genome mining for natural product discovery. He then was trained as a postdoc in Dr Joseph Noel's and Dr Jing-Ke Weng's labs at Salk Institute and Whitehead Institute in plant and algal natural product biosynthesis. Roland's lab develops gene-guided discovery approaches for plant natural products.*



defining RiPP biosynthetic feature is the ribosomal generation of a precursor peptide which is post-translationally modified (PTM) and proteolytically cleaved to yield a mature RiPP natural product<sup>1,2</sup> (Fig. 1A). Plants are a prolific source of peptide chemistry<sup>3</sup> and the investigation of underlying peptide biosynthetic genes has revealed that plants produce peptide natural products exclusively *via* the ribosomal pathway based on current knowledge.<sup>4–21</sup> Plant RiPPs have yielded sustainable pest control agents in agriculture such as Sero-X,<sup>22,23</sup> and they have emerged as promising lead compounds for medicinal applications exemplified by the clinical lead immunosuppressant T20K.<sup>24</sup> With growing plant genetic resources<sup>25</sup> and recent advances in identification of precursor peptides and biosynthetic enzymes involved in plant RiPP biosynthesis, the field of plant RiPPs is primed for an expansion of plant RiPP chemistry and enzymology that will further advance research about their agricultural and therapeutic applications. This review aims to give an overview of plant RiPP chemistry, biosynthesis, and bioactivity while also recommending nomenclature for the recent discovery of side-chain-macrocylic RiPPs produced by BURP-domain peptide cyclases.

In this review, we will first introduce plant peptide chemistry to highlight characterized ribosomal peptides with post-translational modifications. We will then give insights into plant RiPP biosynthetic trends based on current biochemical knowledge (Section 2). Next, plant RiPP classes will be reviewed in structure, biosynthesis, and bioactivity (Section 3). We summarize the distribution of these plant RiPPs across vascular plants in Section 5. Finally, discovery approaches for plant RiPPs are described (Section 6) and we look ahead in future research directions of plant RiPPs (Section 7). We also review biosynthetically undefined peptides derived from plants (Section 4). This review has a particular focus on BURP-domain-derived RiPPs due to their recent biosynthetic discovery. For the other plant peptide classes, we further recommend excellent reviews discussed in their respective sections.

## 2. Plant peptide chemistry and biosynthesis

To date, over 1500 modified peptides have been isolated from plants.<sup>3,18,26,27</sup> Detailed biosynthetic studies have firmly established the biosynthetic basis for many of these peptides, allowing for classification based on both structure and construction (Fig. 1B). The plant itself is typically responsible for biosynthesis of the RiPP, which is contrary with other sessile organisms like sponges that rely on a complex microbiome to produce the observed bioactive metabolite.<sup>28–30</sup> There are notable exceptions to this trend in plants, such as the macrocyclic depsipeptide FR900359,<sup>31</sup> which is produced by the endosymbiotic bacterium *Candidatus Burkholderia crenata* in the leaves of the plant *Ardisia crenata*. FR900359 is a potent inhibitor of the Gq subfamily of G protein signaling, a novel therapeutic target for treating asthma, inflammation, and cancer.<sup>32–35</sup> Other peptidic natural products derived from plant

sources such as the bouvardins still await biosynthetic classification (Fig. 1B and Section 4).

Plant RiPPs have a wide range in size, from four amino acids (cyclopeptide alkaloids)<sup>36</sup> up to 66 amino acids (cysteine-rich peptides)<sup>37</sup> or 430 Da to 7 kDa (Fig. 1B). A common structural feature of plant peptide natural products is macrocyclization. Macrocylic plant RiPPs can be constructed in a variety of ways: head-to-tail (cyclotide), side-chain-to-side-chain (cyclopeptide alkaloids) or side-chain-to-backbone (lyciumins). Below we describe important features of plant RiPP pathways following general RiPP biosynthetic order of precursor peptide formation, post-translational modification, and proteolysis.

### 2.1. Precursor peptides

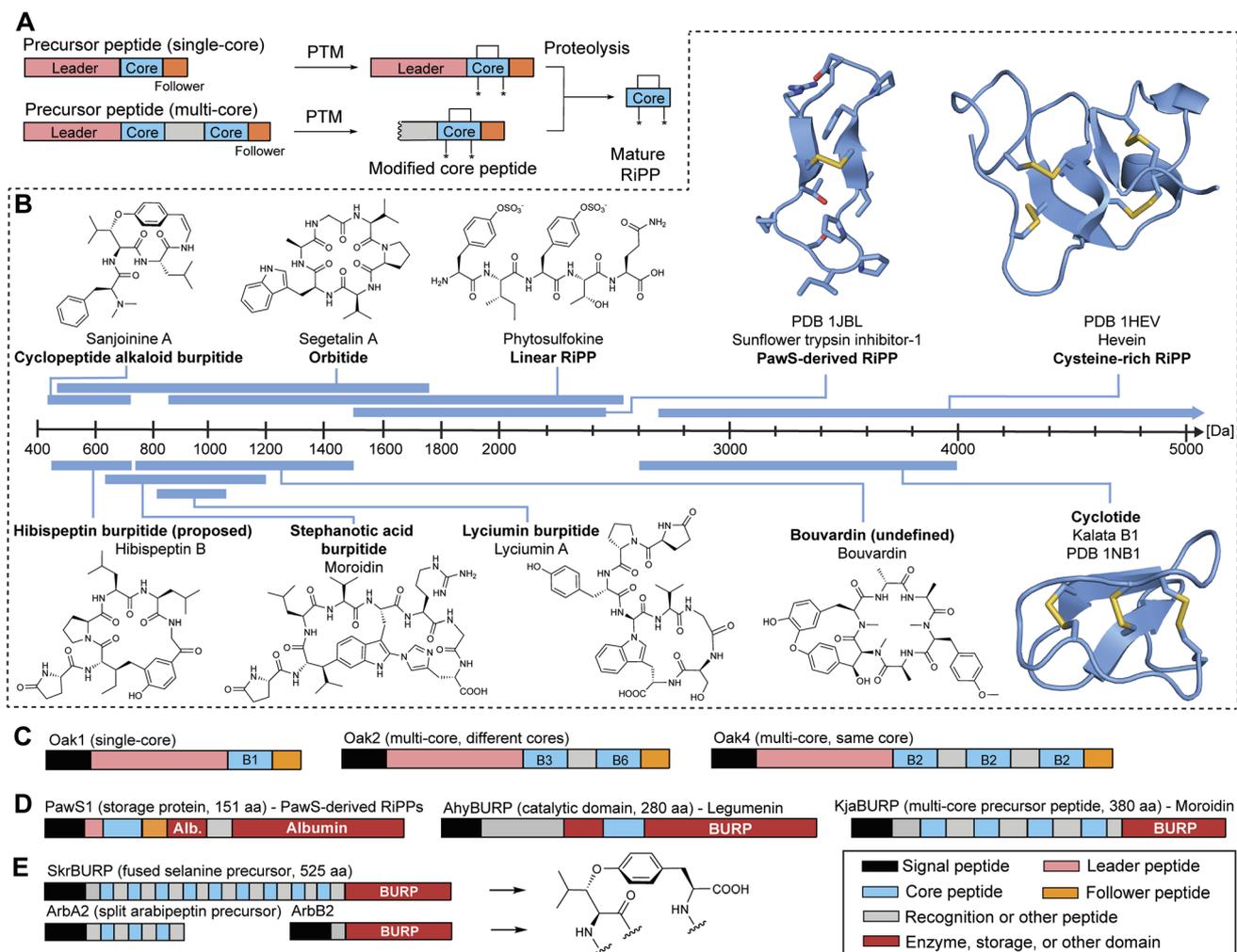
As characteristic of RiPPs, genetically encoded precursor peptides serve as the substrate for the biosynthetic pathways. Precursor peptide sequences can often be subdivided by their functional role. Since many contrasting definitions have been used for plant precursor peptide organization over the past decades, we propose the following in addition to accepted RiPP precursor peptide definitions:<sup>1,2</sup>

- Leader (peptide) is the non-repeating N-terminal portion of the precursor peptide following the signal sequence (if a signal peptide is present) before the first N-terminal core peptide.
- Core (peptide) is a sequence of the precursor peptide which is transformed to the RiPP natural product.
- Recognition sequence is the repeating unit demarcating core sequences in multi-core precursor peptides.
- Follower (peptide) is the non-repeating C-terminal portion of the peptide following the core sequence.

Leader, follower and recognition sequences often function as regions for interactions with PTM enzymes or domains and proteases in RiPP pathways.<sup>1,2</sup> The presence of leader peptide, follower peptides, or recognition sequences is not a requirement for plant RiPP precursor peptides.

Plant RiPPs can be derived from single-core and multi-core precursor peptides (Fig. 1C). Most plant RiPP classes have precursor peptides that have been characterized to include more than one core peptide in sequence repeats, known as multi-core precursor peptides (Table 1).<sup>38</sup> For example, two of the originally reported cyclotide precursor peptides (Oak2, Oak4) from *Oldenlandia affinis* include 2 or 3 core peptides, respectively, but precursor peptides with only one cyclotide core peptide (Oak1, Oak3) were reported from the same plant.<sup>4</sup> Other multi-core precursor peptides have been described for orbittides,<sup>39</sup> linear RiPPs,<sup>11</sup> several cysteine-rich peptides,<sup>40–42</sup> lyciumins,<sup>17</sup> moroidins<sup>19</sup> and cyclopeptide alkaloids<sup>18,20</sup> (Table 1). Multi-core precursor peptides are also in biosynthetic pathways of fungal RiPPs such as dikaritin<sup>43–46</sup> and borosins,<sup>47,48</sup> and cyanobacterial RiPPs such as cyanobactins<sup>49,50</sup> and microviridins.<sup>51</sup> A multi-core precursor in plant RiPPs could have the advantage over a single-core precursor in that RiPP production could be increased faster and more efficient, *i.e.* transcription and translation of one precursor gene can yield multiple plant RiPP products. The evolution of multi-core precursors has been hypothesized to occur through internal gene duplication and





**Fig. 1** Plant RiPP chemistry and biosynthesis. (A) RiPP biosynthetic dogma.<sup>1,2</sup> (B) Representative structures of plant RiPPs and their corresponding size ranges. Peptide mass ranges were determined based on plant peptide databases and studies.<sup>18,26,27</sup> (C) Single-core and multi-core precursor peptides of kalata cyclotides. B1 represents kalata B1 core peptide. (D) Precursor proteins with storage, catalytic, and multi-core domains. (E) Fused and split precursor peptides in cyclopeptide alkaloid RiPP biosynthesis. Disulfide bond formation and pyroglutamate formation can occur enzymatically or spontaneously. Abbreviations: PTM – post-translational modification, alb. – albumin, core – core peptide, Leader – leader peptide, Follower – follower peptide.

neofunctionalization in the form of core peptide diversification, as observed for squash trypsin inhibitor cyclotides.<sup>5</sup> Many precursor peptides from plant RiPP pathways are above 100 amino acids in length (Table 1 and Fig. 1D) and therefore larger than typical precursor peptides from bacterial RiPPs.<sup>1,2</sup> The large size of plant precursor peptides arises for several reasons: (a) plant RiPP precursor peptides often include signal peptides for processing through the secretory pathway in plant cells, (b) they can include multiple core peptides,<sup>4,5,17</sup> (c) core peptides have evolved into storage proteins<sup>22,52,53</sup> and (d) precursor peptides can be fused to catalytic domains.<sup>18,54</sup> In some cases, the large size of plant RiPP precursors has been hypothesized to be a result of cyclic peptide motifs ‘hijacking’ plant proteins and their processing enzymes for biosynthesis.<sup>52</sup> Such cyclic peptide motifs can function as stand-alone inhibitor protein loops like in Bowman–Birk-inhibitors<sup>55</sup> or serve as RiPP core peptides for

post-translational modification and proteolysis during cyclic peptide biosynthesis.<sup>6</sup>

## 2.2. Post-translational modification

Bacterial RiPPs are often heavily modified to contain many non-proteogenic amino acids, whereas plant RiPPs are generally comprised of proteinogenic amino acids with only a limited number of post-translationally modified amino acids. Plant RiPP PTMs can be differentiated into terminal modifications such as N-to-C-macrocyclization or pyroglutamate formation and into side-chain-modifications such as crosslinking of tryptophan and tyrosine side chains with other amino acids (Fig. 2). Multiple plant RiPP classes feature PTMs common to eukaryotic proteins, including cystines, pyroglutamate, tyrosine sulfation, and proline hydroxylation.<sup>56</sup> We rationalize the inclusion of common protein PTMs in plant RiPPs because of the general definition of RiPPs as



**Table 1** Plant RiPP biosynthetic genes. Representative precursor peptides and PTM enzymes of plant RiPP classes are described in source organism, size, Pfam accession, GenBank accession and reference. Abbreviations: TIPTOP – two inhibitor peptide topologies precursor, aa – amino acid

Name	Organism	Length [aa]	RiPP class	Core peptide #	Pfam/InterPro accession <sup>64</sup>	GenBank accession	Reference
<b>Precursor peptides</b>							
Oak1	<i>Oldenlandia affinis</i>	124	Cyclotide	1	n/a	AAI05477	4
Oak2	<i>Oldenlandia affinis</i>	158	Cyclotide	2	n/a	AAI05478	4
Oak3	<i>Oldenlandia affinis</i>	111	Cyclotide	1	n/a	AAI05479	4
Oak4	<i>Oldenlandia affinis</i>	210	Cyclotide	3	n/a	AAI05480	4
TIPTOP1	<i>Momordica cochinchinensis</i>	281	Cyclotide, knottin	5	n/a	AEK70372	5
TIPTOP2	<i>Momordica cochinchinensis</i>	331	Cyclotide, knottin	5	n/a	AEK70373	5
TIPTOP3	<i>Momordica cochinchinensis</i>	431	Cyclotide, knottin	8	n/a	AEK70374	5
PawS1	<i>Helianthus annuus</i>	151	PawS-derived peptide	1	PF00234	ACT34883	52
PawS2	<i>Helianthus annuus</i>	137	PawS-derived peptide	1	PF00234	ACS74804	52
CterM precursor	<i>Clitoria ternatea</i>	127	Cyclotide	1	PF16720	AEB92229	22
PawL1a	<i>Senecio pinnatifolius</i> var. <i>maritimus</i>	175	Orbitide	2	PF00234	ARD06052	6
PawL1b	<i>Senecio pinnatifolius</i> var. <i>maritimus</i>	160	Orbitide	1	PF00234	ARD06067	6
Presegetalin A1	<i>Gypsophila vaccaria</i>	32	Orbitide	1	n/a	AEG75782	7
Linusorb A1–A3 precursor protein	<i>Linum usitatissimum</i>	219	Orbitide	5	n/a	FAA04139	8
Phytosulfokine precursor protein (PSK)	<i>Asparagus officinalis</i>	75	Phyto-sulfokines (linear RiPP)	1	PF06404	Q9FS10	9
CLAVATA3	<i>Arabidopsis thaliana</i>	96	CLE family peptides (linear RiPP)	1	n/a	NP_001118398	10
Systemin-like precursor proTOBSYS-A	<i>Nicotiana tabacum</i>	165	Systemins (linear RiPP)	2	n/a	AAK52096	11
Metallo-carboxypeptidase inhibitor IIA precursor	<i>Solanum tuberosum</i>	102	GRP (knottin)	1	PF02977	NP_001275048	12
AcAMP-2 precursor	<i>Amaranthus caudatus</i>	86	GRP (6C hevein)	1	PF00187	P27275	13
Ginsentide precursor	<i>Panax ginseng</i>	121	GRP (8C hevein)	1	PF18687	AAX40471.1	65
Hevein precursor	<i>Hevea brasiliensis</i>	204	GRP (8C hevein)	1	PF00187, PF00967	AAA33357	14
EeCBP precursor	<i>Euonymus europaeus</i>	305	GRP (10C hevein)	1	PF00187, PF00182	AAP35270	16
WAMP-1a precursor	<i>Triticum kiharae</i>	116	GRP (10C hevein)	1	PF00187	P85966	66
EAPP1 precursor	<i>Eucommia ulmoides</i>	307	GRP (10C hevein)	1	PF00187, PF00182	WED30098	54
Defensin-like protein 1	<i>Triticum aestivum</i>	77	GRP (8C defensin)	1	PF00304	XP_044389906	67
Petunia defensin precursor	<i>Petunia × hybrida</i>	103	GRP (10C defensin)	1	PF00304	ADV59771	68
Crambin precursor	<i>Crambe hispanica</i> subsp. <i>abyssinica</i>	125	GRP (6C thionin)	1	PF00321	S52550	69
$\alpha$ 1 purothionin precursor	<i>Triticum aestivum</i>	136	GRP (8C thionin)	1	PF00321	BAA12336	70
Jasminide S1 precursor	<i>Jasminum sambac</i>	100	GRP (jasminotide)	1	n/a	ALO52196	71
Potentilla pA1 precursor	<i>Potentilla anserina</i>	285	GRP (potentillide)	4	n/a	XP_050382506	40
$\alpha$ -hairpinin EcAMP2 precursor	<i>Echinochloa crus-galli</i>	361	GRP ( $\alpha$ -hairpinin)	3	PF14861	B3EWR6	41
IbAMP precursor	<i>Impatiens balsamina</i>	333	GRP (AMP)	6	n/a	O24006	42
Nodule-specific cysteine-rich peptide 169	<i>Medicago truncatula</i>	61	GRP (NCR)	1	PF07127	ABS31414	72
Snakin-1 precursor	<i>Solanum tuberosum</i>	88	GRP (snakin)	1	PF02704	CAC44032	37
RALF precursor	<i>Nicotiana tabacum</i>	115	GRP (RALF)	1	PF05498	AAI26478	73
LbaLycA	<i>Lycium barbarum</i>	543	Lyciumin-type RiPP (lyciumin)	12	PF03181	AYN06992	17
KjaBURP	<i>Kerria japonica</i>	380	Stephanotic acid-type RiPP (moroidin)	4	PF03181	QIG55799	19

Table 1 (Contd.)

Name	Organism	Length [aa]	RiPP class	Core peptide #	Pfam/InterPro accession <sup>64</sup>	GenBank accession	Reference
SkrBURP	<i>Selaginella kraussiana</i>	525	Cyclopeptide alkaloid	8	PF03181	QXY82431	18
CcaBURP1	<i>Cercis canadensis</i>	630	Lyciumin-type RiPP (cercic acid)	16	PF03181	QXY82432	18
CcaBURP2	<i>Cercis canadensis</i>	458	Stephanotic acid-type RiPP (stephanotic acid)	8	PF03181	QXY82433	18
ElaBURP	<i>Elaeagnus pungens</i>	505	Cyclopeptide alkaloid	6	PF03181	OR257605	21
ArbA2	<i>Coffea arabica</i>	162	Cyclopeptide alkaloid	3	PF10950	XP_027066141	20
<b>PTM &amp; processing enzymes</b>							
Butelase (asparaginyl endopeptidase/peptide macrocyclase)	<i>Clitoria ternatea</i>	482	Cyclotide	n/a	PF01650	AIB06797	74
OaAFP1 (asparaginyl endopeptidase/peptide macrocyclase)	<i>Oldenlandia affinis</i>	474	Cyclotide	n/a	PF01650	ALG36103	75
OaPDI	<i>Oldenlandia affinis</i>	531	Cyclotide	n/a	PF00085	ABS11216	59
OaRD21a	<i>Oldenlandia affinis</i>	465	Cyclotide	n/a	PF00112	QBH22534	76
PCY1	<i>Gypsophila vaccaria</i>	724	Orbitide	n/a	n/a	AGL51088	77
Tyrosine sulfotransferase	<i>Arabidopsis thaliana</i>	500	Linear RiPP	n/a	PF03567	Q3EDG5	78
Hydroxyproline O-arabinosyltransferase	<i>Arabidopsis thaliana</i>	366	Linear RiPP	n/a	n/a	Q8W4E6	79
ArbB2	<i>Coffea arabica</i>	320	Cyclopeptide alkaloid	n/a	PF03181	XP_027066250	20

ribosomal peptides with a post-translational modification and an altered structure–activity-relationship (e.g. a change in biological target affinity, stability, conformation, charge state) of the unmodified ribosomal peptide through the PTM. For example, cystines were not detailed in seminal RiPP reviews as PTMs defining a ribosomal peptide as post-translationally modified, but one of these reviews mentions ribosomal peptides with disulfide bonds, so called cysteine-rich peptides (CRPs), following RiPP biosynthetic logic.<sup>1,2</sup> We suggest including cystines as RiPP-defining PTMs given their nature as a covalent precursor peptide modification and their impact on peptide SAR (Section 3.4). Plant RiPPs with cystines are cyclotides, PawS-derived peptides, and CRPs. In contrast to established RiPP-defining PTMs which are only formed enzymatically, cystines can form non-enzymatically through oxidative folding<sup>57,58</sup> or enzymatically by protein disulfide isomerases.<sup>59</sup>

Plant RiPP biosynthesis can occur through a split pathway or an autocatalytic (fused) pathway (Fig. 1E). Most plant RiPPs such as cyclotides, orbitides, PawS-derived peptides and cyclopeptide alkaloids are derived from precursor peptides, which are expressed as separate proteins from their PTM enzymes (Table 1). Recently, several BURP-domain-containing proteins (Section 3.6) have been characterized which can catalyze macrocyclic PTMs in core peptides within the same protein as the cyclase domain. This discovery showed that plant RiPP biosynthesis can involve autocatalysis,<sup>18</sup> a biochemical mechanism first observed for peptide-*N*-methyltransferases of fungal RiPP pathways in borosin biosynthesis.<sup>60,61</sup>

### 2.3. Proteolysis

Plant RiPP biosynthesis utilizes protein processing enzymes found in secretory pathways. For example, the biosynthesis of cyclotides involves asparaginyl endopeptidases (AEP), proteolytic enzymes often involved in processing vacuolar proteins like vacuolar storage protein albumins.<sup>62,63</sup> Precursor peptides with albumin domains are indeed observed for head-to-tail-macrocyclic plant RiPPs PawS-derived peptides, PawL-type orbitides, and cyclotides. The genetic, structural, and biochemical connection of plant RiPP chemistry to larger plant proteins indicates evolutionary roots of some plant RiPPs from functional motifs in such proteins.<sup>52,55</sup>

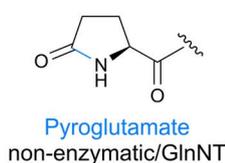
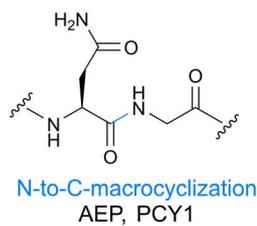
## 3. Plant RiPP classes

### 3.1. Cyclotides

Cyclotides are defined by a head-to-tail-macrocyclization and the presence of a cystine knot motif in their structure (Fig. 3). The founding member kalata B1 was discovered as an oxytocic agent in extracts of the Congolese plant *Oldenlandia affinis*, which is used as an herbal medicine to facilitate child birth during labor.<sup>80–83</sup> Since the discovery of *O. affinis* cyclotides, many other cyclotides have been characterized with 761 cyclotides documented in the cyclic plant peptide database CyBase<sup>26,84</sup> that establishes cyclotides as the largest plant RiPP family to date. For more details about cyclotides, we direct readers to several dedicated reviews.<sup>85,86</sup>



## Terminal modifications



## Side-chain-modifications

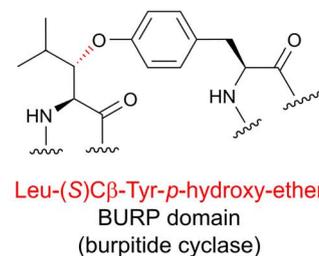
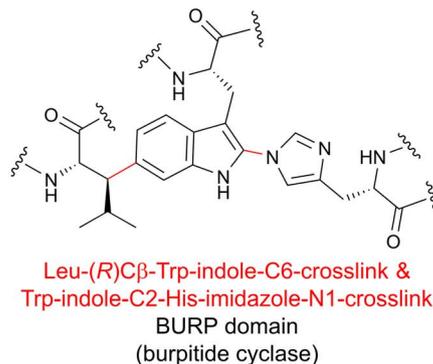
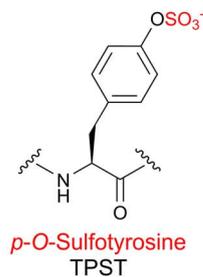
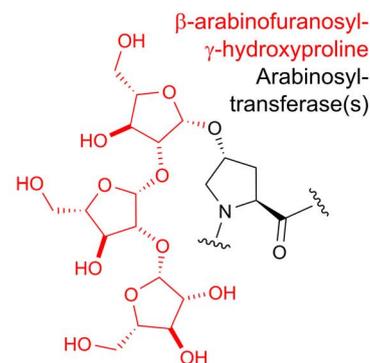


Fig. 2 Post-translational modifications in plant RiPPs. Terminal modifications are in blue and side-chain-modifications are shown in red. Abbreviations: AEP – asparaginyl endopeptidase, PCY1 – peptide cyclase 1, N-MT – N-methyltransferase, GlnNT – glutamine aminotransferase, PDI – protein disulfide isomerase, TPST – tyrosylprotein sulfotransferase.

**3.1.1 Structure.** Cyclotides are 28–37 amino acid-long, head-to-tail-macrocytic plant peptides with a ‘cyclic cystine knot’-motif formed by three disulfide bonds<sup>87</sup> (Fig. 3). The characteristic macrocyclic structure of kalata B1 was confirmed by NMR analysis in 1995<sup>88</sup> and the name cyclotide and cyclic cystine knot were subsequently introduced.<sup>87</sup> Kalata B1 and other cyclotide structures were further confirmed *via* synthetic studies.<sup>89</sup> Notably, the cystine knot of cyclotides was previously characterized in knottin plant peptides which have no head-to-tail-macrocytization.<sup>90</sup> Cyclotides have high stability against heat and enzymatic and chemical degradation.<sup>91</sup> Three subgroups of cyclotides have been defined based on primary and secondary structure: bracelet cyclotides have an all-*trans* backbone, Möbius cyclotides have a *cis*-proline residue, and trypsin inhibitor-type cyclotides have high sequence similarity to knottin-type trypsin inhibitors. An additional proposed subgroup are lysine-rich cyclotides.<sup>86,87,92</sup>

**3.1.2 Biosynthesis.** The ribosomal biosynthesis of cyclotides was first established in kalata B1.<sup>4</sup> In 2001, the Craik and Anderson labs examined the mRNA of *Oldenlandia affinis* and identified a cDNA termed *Oak1* that contained a core peptide sequence that perfectly matched kalata B1 (Fig. 3). The clone possessed an N-terminal endoplasmic reticulum signal sequence followed by an N-terminal leader peptide. Three additional clones, termed *Oak2–4*, were also found to code for alternative cyclotides. Notably, *Oak2* and *Oak4* possessed two and three separate core peptides, respectively (Fig. 1C). Each core was separated by a conserved N-terminal recognition sequence. *Oak1–4* also contains a conserved C-terminal follower

peptide after the core peptide sequence essential for processing. Shortly thereafter, additional cyclotide precursor peptides were discovered from different plants, each displaying a similar arrangement of conserved N-terminal leader peptide or recognition sequence, core peptide, and short C-terminal follower peptide.<sup>93–95</sup> Cyclotide precursor peptides range from one to seven core peptide sequences.<sup>5</sup> The core peptides usually start with a glycine at the N-terminus, end with an asparagine or aspartate at the C-terminus, and are flanked by N-terminal and C-terminal recognition sequences.<sup>4,86,93</sup> Cyclotide precursor peptides are mostly stand-alone precursor peptides, however a precursor peptide from *Clitoria ternatea* (Fabaceae) includes a proalbumin subunit.<sup>96,97</sup>

Once the signal sequence is cleaved, it is thought the disulfide bond formation is the first step of the biosynthesis. The cyclic cystine knot is proposed to be formed by a protein disulfide isomerase in the secretory pathway<sup>59</sup> (Fig. 3). Oxidative folding studies on kalata B1 core peptide determined that the cystine knot can also form correctly *via* a non-enzymatic pathway in the presence of reducing agents such as glutathione.<sup>57</sup> Next, the precursor peptides need to be processed to expose the primary amine of the N-terminal amino acid, which is typically glycine.<sup>26</sup> The Craik and Durek Labs used an activity-based isolation approach to identify a papain-like cysteine protease (*NbCysP6*) from *Nicotiana benthamiana* that could hydrolyze a short kalata B1 precursor peptide (LQLK-kB1) to acyclic kalata B1.<sup>76</sup> A homologous enzyme identified in the transcriptome of *O. affinis* named *OaRD21A* was also able to process the kalata B1 precursor peptide. Notably, the kalata B1



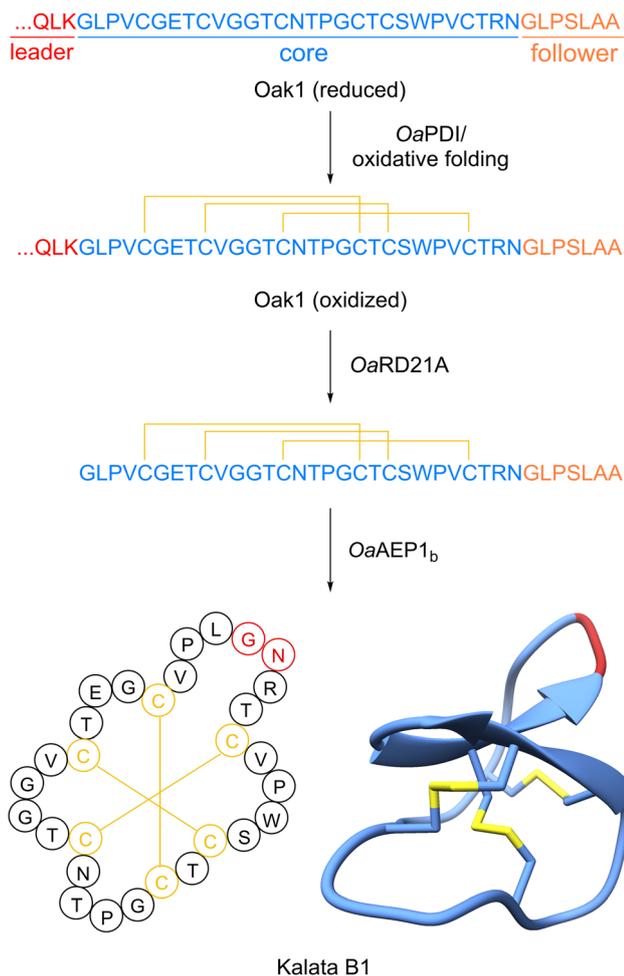


Fig. 3 Biosynthesis of cyclotide kalata B1. PDB accession: 1NB1.

core sequence needed to be properly folded with the requisite three disulfide bonds to prevent further degradation by *OaRD21A*. These experiments with *OaRD21A* support that the cyclic cystine knot forms first in the pathway.

The final step of the biosynthesis is the N- to C-terminus macrocyclization. The first cyclase was identified by focusing on the cyclotide producer *Clitoria ternatea*. Tam and colleagues conducted an activity-guided fractionation and isolation of an enzyme that could cyclize a short model peptide substrate to produce kalata B1.<sup>74</sup> Notably, this is not the native cyclotide produced by *C. ternatea*. Using this approach, they identified an asparaginyl endopeptidase (AEP) they called butelase 1 that efficiently cyclized a wide range of peptide substrates to make cyclotides, PawS-derived peptides, and peptides from animals. Detailed substrate scope assays showed this cysteine protease requires a C-terminal NHV motif to cyclize with an N-terminal glycine. Additional cyclizing AEPs were later screened and characterized from other plants. From the kalata B1 producer *Oldenlandia affinis*, the Anderson and Craik labs identified *OaAEP1<sub>b</sub>* (Fig. 3).<sup>75</sup> This enzyme was successfully heterogeneously expressed in *E. coli* and its activity was reconstituted *in vitro*. *OaAEP1<sub>b</sub>* efficiently cyclized the kalata B1 precursor peptide along with a variant that lacked the characteristic

disulfide bonds of cyclotides. To understand the structural basis that differentiates a normal AEP from a cyclase, detailed structural and mutagenesis studies were undertaken. A crystal structure of *OaAEP1* identified a mutation of a single Cys “Gatekeeping” residue to Ala that improves the cyclization  $k_{\text{cat}}$  by 160 fold.<sup>98</sup> Further studies in AEPs from the garden petunia (*Petunia × hybrida*) demonstrated that this Gatekeeping Cys mutation can be combined with swapping a short “marker of ligase activity” (MLA) loop from a robust AEP cyclase into a weak AEP cyclase to dramatically increase cyclase activity.<sup>99</sup> Moreover, this MLA region can be used diagnostically to predict if a given AEP is a cyclase.

Instead of separate enzymes for N-terminal processing and cyclizations, some cyclotides may employ a bifunctional enzyme. For example, examination of trypsin inhibitor cyclotides from *Momordica cochinchinensis* (MCoTI-I and MCoTI-II) by the Craik and Durek Labs identified an AEP named MCoAEP2.<sup>100</sup> This enzyme was able to efficiently both cleave the N-terminus of the MCoTI-II precursor peptide to expose the conserved Gly and then cyclize to generate MCoTI-II. These two reactions were pH dependent, so it was hypothesized the cellular trafficking of the precursor peptide into compartments with different pHs may be important in the maturation process.

**3.1.3 Bioactivity.** Insecticidal activities of cyclotides have been well characterized and indicate that these plant natural products function endogenously as defense compounds in source plants.<sup>4</sup> A common mode of action for insecticidal cyclotide activity is the disruption of plasma membranes in insect gut epithelial cells<sup>101</sup> or in insect cell culture.<sup>102</sup> In addition, cyclotides have shown molluscicidal activity against the rice pest golden apple snail.<sup>103</sup> The insecticidal activities of cyclotides have led to development and approval of a cyclotide-based insecticide Sero-X that includes a mixture of cyclotides from the seeds of the legume *Clitoria ternatea*.<sup>22,23</sup> Sero-X is a sustainable agrochemical pest control agent as it does not show cytotoxicity against pollinators. This success demonstrates the potential of cyclic plant peptides for sustainable agriculture.

Besides agricultural applicability, cyclotides have a broad range of medicinal activities with several pre-clinical and clinical lead structures to treat human diseases. Cyclotides have anthelmintic activity against gastrointestinal nematodes of sheep<sup>104</sup> and humans.<sup>105</sup> In addition, cytotoxicity of multiple cyclotides against cancer cells and in *in vivo* cancer models has been reported.<sup>92,106–114</sup> Several cyclotides have *in vitro* antibacterial activity, yet less activity in mouse models.<sup>89,115–118</sup> Antiviral activities, such as anti-HIV function, was determined early in the medicinal exploration of cyclotides<sup>119</sup> and a recent kalata B1 analog, T20K, has reached clinical trials as an immunosuppressive drug candidate for the treatment of multiple sclerosis (MS) based on promising immune modulatory activity in MS mouse models.<sup>24</sup>

The mode of action of cyclotides is based on their stabilized cyclic cystine knot structure and conserved residues that promote binding to therapeutic targets. The conserved residues found in most cyclotides are a positively charged residue and a glutamate. The positively charged residue can promote binding to target proteases, such as trypsin.<sup>120</sup>



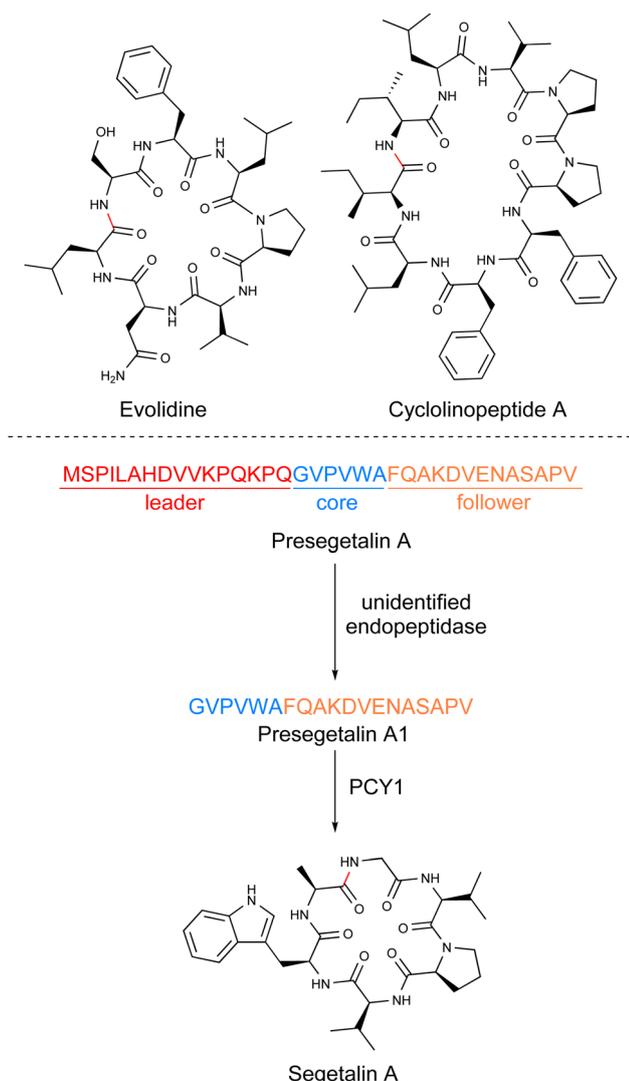


Fig. 4 Structures of iconic orbitides and proposed biosynthetic route for segetalin A.

### 3.2. Orbitides

Orbitides are small head-to-tail-macrocylic RiPPs with no disulfide bond (Fig. 4). They were originally described as Caryophyllaceae-type cyclic peptides because some founding members of these compounds, such as nonapeptide cyclolinopeptide A from flax (*Linum usitatissimum*), were discovered from plants in the pink family.<sup>121</sup> A wider taxonomic distribution of these compounds was revealed in the same and following decades<sup>122–124</sup> and the precursor genes of several orbitides were sequenced and confirmed recently which defined them as RiPPs.<sup>6,7,39,77,123,125,126</sup> In 2013, the name orbitides was proposed for these head-to-tail-cyclic RiPPs without disulfide bonds.<sup>1</sup> To date, nearly 200 orbitides have been deposited to CyBase<sup>26,84</sup> and are broadly distributed across plants as they are found in the Asterales, Caryophyllales, Lamiales, Magnoliales, Malpighiales, and Sapindales orders.

**3.2.1 Structure.** Orbitides are head-to-tail cyclized peptides between 5 and 16 amino acids in length.<sup>127</sup> They lack disulfide

bonds and are largely devoid of additional post-translational modifications outside a few examples including isoleucine hydroxylation<sup>128</sup> and *N*-methylation.<sup>129</sup> The core sequences are highly enriched in uncharged amino acids, with Gly and Pro particularly well represented. Glycine often serves as the *N*-terminus of the core peptide and, as such, is an important determinant in the macrocyclization reaction.

**3.2.2 Biosynthesis.** As orbitides are defined based on structural features and not biosynthetic enzymes, multiple distinct pathways appear to have evolved in plants. The biosynthesis for the Caryophyllaceae-type orbitides is the most well characterized with multiple detailed studies of the orbitide cyclase available. The PawL-type appears to mirror the PawS-derived peptide biosynthesis pathway (Section 3.3), however specific identification of biosynthetic enzymes remains to be completed. Finally, other orbitides may be synthesized by undiscovered pathways.

**3.2.2.1 Caryophyllaceae-type.** Based on the early work on cyclotides, it was hypothesized that orbitides would also be produced from a linear precursor peptide. To investigate this, the Covello lab focused on *Vaccaria hispanica* (*Saponaria vaccaria*), a well-known and prolific producer of orbitides.<sup>7</sup> They examined a cDNA-derived expressed sequence tags library for the presence of core sequences that directly mapped onto the structures of segetalins A through H. A total of 12 core peptide-containing precursor peptides about 35 amino acids in length were discovered. Each precursor contained conserved leader and follower regions flanking the variable core (Fig. 4). While the segetalins are found in precursors with only a single core, other orbitides such as those from *Annona muricata* (soursop),<sup>39</sup> appear to be biosynthesized from precursor peptides with multiple cores. The segetalin A precursor peptide was expressed in *Vaccaria hispanica* root cultures and detected by LC-MS to confirm the authentic precursor with multiple core peptides. The root extracts additionally processed the linear presegetalin A precursor peptide into the cyclic segetalin A product.<sup>7</sup>

The identification of biosynthetic enzymes of orbitides was done by the Covello lab, who used activity guided fractionation of *V. hispanica* seeds to look for processing of presegetalin A.<sup>77</sup> Their approach led to two enzyme activities: (1) cleavage of the conserved *N*-terminus by an unidentified oligopeptidase 1 (OLP1) and (2) cyclization of the presegetalin A1 by peptide cyclase 1 (PCY1)<sup>130</sup> (Fig. 4). *In vitro* assays of heterologously expressed and purified PCY1 validated that this enzyme was responsible for cyclization to form the final segetalin A product. PCY1 could process a diverse array of orbitide precursor peptides into the cyclic products and was tolerant to mutations in the core region.

While the identity of OLP1 has not been firmly established, PCY1 has been extensively studied. Crystallization by the Nair group revealed PCY1 to be a member of the prolyl oligopeptidase (POP) family of serine enzymes.<sup>130</sup> Co-crystal structures of PCY1 with the presegetalin A1 substrate along with mutagenesis studies demonstrated that the conserved C-terminal follower peptide was important for recognition by PCY1.<sup>130</sup> They hypothesized that PCY1 cleaves the core peptide from the follower to form an acyl-enzyme intermediate to facilitate



macrocyclization with the N-terminal amine of the core peptide to form the final product. Further crystallographic and mechanistic studies by the Naismith group gave detailed insights into substrate recognition by PCY1.<sup>131</sup> Notably, their work has characterized substrate tolerance for PCY1 to function as a versatile peptide cyclase.

**3.2.2.2 PawL-type.** The PawL-type orbitides were discovered using a bioinformatic approach. While characterizing the evolution of the PawS core sequence (Section 3.3), the Mylne group identified the PawS-Like genes (PawL).<sup>124</sup> Similar to PawS, the PawL core peptide is contained within a two-subunit seed storage albumin (Fig. 5). Post-translational processing by asparaginyl endopeptidases separate the seed storage albumin proteins and cyclize these core sequences. A combination of both transcriptomics and metabolomics validated that these PawL core peptides were processed into cyclic peptides.<sup>123</sup> The biosynthesis is distinct from the Caryophyllaceae-type orbitides, yet the N- to C- macrocyclic nature and lack of disulfide bonds define PawL-derived peptides as orbitides.

**3.2.2.3 Others.** While biosynthetic enzymes have been demonstrated or proposed for the maturation of Caryophyllaceae- and PawL-type orbitides, other orbitides do not have a clearly defined pathway. For example, transcriptomics-based analysis of *Melicope xanthoxyloides* by the Mylne group enabled the discovery of the evolidine precursor peptide.<sup>122</sup> Characteristic of orbitides, evolidine is a small N- to C-macrocyclic peptide without any disulfide bonds. Like many of the Caryophyllaceae-type orbitides, the evolidine is

biosynthesized from a stand-alone precursor peptide with a single core sequence homologous to *Citrus clementina*. However, a lack of conserved proline/alanine in the core region along with differences in the C-termini compared to the precursor peptides from *V. hispanica* led the authors to suggest that a PCY1 homolog is likely not responsible for cyclization of the core sequences, and the cyclase for evolidine remains to be determined.<sup>122</sup>

Similarly, the precursor peptides in *Linum usitatissimum* (flax) are distinct from those processed by PCY1. *L. usitatissimum* is a prolific producer of orbitides, specifically cyclolinopeptides/linusorbs. Efforts from the Reaney group to link the isolated orbitides to precursor peptides have identified genes that encode for multiple core sequences flanked by conserved recognition sequences.<sup>125,132,133</sup> A recent bioinformatic study has explored these results to identify five separate precursors peptides from *L. usitatissimum*, along with 25 additional linusorb-like domains, that have features consistent with orbitide precursor peptides;<sup>8</sup> a biosynthetic route remains to be elucidated for these orbitides.

**3.2.3 Bioactivity.** The physiological function of orbitides in plants is unknown, yet they have demonstrated a wide range of biological activities. In the 1990s, studies completed with cyclolinopeptide A ([1-9-N $\alpha$ C]-linusorb B3) to investigate its immunosuppressive activity found it was comparable to the clinically approved cyclosporine A in mouse models.<sup>134</sup> Detailed mechanistic studies later indicated that cyclolinopeptide A inhibits peptidyl-prolyl *cis-trans* isomerase (PPIase) that has the downstream effect of blocking the transcription of interleukin-2 and other cytokines.<sup>135,136</sup> Other orbitides, hymenistatin I<sup>137</sup> and cycloleonorin,<sup>138</sup> were also found to be immunosuppressive, with the latter demonstrating inhibition of human peripheral-blood lymphocytes response (IC<sub>50</sub> of 28 ng mL<sup>-1</sup>).

The segetalin series of orbitides were originally discovered based on their estrogen-like bioactivity. The seeds from *V. hispanica* (*Vaccaria segetalis*) have been traditionally used to activate blood flow, promote milk secretion, and treat amenorrhea, and breast infections. Using these purported effects as a guide, segetalin A and B were isolated.<sup>139,140</sup> Subsequent bioactivity testing of these two orbitides demonstrated a positive effect on rat uterine weight and contraction strength, further supporting their estrogen-like effects. Another study in 1997 found that two new segetalins (G and H) also isolated from *V. hispanica* possessed estrogen-like activity due to their ability to increase uterus growth in rats.<sup>141</sup>

Numerous orbitides have been shown to have anti-inflammatory activity.<sup>142,143</sup> Of particular note are the cyclomontanin series with cyclomontanin A active in anti-inflammatory activity assays at 1 and 3  $\mu\text{g mL}^{-1}$ .<sup>143</sup> Another recent study on linusorbs (flaxseed orbitides) showed that these orbitides can inhibit the NF- $\kappa$ B pathway and production of pro-inflammatory cytokines at concentrations ranging from 1–4  $\mu\text{M}$ .<sup>144</sup>

Orbitides have also been investigated for anticancer properties. Glauccyclopeptide B and cherimolacyclopeptide G, were found to be cytotoxic against a nasopharyngeal carcinoma cell line with IC<sub>50</sub> values of 16.5  $\mu\text{M}$  and 0.52  $\mu\text{M}$ , respectively.<sup>145</sup> Further examples of cytotoxic orbitides including yunnanin A

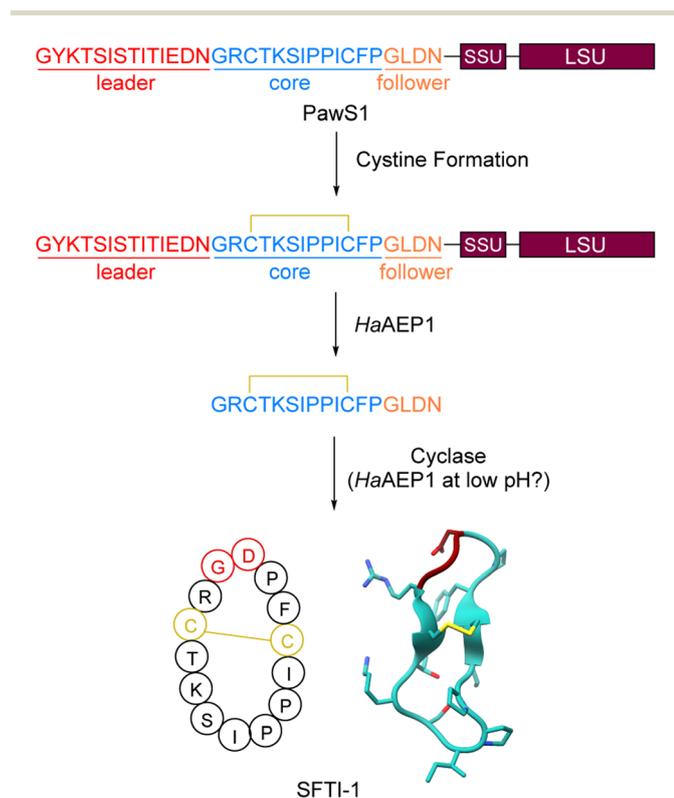


Fig. 5 Biosynthesis of SFTI-1. SSU and LSU are small and large subunits, respectively, of albumin domain. PDB accession: 1JBL.



and yunnanin C showed cytotoxicity with  $IC_{50}$  values of 2.1 and 2.2  $\mu\text{g mL}^{-1}$ , respectively.<sup>146</sup> Other orbitides such as [1–9-N $\alpha$ C]-linusorb (LO) B2 and [1–9-N $\alpha$ C]-linusorb B3 (cyclolinopeptide A) demonstrated anti-proliferative properties by the reduction of cancer cell viability and increased apoptosis, and induction of cell cycle arrest.<sup>147,148</sup> A later study found that linusorb B3 modulated the expression of apoptosis related genes and suppression of cell motility by inhibiting actin polymerization in glioblastoma cells.<sup>149</sup> Numerous other bioactivity studies have been completed with orbitides and variety of properties relevant to human health have been observed including vasorelaxant activity (segetalin F–H),<sup>150</sup> moderate antibacterial activity (evoludine),<sup>142</sup> and weak anti-plasmodial activity (ribifolin).<sup>151</sup>

### 3.3. PawS-derived peptides

PawS-derived peptides (PDPs) are head-to-tail-macrocylic RiPPs with one disulfide bond. The founding member of the PDPs, sunflower trypsin inhibitor (SFTI-1 Fig. 5), was discovered in 1999 from sunflower seeds (*Helianthus annuus*).<sup>152</sup> Additional isolation and genetic mining efforts have revealed around 20 total PDPs.<sup>52,124,153</sup> These natural products do not appear to be widespread in plants and are largely localized around species phylogenetically related to *H. annuus*.<sup>124</sup> We recommend a review that focuses on SFTI-1 with many details on the extensive engineering efforts of SFTI-1<sup>154</sup> and another review that discusses PDPs more broadly<sup>53</sup> for further reading.

**3.3.1 Structure.** PDPs are head-to-tail-macrocylic peptides containing a single disulfide bond. All cyclic PDPs to date are formed between an N-terminal Gly and a C-terminal Asp (Fig. 5). While they share conceptual similarities to cyclotides, they are distinct in that they are smaller than cyclotides (12–18 amino acids) and contain only one disulfide bond compared to the three found in cyclotides. Many PDPs are rich in proline and contain a tandem *cis* and *trans* proline pair. PDPs often possess a Bowman–Birk inhibitor<sup>155</sup> sequence loop mimic that is responsible for the observed trypsin inhibition bioactivity.<sup>124</sup>

**3.3.2 Biosynthesis.** By searching for the amino acid sequence of SFTI-1 within *H. annuus* cDNA-derived expression sequence tags, Mylne *et al.* were able to identify a region within a protein they termed PawS1 (preproalbumin with SFTI-1) that was an exact match to the SFTI-1 sequence.<sup>52</sup> PawS1 is a small 151-amino acid protein that contains three features: an ER signal peptide, the SFTI-1 precursor peptide, and a 2S seed albumin (Fig. 5). Subsequent NMR studies on PawS1 confirmed this structural organization.<sup>156</sup> Seed albumins are used as a nutrient source during development.<sup>157,158</sup> Bioinformatic searching revealed a second gene termed *pawS2* with the same structure that also matched the same tripartite organization. Mass spectrometric analysis confirmed that the SFTI-Like1 precursor peptide of PawS2 matured into a cyclic peptide like SFTI-1. Using an *A. thaliana* expression system, the authors demonstrated that asparaginyl endopeptidases mature PawS1 into both albumin and SFTI-1. Notably, the native *A. thaliana* AEPs could accomplish the SFTI-1 production and a specialized AEP is not essential.

Additional work from the Mylne and Craik labs further dissected the maturation process of SFTI-1.<sup>159</sup> Examination of the *H. annuus* transcriptome revealed at least five AEPs, and one was particularly upregulated, *HaAEP1*. *In vitro* reconstitution of *HaAEP1* showed hydrolysis after Asn residues to generate a SFTI-1+Follower peptide, but no cyclization (Fig. 5). While the AEP responsible for cyclizing SFTI-1 remained elusive in *H. annuus*, the authors also screened other known AEPs from different organisms. Notably, the AEP from *Canavalia ensiformis* (jack bean) termed *CeAEP1* was able to fully process a 25-amino acid peptide into cyclic SFTI-1. *CeAEP1* acts first as a typical AEP peptidase to trim the peptide to reveal the N-terminal Gly and then as a macrocyclase to form the Gly1-Asp14 peptide bond. While *CeAEP1* only forms the cyclic SFTI-1 product 15% of the time, the acyclic SFTI-1 is rapidly degraded *in vivo*. It remains unclear whether the authentic *H. annuus* cyclase would have similar cyclization efficiency.

Later studies examined processing of the full length PawS1 protein.<sup>156</sup> As before, *HaAEP1* was able to function as a protease and cleave both the proalbumin and liberate a linear SFTI-1 containing peptide. More recently, a crystal structure of *HaAEP1* and subsequent characterization revealed that the maturation process of *HaAEP1* and assay conditions strongly influence cyclization activity.<sup>160</sup> When *HaAEP1* is purified as a zymogen, matured by autoactivation at pH 4.0, and assayed at pH 6.5, it is able to catalyze the formation of cyclic SFTI-1 (Fig. 5).<sup>160</sup> This pH dependent cyclization appears to be a conserved feature of AEPs as it was observed in AEPs from other organisms.

**3.3.3 Bioactivity.** SFTI-1 is a potent trypsin inhibitor with a  $K_i$  of 0.1 nM.<sup>152</sup> Initial isolation and co-crystallization with bovine trypsin revealed binding in the active site similar to the Bowman–Birk inhibitor family of natural products.<sup>152</sup> Specifically, the lysine of SFTI-1 fits into the P1 site of bovine trypsin and adopts a conformation favorable for hydrolysis. Upon cleavage of the labile Lys5–Ser6 bond of SFTI-1 by trypsin, cyclization of the acyl-enzyme intermediate to re-form intact SFTI-1 outcompetes hydrolysis.<sup>154,161</sup> To evaluate the physiological function of PDPs, bioactivity assays of SFTI-1 with gut extracts of the cotton bollworm (*Helicoverpa armigera*) also showed inhibition indicating that these natural products may serve to protect the plant from insect predation.<sup>124</sup>

### 3.4. Cysteine-rich peptides

In addition to cyclotides, plants produce many peptides without head-to-tail-macrocylicization, but with multiple cysteines that engage in macrocylicization *via* disulfide bonds. The classification of cysteine-rich peptides (CRPs) is historically based on activity (antimicrobial peptides, defensins, rapid alkalization factors), structure (knottins, thionins, snakins), or chemotaxonomy of founding members (heveins). Given current knowledge of cysteine-rich peptide (CRP) chemistry and biosynthesis, the most applicable classification scheme for CRPs is the number and pattern of disulfide bonds, which translates into distinct structures (Fig. 6). CRP precursor peptide sequences can differ significantly in size



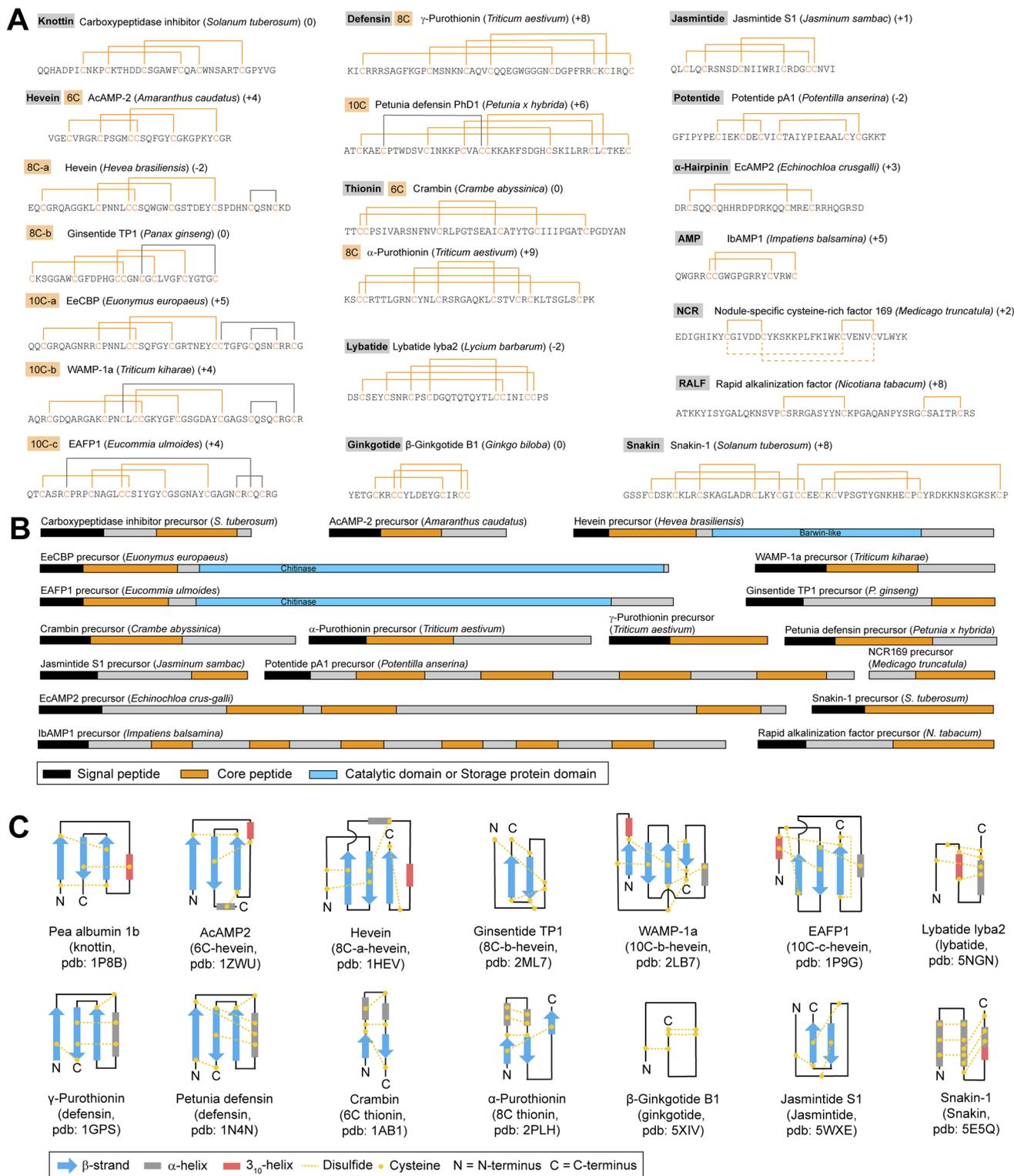


Fig. 6 Chemical structures and precursor peptides of cysteine-rich plant peptides. (A) Sequences and disulfide bond patterns of representatives of CRP subclasses. Class-defining DSBs and cysteines are highlighted in orange. Solid lines in gray are non-class-defining DSBs to the CRP. Peptide net charges are shown behind species names. (B) Precursor peptides of CRP subclass representatives. (C) Topology diagrams of CRP subclass representatives based on experimentally determined structures.

and domain structure within individual classes so that core peptide sequence and its cysteine pattern appear to be the best features for CRP prediction (Fig. 6). In addition, the net charge

of a core peptide sequence is often considered for CRP characterization (Fig. 6A). Multiple excellent reviews have been published for CRPs.<sup>162–164</sup>



**3.4.1 Structure.** The number of disulfide bonds in plant CRPs ranges from two (rapid alkalization factors) to six (snakins), and from 18 to >50 amino acids total. CRPs discussed in this review include the following subclasses: knottins,<sup>90,165,166</sup> thionins,<sup>167</sup> defensins,<sup>67,168</sup> snakins,<sup>37,169</sup> heveins,<sup>170,171</sup>  $\alpha$ -hairpinins,<sup>172</sup> potentides,<sup>40</sup>  $\beta$ -ginkgotides,<sup>173</sup> jasmintides,<sup>174</sup> lybatides,<sup>175</sup> impatiens AMPs,<sup>42</sup> nodule cysteine-rich peptides,<sup>176</sup> and rapid alkalization factors.<sup>73</sup> Another antimicrobial peptide class which is often included in CRP classification is lipid-transfer proteins.<sup>162,177</sup> Lipid-transfer proteins are 9–10.5 kDa and more appropriately classified as a post-translationally modified protein, as they exceed the upper limit of 10 kDa for RiPPs.<sup>1</sup>

Knottins feature a cystine knot motif, also called inhibitor cystine knot<sup>178</sup> or knottin fold,<sup>166</sup> based on three cystines in cyclotides: two disulfide bonds form a loop and the third one passes through the first two disulfides, producing a knot with five loops between the disulfides in the RiPP.<sup>90</sup> The core peptide cysteine pattern of knottins is C(1)X<sub>3–6</sub>C(2)X<sub>5</sub>C(3)X<sub>3–5</sub>C(4)X<sub>1–2</sub>C(5)X<sub>5–6</sub>C(6), where the numbers in parentheses indicate the order of cysteines in the core and X<sub>n</sub> between Cys indicates n non-cysteine residue(s) between the next cysteine based on characterized CRP structures. The knottin cysteine pattern yields disulfide bonds between C(1)–C(4), C(2)–C(5) and C(3)–C(6) (Fig. 6A). Knottins are 26 to 37 amino acids long (~2.7 to 5 kDa). The founding member of the knottins is the potato carboxypeptidase inhibitor which yielded the first structure of a cyclic cystine knot peptide.<sup>90,179</sup> The name ‘knottin’ was introduced in a study of a microprotein with knotted cystine topology from *Ecballium elaterium* seeds.<sup>166,180</sup> Knottins such as trypsin inhibitors from *Momordica cochinchinensis* (MCoTI-3/5/6) can have an N-terminal pyroglutamate derived from an N-terminal glutamine<sup>5</sup> and they are generally neutral or positively charged. Knottin structures such as pea albumin 1b knottin consist of three antiparallel  $\beta$ -strands and one  $3_{10}$ -helix crosslinked by the three disulfide bonds<sup>181</sup> (Fig. 6C). Knottins have been characterized from Solanaceae,<sup>90,182</sup> Fabaceae,<sup>181,183,184</sup> Cucurbitaceae,<sup>120</sup> Apocynaceae,<sup>185</sup> Rubiaceae,<sup>186,187</sup> and Poaceae.<sup>188</sup>

Thionins belong to a group of CRPs termed antimicrobial peptides (AMPs) that are characterized by positive charge and antimicrobial activity due to interaction with negatively charged bacterial membranes.<sup>189</sup> Thionins are a plant AMP class with three to four disulfide bonds. The core cysteine pattern of thionins with three disulfide bonds (DSBs) is C(1)C(2)X<sub>11</sub>C(3)X<sub>9</sub>C(4)X<sub>5</sub>C(5)X<sub>7</sub>C(6) with DSB between C(1)–C(6), C(2)–C(5) and C(3)–C(4). The core cysteine pattern of thionins with four DSBs is C(1)C(2)X<sub>7</sub>C(3)X<sub>3</sub>C(4)X<sub>8</sub>C(5)X<sub>3</sub>C(6)X<sub>1</sub>C(7)X<sub>7</sub>C(8) with DSBs between C(1)–C(8), C(2)–C(7), C(3)–C(6) and C(4)–C(5) (Fig. 6A). Thionins are 44 to 47 amino acids long (4.7 to 5.3 kDa). In addition, several variants have been identified or predicted that have uneven numbers of cysteines in the core peptides, and it remains unclear how the structure would be affected in this case.<sup>190</sup> Thionins are generally positively charged with few neutral exceptions.<sup>191</sup> The first characterized plant thionins were  $\alpha/\beta$ -purothionin isolated from wheat (Poaceae) as agents toxic to baker's yeast<sup>192,193</sup> with several others identified from

Brassicaceae,<sup>194–196</sup> Liliaceae,<sup>197</sup> Papaveraceae,<sup>190</sup> Ranunculaceae,<sup>198</sup> and Santalaceae.<sup>199–202</sup> Thionins share a fold consisting of two antiparallel  $\beta$ -strands and two  $\alpha$ -helices stabilized by disulfide bonds (Fig. 6C).<sup>167,193</sup>

Defensins are host defense peptides with four disulfide bonds and general positive charge which have also been called ‘ $\gamma$ -thionins’ in the literature.<sup>67</sup> Plant defensins belong to the *cis*-defensin class which has two conserved disulfides crosslinking a central  $\beta$ -strand to an  $\alpha$ -helix<sup>203</sup> (Fig. 6C). Plant defensins usually have four DSBs with a cysteine pattern C(1)X<sub>10–11</sub>C(2)X<sub>3–5</sub>C(3)X<sub>3</sub>C(4)X<sub>9–11</sub>C(5)X<sub>4–9</sub>C(6)X<sub>1</sub>C(7)X<sub>2–3</sub>C(8) with DSBs between C(1)–C(8), C(2)–C(5), C(3)–C(6) and C(4)–C(7). Phd1, a *Petunia* defensin, has an additional DSB and a cysteine pattern of C(1)X<sub>3</sub>C(2)X<sub>6</sub>C(3)X<sub>5</sub>C(4)X<sub>2</sub>C(5)C(6)X<sub>9</sub>C(7)X<sub>6</sub>C(8)X<sub>1</sub>C(9)X<sub>3</sub>C(10) with DSBs between C(1)–C(10), C(2)–C(5), C(3)–C(7), C(4)–C(8) and C(6)–C(9) (Fig. 6A). The general size of plant defensins is 45 to 54 amino acids but several representatives with up to 71 amino acids are known.<sup>204,205</sup> The plant defensin fold contains a  $\beta$ -sheet and an  $\alpha$ -helix and is also referred to as a cystine-stabilized  $\alpha\beta$ -motif<sup>203,206</sup> (Fig. 6B). The first plant RiPPs recognized as plant defensins were two  $\gamma$ -purothionins that were initially classified as thionins and later structurally differentiated as defensins.<sup>67,168</sup> To date, plant defensins were characterized from Asteraceae,<sup>207</sup> Brassicaceae,<sup>208–212</sup> Caryophyllaceae,<sup>213</sup> Fabaceae,<sup>213–218</sup> Oleaceae,<sup>219</sup> Pentadiplandraceae,<sup>204</sup> Pinaceae,<sup>220</sup> Poaceae,<sup>67,168,205,221–223</sup> Ranunculaceae,<sup>224</sup> Sapindaceae,<sup>225</sup> Saxifragaceae,<sup>226</sup> Solanaceae,<sup>227,228</sup> and Vitaceae.<sup>229</sup>

Heveins are a diverse class of CRPs which derive their name from the founding member hevein, a wound-induced anti-fungal peptide from the rubber tree (*Hevea brasiliensis*).<sup>170</sup> Heveins have three to five cysteines with a foundational cysteine pattern C(1)X<sub>4–8</sub>C(2)X<sub>4–6</sub>C(3)C(4)X<sub>4–5</sub>C(5)X<sub>4–7</sub>C(6), connected by the three DSBs *via* C(1)–C(4), C(2)–C(5) and C(3)–C(6). Heveins have a general size of 29 to 45 amino acids (3 to 5 kDa) and are usually positively charged except for a few neutral representatives. Two subclasses of heveins, hevein-type heveins and ginsentides, have eight cysteines (8C-hevein) in the core peptide and thus an additional DSB with variable positioning to the core DSB motif (Fig. 6A). An additional three subclasses of heveins are represented by EeCBP from *Euonymus europaeus*, WAMP-1a from *Triticum kiharae*, and EAFP1 from *Eucommia ulmoides*, which all have ten cysteines (10C-heveins) in their core peptides and form five DSBs. The additional cysteines in 8C- and 10C-heveins are generally located C-terminally of the core hevein cysteine pattern (Fig. 6A). Heveins typically contain an antiparallel  $\beta$ -sheet, a  $3_{10}$ -helix, and an  $\alpha$ -helix, whereas ginsentides thus far lack helices (Fig. 6C). Hevein CRPs have been characterized from Amaranthaceae,<sup>230–235</sup> Araliaceae,<sup>65</sup> Asteraceae,<sup>236</sup> Cactaceae,<sup>237</sup> Caryophyllaceae,<sup>238,239</sup> Celastraceae,<sup>240</sup> Convolvulaceae,<sup>241</sup> Cycadaceae,<sup>242</sup> Eucommiaceae,<sup>54,243</sup> Ginkgoaceae,<sup>173</sup> Malvaceae,<sup>244,245</sup> Nyctaginaceae,<sup>246</sup> Phytolaccaceae,<sup>247</sup> Poaceae,<sup>248–250</sup> and Polygonaceae.<sup>251</sup>

Several CRP subclasses with unique cysteine patterns of three or four DSBs were discovered recently in plants. Lybatides are 32 to 33 amino acid long CRPs (~3.5 kDa) with an 8-cysteine pattern of C(1)X<sub>3</sub>C(2)X<sub>3</sub>C(3)X<sub>2</sub>C(4)X<sub>10</sub>C(5)C(6)X<sub>3</sub>C(7)C(8) forming the four DSBs C(1)–C(6), C(2)–C(8), C(3)–C(7) and C(4)–



C(5)<sup>175</sup> (Fig. 6A). Lybatides were discovered from *Lycium barbarum* (Solanaceae). The structure of lybatide lyba2 revealed a fold with disulfide-stapled helices (Fig. 6C).  $\beta$ -Ginkgotides are 18–20 amino acid-long CRPs (~2.5 kDa) from *Ginkgo biloba* (Ginkgoaceae) with three DSBs of C(1)–C(4), C(2)–C(6) and C(3)–C(5) derived from the cysteine pattern C(1)X<sub>2</sub>C(2)C(3)X<sub>6</sub>C(4)X<sub>2</sub>C(5)C(6) (Fig. 6A). The NMR structure of  $\beta$ -ginkgotide B1 did not show any helices or  $\beta$ -strands<sup>252</sup> (Fig. 6C). Jasmintides are 27 amino acid long CRPs (3.1 kDa) isolated from *Jasminum sambac* (Oleaceae), with six cysteines and three disulfides between C(1)–C(5), C(2)–C(4) and C(3)–C(6) from cysteine pattern C(1)X<sub>2</sub>C(2)X<sub>5</sub>C(3)X<sub>6</sub>C(4)X<sub>3</sub>C(5)C(6) (Fig. 6A) with more predicted variations.<sup>174</sup> Two jasmintides have been elucidated by solution NMR that are characterized by two antiparallel  $\beta$ -strands<sup>71,174</sup> (Fig. 6C). A jasmintide from *Achyranthes bidentata* was identified with an unusual N-terminal fructosylation and three antiparallel  $\beta$ -strands.<sup>253</sup> Lastly, a subclass of CRPs from the roots of *Potentilla anserina* has three DSBs called potentides,<sup>40</sup> characterized by C(1)–C(3), C(2)–C(6), and C(4)–C(5) DSBs in a C(1)X<sub>3</sub>C(2)X<sub>2</sub>C(3)X<sub>2</sub>C(4)X<sub>10</sub>C(5)X<sub>1</sub>C(6) cysteine-pattern (Fig. 6A). Potentides are 35 amino acid long and to date no structure has been determined. While lybatides and  $\beta$ -ginkgotides are slightly acidic, jasmintides and potentides are neutral to slightly basic.

Snakins are a CRP subclass that was discovered in the late 1990s from potato. Snakins have 12 cysteines and six disulfide bonds (C(1)–C(7), C(2)–C(5), C(3)–C(4), C(6)–C(12), C(8)–C(11), C(9)–C(10)) in the cysteine pattern C(1)X<sub>3</sub>C(2)X<sub>3</sub>C(3)X<sub>8</sub>C(4)X<sub>3</sub>C(5)X<sub>2</sub>C(6)C(7)X<sub>2</sub>C(8)X<sub>1</sub>C(9)X<sub>11</sub>C(10)X<sub>2</sub>C(11)X<sub>12</sub>C(12).<sup>37</sup> Snakins are large peptides of 63 to 66 amino acids and are positively charged. The first characterized snakin CRP was snakin-1, with antifungal properties. The structure revealed a fold with multiple DSB-stapled helices<sup>169</sup> (Fig. 6C).

Finally, plants produce several RiPPs with two DSBs.  $\alpha$ -Hairpinins and impatiens AMPs have both two DSBs with the pattern C(1)–C(4), C(2)–C(3) and C(1)–C(3), C(2)–C(4), respectively, from cysteine patterns C(1)X<sub>3</sub>C(2)X<sub>11</sub>C(3)X<sub>3</sub>C(4) and C(1)C(2)X<sub>8</sub>C(3)X<sub>3</sub>C(4), respectively (Fig. 6A)<sup>42,172</sup> with more predicted variations for  $\alpha$ -hairpinins.<sup>254</sup>  $\alpha$ -Hairpinins such as EcAMP2 were discovered from *Zea mays* (Poaceae),<sup>255</sup> later described from Proteaceae,<sup>254</sup> are 26–67 amino acids long, and generally positively charged with a few negatively charged exceptions. AMPs from *Impatiens balsamina* are 20 amino acid long, positively charged, and have a C-terminal glutamine-derived pyroglutamate.<sup>42</sup> Another plant CRP subclass with two DSBs are rapid alkalization factors (RALFs) (Fig. 6A). RALF from *Nicotiana tabacum* was characterized to have a C(1)–C(2) and C(3)–C(4) DSB pattern in a C(1)X<sub>9</sub>C(2)X<sub>12</sub>C(3)X<sub>5</sub>C(4) cysteine pattern.<sup>73</sup> RALF peptides are important signaling peptides, and it was discovered based on its activity to alkalize tobacco cell suspension cultures and inhibit root growth.<sup>73,256</sup> Nodule-specific cysteine-rich peptides (NCR) of 36–58 amino acids have been characterized from Fabaceae such as *Medicago truncatula* and *Pisum sativum*.<sup>72</sup> The oxidative folding and solution NMR analysis of the RiPP NCR169 showed that two DSB patterns can form, *i.e.* C(1)–C(2), C(3)–C(4) or C(1)–C(3), C(2)–C(4), from the C(1)X<sub>5</sub>C(2)X<sub>12</sub>C(3)X<sub>4</sub>C(4) cysteine pattern *in vitro*,

where the former DSB pattern was found only in the presence of reduced glutathione<sup>176</sup> (Fig. 6A, solid lines).

**3.4.2 Biosynthesis.** Most precursor peptides of CRPs are single-core precursors (Fig. 6B and Table 1) and have an N-terminal signal peptide indicating processing through the secretory pathway. The multi-core exceptions are precursor peptides of potentides,  $\alpha$ -hairpinins and AMPs from *Impatiens balsamina*. Several hevein precursor peptides include additional protein domains: the hevein precursor includes a C-terminal Barwin-like domain,<sup>14,257</sup> whereas the precursor peptides of 10C-heveins from *Eucommia ulmoides* and *Euonymus europaeus* include C-terminal chitinase domains.<sup>16,54</sup> Thionin precursor peptides contain a C-terminal acidic domain without a known function.<sup>190</sup> Some CRPs, including several knottins, heveins, and jasmintides, possess an N-terminal pyroglutamate which is derived from an N-terminal core glutamine.

The PTM event in CRPs is the formation of disulfide bonds during the folding process of a core peptide, whereby the peptide sequence and cellular environment dictates folding and disulfide formation. DSBs can form non-enzymatically under increasingly oxidizing conditions of the secretory pathway in plants and enzymatically by protein disulfide isomerases (PDI), which can reshuffle established disulfides, typically from incorrectly formed DSBs to the native DSBs in the peptide. The improvement of CRP disulfide formation was shown *in vitro* on a core peptide substrate of *Amaranthus*  $\alpha$ -amylase knottin-type inhibitor (AAI) by PDI DsbC. As for kalata B1 disulfide formation by OaPDI, the folding process towards  $\alpha$ -amylase knottin-type inhibitor does not require a PDI, however its inclusion dramatically increases forming a correctly folded RiPP.<sup>58,59</sup> Transformation of a linear reduced CRP core towards the most thermodynamically stable oxidized (native) structure is generally catalyzed by the presence of a redox pair, such as reduced and oxidized glutathione or cysteine and cystine.<sup>258,259</sup> Two general oxidative folding pathways have been described for CRPs: scrambled and native folding. In the scrambled pathway, intermediates with non-native DSBs are formed and reshuffled to native DSBs of the native structure. Examples for the scrambled pathway are knottins such as *Amaranthus*  $\alpha$ -amylase inhibitor and potato carboxypeptidase inhibitor.<sup>58,259</sup> In the native folding pathway, intermediates with native DSBs are formed without the need for reshuffling. Examples include *Echallium elaterium* trypsin inhibitor II (EETI II)<sup>260</sup> and *Momordica charantia* trypsin inhibitor MCh-I.<sup>261</sup> In the knottin EETI-II folding pathway, two intermediates were characterized which were derived through sequential formation of the C(3)–C(6) and then C(2)–C(5) disulfide bridges before the final formation of the C(1)–C(4) disulfide. The investigation of such CRP folding pathways can inform their synthesis from synthetic linear peptide substrates with protection strategies mimicking the native folding pathway of a CRP sub-class.<sup>262</sup>

Another important step in CRP biosynthesis is proteolytic processing of precursor peptides. It is hypothesized that proteolysis occurs after disulfide formation by proteases in the secretory pathway for most of the CRPs. An example is the characterization of a thionin-processing proteinase from barley



leaf. The protease responsible for cleavage of an oxidized core peptide in a precursor peptide *in vitro* to release a mature thionin is a subtilisin-like vacuolar protease. The corresponding studies revealed that the thionin precursor was processed in the vacuoles of barley leaves from which the precursor processing protease was isolated.<sup>263,264</sup> Mature barley thionins have been localized in vacuoles and in cell walls of leaf cells.<sup>265,266</sup>

**3.4.3 Bioactivity.** Many cysteine-rich plant peptides are part of the innate immune response of plants against fungal and bacterial pathogens and insect attacks. Knottins are potent inhibitors of fungal and insect proteases.<sup>267–270</sup> Several heveins have inhibitory activity against phytopathogenic fungi as well. For instance, hevein has a chitin-binding domain, which can interact with chitin of pathogenic fungi, and further inhibit fungal metalloproteases like chitinases that target plant defense proteins.<sup>232,271</sup> Thionins also have inhibitory activity against phytopathogenic bacteria and fungi,<sup>266,272,273</sup> which was first realized when  $\alpha/\beta$ -purothionins were discovered as inhibitory agents of baker's yeast.<sup>274</sup> Thionin is hypothesized to target and disrupt negatively charged phospholipids in cell membranes as mode of action.<sup>275</sup> Plant defensins exhibit antifungal, antibacterial and insecticidal activity,<sup>276–278</sup> and can inhibit insect  $\alpha$ -amylase *in vitro*.<sup>279</sup> Jasmintides have antifeedant activity against mealworm *Tenebrio molitor*,<sup>174</sup> and maize  $\alpha$ -hairpinins and snakins also showed antibacterial and antifungal activity against phytopathogens.<sup>37,255,280</sup> The main mode-of-action hypothesis for snakins is non-specific pore-forming activity in tested bacterial and fungal membranes.<sup>281</sup> This mode-of-action is supported by perforation of fungal biomembranes in the presence of snakin-2.<sup>282</sup> Full antifungal bioactivity is achieved by enantiomer D-snakin-1 compared to L-snakin-1 indicating a non-chiral mode-of-action such as interaction with negatively charged membrane lipids.<sup>169</sup> The snakin target interaction for membrane pore formation remains to be determined. Several nodule-specific cysteine-rich peptides function to control bacterial differentiation during symbiosis of nitrogen-fixing bacteria in legume nodules for nitrogen fixation<sup>283–285</sup> despite having *in vitro* antibacterial activity. For example, NCR247 from *Medicago truncatula* was characterized to have nanomolar affinity to haem and therefore iron sequestration activity in root nodules which triggered uptake of iron in nodule symbiont *Sinorhizobium meliloti*, a cofactor for its nitrogen fixation.<sup>286</sup> RALF peptides are involved in regulation of plant development and have several cellular effects, including MAP kinase activation and blockage of a proton pump, the latter implicating RALF as an alkalization factor in tobacco cells. As an alkalization factor, RALF has been observed to increase the pH that causes an observed stop in root growth.<sup>73</sup>

The endogenous bioactivity against phytopathogens and herbivores from CRP subclasses such as heveins, snakins, knottins, and defensins, has been successfully applied in transgenic plants to generate more disease- and pest-resistant plants.<sup>276,287–290</sup> In addition, CRPs such as knottins have potential as therapeutics, drug delivery agents, and diagnostics. For example, knottin–drug conjugates targeting integrins have shown potent inhibition of tumor cell proliferation<sup>291</sup> and knottins have inhibitory activity against cancer-relevant

proteases such as matrilysin-1.<sup>292</sup> Recently, the knottin excelsatoxin A from nettle *Dendrocnide excelsa* was described as a pain-causing agent. Excelsatoxin A targets a transmembrane protein, TMEM233, that modulates the voltage-gated sodium channel Na<sub>v</sub>1.7 involved in pain sensation.<sup>293,294</sup> Pain sensations of *Dendrocnide* plants have been described as long-lasting and intense; excelsatoxin A represents a major pain-causing agent from these plants with potential as an analgesic lead structure. Lastly, a mutant form of knottin can be labeled with <sup>18</sup>F and used in Positron Emission Topography (PET) imaging studies by binding to integrin positive tumors with high affinity.<sup>295</sup>

### 3.5. Linear plant RiPPs

In addition to cyclic RiPPs, multiple linear plant peptides with post-translational modifications are produced by plants. Many of these linear RiPPs have roles in plant development and growth and are often secreted. The most common post-translational modifications in linear plant RiPPs are known from protein biosynthesis:<sup>56</sup> tyrosine sulfation,<sup>296</sup> proline hydroxylation, and hydroxyproline arabinosylation (Fig. 2).<sup>297</sup> We direct readers to dedicated reviews for additional details<sup>296,298,299</sup> as we discuss them only briefly.

**3.5.1 Structure.** The discovery of the first linear plant RiPP was based on a desire to understand the molecular components responsible for cellular division in the conditioned growth media for plant cell culture. The Matsubayashi group conducted activity-guided fractionation from the condition media of *Asparagus officinalis* and identified a five amino acid peptide they named phytosulfokine- $\alpha$  (PSK) as a growth-promoting signaling molecule for *in vitro* asparagus cell culture.<sup>9</sup> Two of the tyrosines in the peptide were sulfated (Fig. 7) based on sequencing and mass spectrometry analysis. The authors validated PSK as a potent mitogen, stimulating cell growth at nanomolar concentrations, when compared to a synthetic standard. PSK was later isolated from *Oryza sativa*<sup>300</sup> with its single-core-containing precursor peptide identified shortly thereafter.<sup>301</sup> Other notable examples of sulfotyrosine-containing linear peptides include Casparian strip integrity factor 1 (CIF1)<sup>302,303</sup> and root meristem growth factors (RGFs).<sup>304</sup> Linear plant peptide hormones called systemins were characterized in tobacco and featured a hydroxyproline PTM and O-arabinosylation on the hydroxyproline hydroxyl group.<sup>11</sup>

**3.5.2 Biosynthesis.** The precursor peptides responsible for linear plant RiPPs are typically around 80–150 amino acids in length and commonly contain only a single core,<sup>301,305,306</sup> although multi-core examples do exist.<sup>11</sup> Elucidating the maturation process of these linear RiPPs is challenging as the observed modifications overlap with typical protein PTMs. However, some of the biosynthetic enzymes have been clearly delineated.

While the sulfotransferase reaction is well known to be catalyzed by tyrosylprotein sulfotransferases (TPST) in mice and humans, no enzyme with a high sequence similarity to TPST has been identified in plants.<sup>78</sup> Therefore, the Matsubayashi group used an enzyme activity-based fractionation approach with the precursor peptide for plant peptide containing sulfated



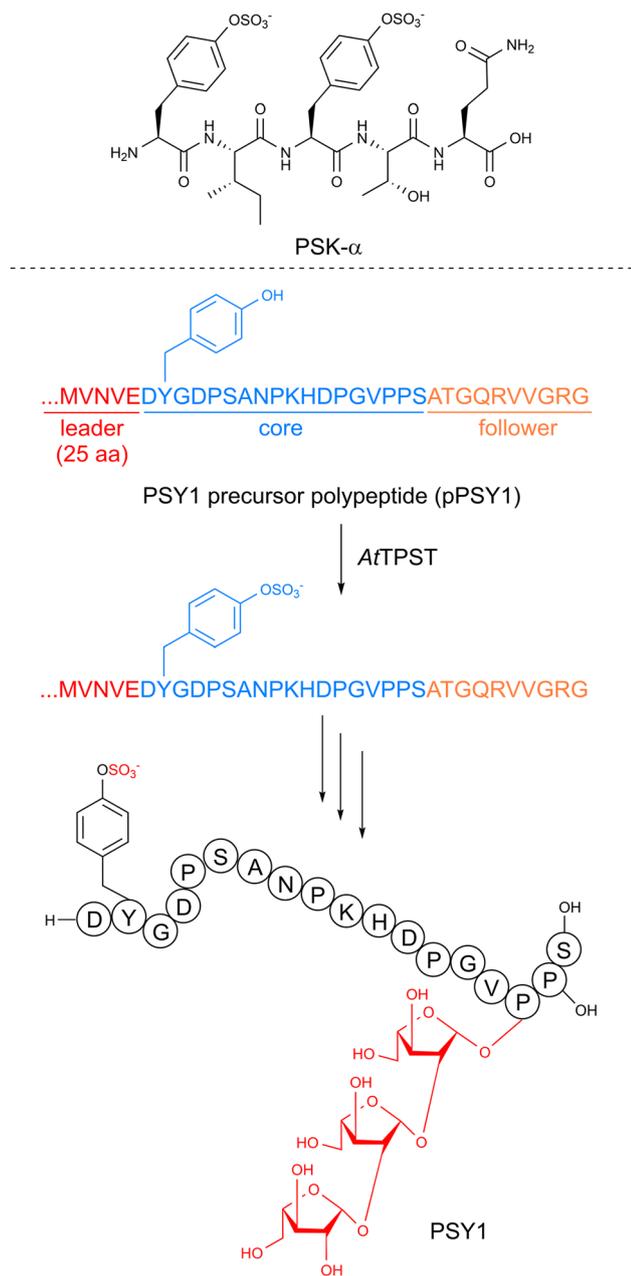


Fig. 7 Biosynthesis of linear RiPP PSY1.

tyrosines 1 (PSY1) (discussed below) as a substrate.<sup>78</sup> They identified a membrane protein they named *AtTPST* that catalyzed the anticipated tyrosine sulfation reaction using 3'-phosphoadenosine 5'-phosphosulfate (PAPS) as the sulfate source (Fig. 7). Subsequent heterologous expression and *in vitro* assays demonstrated that *AtTPST* could also utilize the PSK precursor peptide, indicating it may be responsible for the *in vivo* maturation of diverse linear plant RiPPs.

The formation of linear RiPP hydroxyproline and the subsequent addition of three arabinose residues (Fig. 7) has been the focus of recent studies.<sup>79</sup> The initial hydroxylation is predicted to be installed by a prolyl-4-hydroxylase. Many members of this enzyme family are found in plants but one

directly responsible for the hydroxylation in these plant RiPPs has not been demonstrated.<sup>307</sup> If present, the three arabinoses are likely added in a stepwise order by one enzyme that catalyzes the initial attachment to hydroxyproline and a second that sequentially adds two more.<sup>79</sup> A candidate for the first step has been identified, but the following arabinose additions have not been defined.

Steps involved in the proteolytic maturation of the core peptide have also been elucidated in several systems. CLE40 is a 12 or 13 amino acid peptide with a single hydroxyproline. Differential expression analysis and *in vitro* assays identified the three subtilisin-like proteases that were able to cleave C-terminal to the core sequence of the precursor peptide.<sup>308</sup> Notably, the presence of the hydroxyproline residue prevented an internal cleavage of the core sequence. Subtilisin-like proteases were also found to be involved in the maturation of PSK by cleaving three amino acids N-terminal and directly C-terminal to the core peptide. Similar proteases are implicated in production of several other linear plant peptides.

**3.5.3 Bioactivity.** In general, these linear plant RiPPs function as important signaling molecules or hormones involved in plant growth and cellular differentiation. PSK was discovered as a component responsible for the conditioned growth media needed for plant cell culture. Using a synthetic standard, the Matsubayashi group validated that PSK was a potent mitogen, simulating cell growth at nanomolar concentrations.<sup>9</sup> PSK was later isolated from *Oryza sativa* and a similar activity was observed.<sup>300</sup> A related approach was used to find PSY1, which also promotes cellular growth.<sup>305</sup> CLV3 suppresses shoot apical meristem cell development and instead upregulates cellular division.<sup>10</sup> CIF1 peptides are necessary for the formation of the Casparian strip in *Arabidopsis* roots, a structure important for ion homeostasis.<sup>302</sup> Other peptides are involved in responses to stress such as dehydration,<sup>309</sup> osmotic stress,<sup>310</sup> and nitrogen starvation.<sup>311</sup> Notably, the fundamental importance of these plant peptides has been targeted by pathogens. The sulfated tyrosine containing bacterial RiPP RaxX is thought to mimic PSY1 and serve as a virulence factor.<sup>312</sup>

### 3.6. BURP-domain-derived RiPPs (burpitides)

BURP-domain-containing proteins were characterized as precursor peptides for lyciumins and therefore for plant RiPP biosynthesis in 2018.<sup>17</sup> Subsequently, several additional RiPP classes with new macrocyclic PTMs were revealed as BURP-domain-derived peptides (Fig. 8).<sup>18–20</sup> Given the rapid expansion of new RiPP classes derived from the BURP domain,<sup>313</sup> we propose the name of burpitide for this family of peptide natural products biosynthesized using a BURP-domain-containing protein.

The BURP-domain-containing protein family was defined by Hattori and co-workers based on a conserved CHX<sub>10</sub>CHX<sub>25–27</sub>CHX<sub>25–26</sub>CH-motif<sup>313</sup> in four founding member plant proteins: microspore-derived embryo protein BNM2 from *Brassica napus*,<sup>314</sup> an unidentified seed protein (USP) from *Vicia faba*,<sup>315</sup> drought-responsive protein RD22 from *Arabidopsis thaliana*<sup>316</sup> and the  $\beta$ -subunit of polygalacturonase 1 (PG1 $\beta$ ) involved in



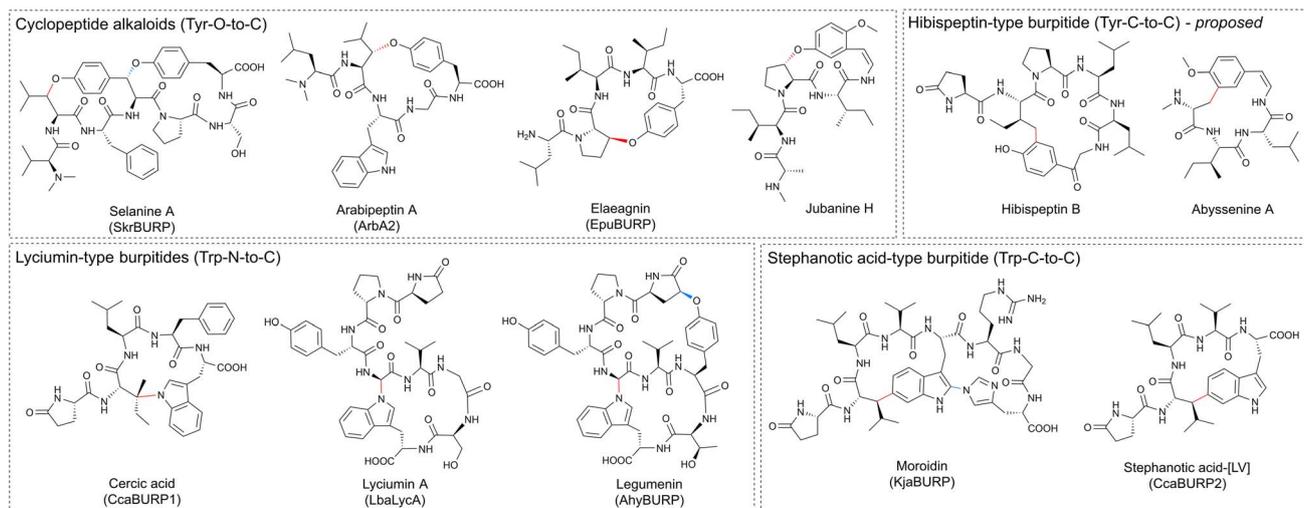


Fig. 8 Proposed burpitude classification. Representative precursor peptide names are in brackets. Class-defining bonds are highlighted in red, other burpitude cyclase-derived PTMs are highlighted in blue.

fruit ripening in *Solanum lycopersicum*.<sup>317</sup> Bioinformatic analyses of plant genomes show an abundance of BURP-domain genes with diverse primary structures and on average twelve BURP-domain genes per plant genome, ranging from one (*Marchantia polymorpha*) to 53 (*Coffea arabica*).<sup>318</sup> BURP-domain protein expression is mainly associated with abiotic stress responses in plants.<sup>319,320</sup> Examples are the expression of Sali3-2 in soybean roots during acidic soil stress.<sup>319</sup> Some BURP domains are potentially involved in biotic plant stress responses, for instance, a BURP-domain gene is up-regulated in a virus-resistant soybean<sup>321</sup> and a BURP-domain gene locus mapped to bruchid resistance in mung beans.<sup>322</sup>

In 2022, several BURP-domain-containing precursor peptides were reported as copper-dependent peptide cyclases, which catalyze macrocyclic bond formations in RiPP biosynthesis and define the autocatalytic or fused burpitude pathway.<sup>18</sup> In 2023, a BURP-domain peptide cyclase with a stand-alone precursor peptide was reported defining the split burpitude pathway.<sup>20</sup>

### 3.6.1 General recommendations for burpitude nomenclature

**3.6.1.1 Macrocyclic bond.** We propose that a macrocyclic bond derived from a BURP-domain-PTM is the primary class-defining feature of a burpitude. Since BURP domains can crosslink aromatic amino acid side chains with many different amino acids *via* diverse chemical bonds, we recommend a burpitude classification based on bond chemistry (C–O, C–N, C–C) and on crosslinked aromatic residues for a given BURP-PTM. For example, cyclopeptide alkaloids would be defined as a burpitude class with C(sp<sup>3</sup>)-O-phenol-ether-crosslinks with the C(sp<sup>3</sup>)-macrocyclization site being any amino acid and the phenol being a tyrosine or similar residue. This classification (Table 2) is inspired by fungal dikaritins RiPPs that are defined by an ether-crosslink between a tyrosine to any amino acid side chain.<sup>2,43,46</sup>

**3.6.1.2 Ring size.** Ring size should not be a class-defining feature of a burpitude class. Traditional classification defines cyclopeptide alkaloids as 13-, 14- or 15-membered peptides.

However, recent discoveries of 17-membered arabipeptins<sup>20</sup> and elaeagnins<sup>21</sup> with cyclopeptide alkaloid-characteristic crosslinks indicate that larger versions of these macrocycles exist in

Table 2 Recommendations for burpitude definition and nomenclature. See main text for rationale and examples

#### Burpitude definition

The macrocyclic bond derived from a BURP-domain-PTM is the primary class-defining feature of a burpitude

In multicyclic burpitudes, the macrocyclic bond, that is formed first during biosynthesis is the class-defining feature

Ring size should not be a class-defining feature of a burpitude

Terminal modifications should not be class-defining features if a macrocyclic BURP-PTM is present in a burpitude

#### Burpitude classification

Class	Class-defining PTM
Lyciumin-type peptides	Crosslink of Trp-indole-N to carbon in another amino acid side chain or peptide backbone
Cyclopeptide alkaloids	Crosslink of Tyr-phenol-O to carbon in another amino acid side chain
Stephanotic acid-type peptides	Crosslink of Trp-indole-C to carbon in another amino acid side chain
Hibispeptin-type peptides (proposed)	Crosslink of Tyr-phenol-C to carbon in another amino acid side chain

#### Burpitude nomenclature

BURP-domain-containing proteins involved in a crosslinking step of a burpitude biosynthesis should be called burpitude cyclases. Fused burpitude cyclases should be named with three letters referring to genus (1 letter) and species (2 letters) and 'BURP' (e.g. SkrBURP for *Selaginella kraussiana* burpitude cyclase).

In split pathways, the burpitude cyclase should be designated with a B (e.g. ArbB2) and the precursor peptide should be designated with an A (e.g. ArbA2).



nature. Similarly, ring size is not applied for any RiPP class with a macrocyclic bond as a class-defining feature.<sup>1,2</sup>

**3.6.1.3 Ring number and mixed modifications.** Several burpitides have two macrocyclic bonds.<sup>18,19</sup> Some identified bicyclic burpitides to date have two different types of macrocyclic bonds, which poses the question as to which PTM should be considered class-defining. In this case, we recommend assigning the macrocyclic bond that is formed first during biosynthesis as the class-defining feature. For example, in legumenin biosynthesis the C-terminal lyciumin-bond is formed first,<sup>18</sup> which defines legumenin as a lyciumin-type burpitide. In moroidin biosynthesis, the N-terminal Leu-Trp-crosslink of the stephanotic acid ring is formed first.<sup>19</sup> Following the logic of legumenin as lyciumin-type, moroidin is classified as stephanotic-acid type.

**3.6.1.4 Terminal modification.** We recommend that terminal modifications should not be class-defining features if a macrocyclic BURP-PTM is present in a burpitide. Common N-terminal modifications of burpitides are N-methylation (cyclopeptide alkaloids) and pyroglutamate formation. They are present in multiple burpitides with different macrocyclic bonds (stephanotic acid-type, lyciumin-type, and hisispeptins).

**3.6.1.5 Biosynthesis.** We recommend that BURP domains involved in a crosslinking biosynthetic step should be called burpitide cyclases to differentiate them from BURP domains for which no function is known. We also recommend that a split and fused burpitide nomenclature be used to define the biosynthetic route. In the split burpitide pathway, the precursor peptide and burpitide cyclase are encoded for by separate genes,<sup>20</sup> whereas in the fused burpitide pathway the core peptides are encoded within the same gene as the burpitide cyclase.<sup>18</sup> To create consistency in split pathways, the burpitide cyclase should be designated with a B (e.g. ArbB2) and the precursor peptide should be designated with an A (e.g. ArbA2). For fused burpitide cyclases, we recommend naming the enzyme with three letters referring to genus (1 letter) and species (2 letters) and 'BURP' (e.g. SkrBURP for *Selaginella kraussiana* burpitide cyclase).

**3.6.2 Lyciumin-type peptides (Trp-indole-N-to-C).** Lyciumins are originally described as monocyclic octapeptides isolated from *Lycium* plants.<sup>17</sup> Lyciumin A and B were isolated by a bioactivity-guided approach from root extract of gojiberry (*Lycium barbarum*), which is used in Chinese herbal medicine to treat hypertension.<sup>323</sup> Since this initial study, an additional 19 lyciumins have been reported including 17 lyciumins from gene-guided discovery approaches.<sup>17,21</sup> Lyciumins have been detected from Fabaceae, Solanaceae, Berberidaceae, and Amaranthaceae. We recommend the macrocyclic bond definition below of lyciumins that includes cercic acid and legumenin into this burpitide class.

**3.6.2.1 Macrocyclic bond.** Lyciumin-type burpitides are defined by a C(sp<sup>3</sup>)-N-macrocyclic bond between a tryptophan-indole-N and the C(sp<sup>3</sup>)-carbon of another amino acid (Fig. 8).

**3.6.2.2 Structure.** Lyciumin-type burpitides can be side-chain-to-backbone-macrocyclic or side-chain-to-side-chain-macrocyclic peptides with a class-defining C(sp<sup>3</sup>)-N-macrocyclic bond between the indole-nitrogen of a C-terminal

tryptophan to an unactivated carbon of another amino acid (Fig. 8). Known side-chain-to-backbone-macrocyclic lyciumins include founding members lyciumin A and B, which are monocyclic octapeptides and have a C(sp<sup>3</sup>)-N-crosslink between the tryptophan-indole-nitrogen to the C $\alpha$  of a glycine at the fourth position of the core peptide. The stereochemistry of the macrocyclic bond in lyciumin A has been characterized as *R* by DFT calculations and comparison to experimental data.<sup>324</sup> Cercic acid is a side-chain-to-side-chain-macrocyclic pentapeptide with a C(sp<sup>3</sup>)-N-crosslink between the tryptophan-indole-nitrogen to the C $\beta$  of an isoleucine at the second position. The stereochemistry of the Ile-C $\beta$  has been determined as *R* based on stereoisomer DFT-calculations and NOE NMR analysis of the purified natural product.<sup>18</sup> The N-terminus of lyciumins is a pyroglutamate, whereas the C-terminus of most lyciumins is unmodified. Exceptions are lyciumins A and C methylates isolated from *Celosia argentea* with a C-terminal methyl ester group at the C-terminal tryptophan carboxyl group.<sup>325</sup>

A bicyclic lyciumin-type burpitide called legumenin was isolated from alfalfa seeds (*Medicago sativa*) based on gene-guided discovery approach. Legumenin is an octapeptide derived from the core peptide sequence QPYGVYTW that has a lyciumin-bond between a C-terminal tryptophan and a glycine at the fourth position and an ether bond between the sixth position Tyr-phenol-OH and the first position PyroGlu-C $\gamma$ . The stereochemistry of the PyroGlu-C $\gamma$  was determined as *S* based on stereoisomer DFT-calculations and comparison to NOE NMR analytical data of the isolated peptide.<sup>18</sup>

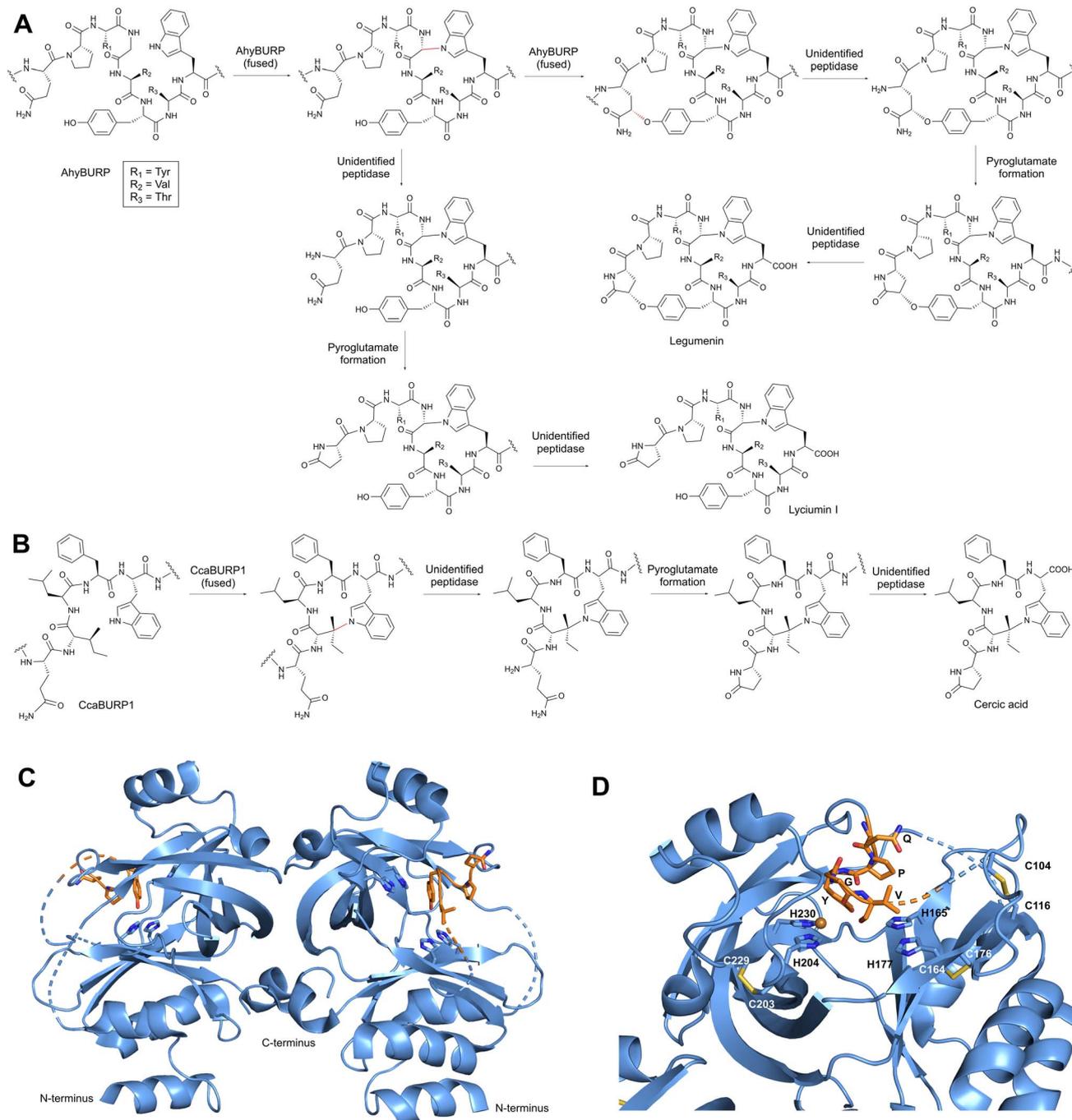
**3.6.2.3 Biosynthesis.** Lyciumins have been characterized as RiPPs by the discovery of precursor peptide gene *LbaLycA* from a root transcriptome of *Lycium barbarum*.<sup>17</sup> *LbaLycA* has twelve core peptides encoding for three lyciumins (A, B, D) in a highly repetitive N-terminal domain and a C-terminal BURP domain. The heterologous expression of *LbaLycA* in *Nicotiana benthamiana* via *Agrobacterium tumefaciens* infiltration and the pEAQ-HT expression system resulted in the formation of the lyciumins corresponding to *LbaLycA*-core peptides in transgenic tobacco leaves after six days. In addition, a BURP-domain-containing protein with an internal lyciumin core peptide called Sali3-2 was identified by bioinformatic searches for BURP-domain lyciumin precursor peptides. Sali3-2 is a homolog of the founding member USP from soybean, comprised of the only one lyciumin I core peptide QPYGVYTW, and has been implicated in increased abiotic tolerance in plants grown in metal-contaminated soil.<sup>326</sup> Lyciumin I was detected in the root and seed pods of soybean plants and in the leaf extracts of transgenic tobacco expressing *Sali3-2*.<sup>17</sup> In addition, bicyclic legumenin was detected in transient gene expression experiments of Sali3-2 homolog AhyBURP with the same core peptide.<sup>18</sup> Transient tobacco expression confirmed a lyciumin precursor peptide from the potato, StuBURP, which is connected to production of lyciumin J (core: QPYGVFAW).<sup>17</sup> Lastly, transient expression of the cercic acid precursor CcaBURP1 in tobacco resulted in the biosynthesis of cercic acid.

A Sali3-2 homolog from peanut called AhyBURP with the same legumenin core peptide was further characterized as a copper-dependent peptide cyclase by bottom-up proteomic



analysis. Herein, a mass loss matching a leguminin macrocyclization in the core peptide region was detected. A lyciumin-type bond formation was detected in the modified bicyclic core peptide after exopeptidase incubation of the purified core peptide and comparison to leguminin. AhyBURP *in vitro* reconstitution further established that the lyciumin bond is

formed before the PyroGlu1-Tyr6-crosslink is formed (Fig. 9A).<sup>18</sup> A direct characterization of a lyciumin PTM in a modified core peptide was done in the cercic acid pathway from Eastern redbud (*Cercis canadensis*). A truncated single core peptide construct of cercic acid BURP-domain precursor CcaBURP1 was reconstituted *in vitro* with copper, treated with exopeptidases,



**Fig. 9** Biosynthetic proposal of lyciumin-type peptides. (A) Proposed formation of lyciumin I and leguminin from AhyBURP (B) Proposed formation of cercic acid from CcaBURP1. (C) The homodimer of AhyBURP, PDB ID 8SY2. Each subunit is in blue, and the core peptides are in orange and colored according to element (oxygen, red; nitrogen, blue). The core peptide and conserved His of BURP-domain-containing proteins are shown as sticks. Main chain atoms are omitted for clarity. (D) Copper-bound structure of AhyBURP, PDB ID 8SY3. The core peptide and conserved Cys and His residues of BURP-domain proteins are shown as sticks (sulfur, yellow). Main chain atoms for Cys-His are omitted for clarity. Dashed lines in (C) and (D) represent disordered regions in the crystal structure.



and the resulting modified QILFW core peptide yielded an analyte matching a cercic acid standard<sup>18</sup> (Fig. 9B).

The AhyBURP protein structure was recently determined by X-ray crystallography.<sup>327</sup> AhyBURP is a homodimer in solution and in the crystal structure (Fig. 9C). The BURP-domain fold represented in the AhyBURP subunit is unique among experimentally determined protein folds. In each AhyBURP subunit, a cleft is found, composed of a  $\beta$ -barrel and a  $\beta$ -sheet that face each other, overlaid by the core peptide substrate. The four conserved Cys–His residues of BURP-domain proteins are in this cleft, which forms the active site of two Type II copper centers.<sup>328</sup> Two conserved Cys residues form a disulfide staple and the sequential, conserved His residues extend into the cleft from the  $\beta$ -sheet, and the same arrangement is found on the facing  $\beta$ -barrel (Fig. 9D). The active sites of the dimer are separated by about 50 Å, and each core peptide is positioned to react in *cis* to its active site subunit. This is contrary to fungal autocatalytic peptide-*N*-methyltransferases, where the substrate core peptides react in *trans* across a dimer interface.<sup>329</sup> In addition to the structural characterization of AhyBURP, functional studies were reported.<sup>327</sup> AhyBURP can use Cu(I) or Cu(II) to facilitate peptide macrocyclization, it requires dioxygen for activity and isotopic labeling with deuterated Gln in the core peptide demonstrated a loss of 1 D and 3 H in the presence of Cu(II), suggesting a radical-based reaction. Isotopic labeling studies further supported an intramolecular reaction within the AhyBURP subunit and its core peptide. Lastly, tandem mass spectrometry identified two covalent radical trap adducts within the core peptide region, indicating that two radicals can be detected within the core peptide in the presence of Cu(II). Many questions emerge from this work, including which steps in the mechanism use single electron transfer, how and when dioxygen is bound, and the copper coordination state throughout the reaction.

The N-terminal pyroglutamate is generated from glutamine, which can spontaneously cyclize to pyroglutamate after proteolytic cleavage at the N-terminus of a lyciumin core peptide. The involvement of glutamine cyclotransferase has been proposed based on co-localization of glutamine cyclotransferase genes with fused lyciumin precursor peptides in the beet genome.<sup>17</sup> To date, all characterized lyciumins are derived through the fused burpitude pathway.

**3.6.2.4 Bioactivity.** Lyciumin A and B showed moderate inhibition of renin and angiotensin-converting enzyme.<sup>323</sup> No endogenous functions of lyciumins have been described yet. They are detected in diverse plant tissues and developmental stages, including in storage tissues such as seeds (amaranth, soybean), roots (gojiberry), and sprouts (potato). The functions of legumenin and cercic acid are also unknown.

**3.6.3 Cyclopeptide alkaloids (Tyr-phenol-O-to-C).** Cyclopeptide alkaloids (CPA)<sup>330</sup> are four to six amino acids in length and represent a large plant peptide class with >230 known compounds.<sup>3,18,331</sup> The high number of cyclopeptide alkaloids is based on their diversity of macrocyclic bonds, and N- and C-terminal functionalities. Traditionally, cyclopeptide alkaloids have been defined by macrocyclization *via* a phenol group of a C-terminal hydroxystyrylamine, octopamine, or tyrosine

residue to an unactivated carbon at an amino acid side chain two positions N-terminally from the phenol-bearing residue. In the traditional classification system, the phenol-crosslinks are usually ethers involving the phenolic hydroxy group with rare exceptions having C(sp<sup>2</sup>)-C(sp<sup>3</sup>)-crosslinks between the meta-carbon of a phenol to a different amino acid side chain. Furthermore, the previous classification limited cyclopeptide alkaloids to 13-, 14- and 15-membered macrocycles.<sup>3,331,332</sup> As burpitudes emerge within the biosynthetic description of cyclopeptide alkaloids,<sup>18,20</sup> and their structural similarity to the fungal RiPP class dikaritins,<sup>2,43–46</sup> we propose a revised classification of cyclopeptide alkaloids.

**3.6.3.1 Macrocyclic bond.** Cyclopeptide alkaloid-type burpitudes are defined by phenol-ether-crosslinks formed by burpitude cyclases (Fig. 1B and 8).<sup>333,334</sup> The phenol group-crosslinks in cyclopeptide alkaloid-type burpitudes are derived from tyrosine residues, however the biosynthesis of cyclopeptide alkaloids containing hydroxystyrylamine and octopamine moieties is largely uncharacterized.<sup>331</sup> Several cyclic plant peptides previously defined as cyclopeptide alkaloids contain C(sp<sup>3</sup>)-C(sp<sup>2</sup>)-crosslinks involving a phenolic carbon at the meta-position, for example, abyssenine A<sup>335</sup> (Fig. 8). Due to the different chemistry of their macrocyclic bond, we suggest excluding these non-ether-crosslinked peptides from the nomenclature of cyclopeptide alkaloids and define them as a separate peptide class (Section 3.6.5).

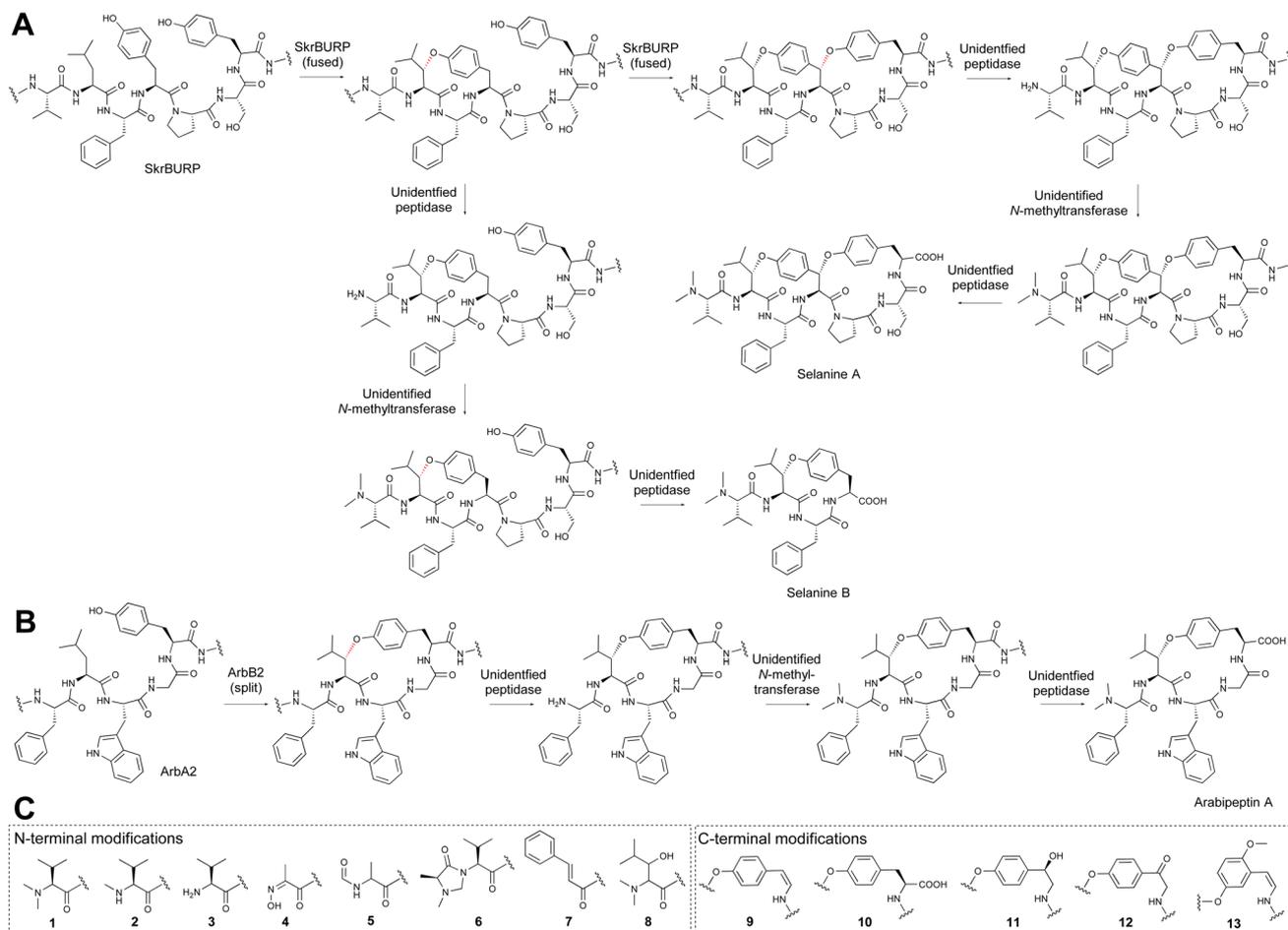
**3.6.3.2 Ring size.** Cyclopeptide alkaloid rings range from 13–17 atoms, but other sizes may exist.

**3.6.3.3 Ring number.** Cyclopeptide alkaloids can have more than one macrocycle. In peptides with different macrocyclic bonds, the first macrocyclic bond formed defines the peptide class. For example, legumenin is not considered as a cyclopeptide alkaloid but a lyciumin-type peptide due to the first formed lyciumin-bond despite having a Tyr-O-PyroGlu-C $\gamma$ -crosslink (see Section 3.6.2). Since many cyclopeptide alkaloids await biosynthetic definition, this nomenclature is subject to refinement.

**3.6.3.4 Structure.** As recommended above, cyclopeptide alkaloids are defined as side-chain-to-side-chain-macrocyclic peptides with a phenol-derived ether bond. The ether bond can involve a C-terminal *meta*- or *para*-phenol hydroxy-group such as in jubanine H or sanjoinine A, respectively (Fig. 10). The C-terminal phenol-containing residue can be a tyrosine in vignatic acid A, a *p*-hydroxystyrylamine in sanjoinine A, a *m*-hydroxystyrylamine in jubanine H (Fig. 8), or an octopamine in pandamine.<sup>336</sup> The residues *p*-hydroxystyrylamine or *m*-hydroxystyrylamine can have additional *o*-hydroxy or *o*-methoxy groups, as observed in ramosine C<sup>337</sup> or in jubanine H,<sup>333</sup> respectively. The cyclopeptide alkaloid encephanine has an O-glycosylation at the *m*-position of a *p*-hydroxystyrylamine and represents the only glycosylated plant cyclopeptide to date.<sup>18</sup>

The double bond configuration of cyclopeptide alkaloid styrylamine is usually *Z*. The macrocyclization site in the other residue of the ether bridge is usually the  $\beta$ -carbon of a leucine, proline, phenylalanine, tyrosine, isoleucine, or valine.<sup>3</sup> 94.1% (64/68) of reported cyclopeptide alkaloids structures have an *S*-configuration at this cyclization site.<sup>3</sup> The N-terminus of





**Fig. 10** Biosynthetic proposal for cyclopeptide alkaloids. (A) Proposed formation of mono- and bicyclic peptide alkaloids from SkrBURP in a fused burpitide pathway. (B) The proposed split burpitide pathway for the production of arabipeptin A. (C) N-terminal modifications observed in cyclopeptide alkaloids: (1) *N,N*-dimethylation, (2) *N*-monomethylation, (3) unmodified, (4) *N*-oxime, (5) *N*-formylation, (6) cyclized *N*-methylation (imidizolidine-4-one), (7) deamination (cinnamic acid), (8) C $\beta$ -hydroxylation, C-terminal modifications observed in cyclopeptide alkaloids: (9) *p*-hydroxy-styrylamine, (10) tyrosine, (11) octopamine, (12) 4-hydroxy- $\alpha$ -aminoacetophenone (13) *m*-hydroxy-*o*-methoxy-styrylamine.<sup>3</sup>

cyclopeptide alkaloids is usually *N*-mono- or *N,N*-dimethylated. Additional N-terminal functional groups include cinnamic acid, deaminated leucine,<sup>338</sup> 2-(hydroxyimino)propanoic acid, *N*-oximes, *N*-formyl groups and cyclized *N*-methylations such as imidizolidine-4-one (Fig. 10C).<sup>3,339</sup> Almost all known cyclopeptide alkaloids are monocyclic, however bicyclic cyclopeptide alkaloids have been determined. Selanine A from African clubmoss has a di-ether-bridge *via* a Leu-Tyr-Tyr-crosslink.<sup>18</sup> Selanine A is similar in overall macrocycle structure to fungal dikaritin RiPP asperipin-2a except for an *S*-configuration at both C $\beta$ -cyclization sites in selanine A.<sup>340</sup>

**3.6.3.5 Biosynthesis.** Recently, a 14-membered cyclopeptide alkaloid with a Tyr-O-Leu-C $\beta$ -crosslink called selanine A was characterized as a burpitide derived from a fused burpitide pathway in African clubmoss (*Selaginella kraussiana*)<sup>18</sup> and a 17-membered cyclopeptide alkaloid with a Tyr-O-Leu-C $\beta$ -crosslink named arabipeptin A was discovered as a burpitide derived from a split burpitide pathway in *Coffea arabica*.<sup>20</sup> Furthermore, 17-membered burpitides with Tyr-O-Pro-crosslinks have been proposed and identified as fused burpitide pathway products in soybean and silverberry (*Elaeagnus pungens*), respectively.<sup>18,21</sup>

The fused cyclopeptide alkaloid pathway from African clubmoss starts with BURP-domain precursor peptide SkrBURP, which has four core peptide sequence repeats of VLFYPSY in the N-terminal domain. Transient expression of native SkrBURP and a truncated SkrBURP with only one core peptide (SkrBURP-1xVLFYPSY) in tobacco led to the production of desmethylselanine A and desmethylselanine B, which matched the purified selanine A and B after reductive *N,N*-dimethylation. This experiment showed that the clubmoss cyclopeptide alkaloids are burpitides. Furthermore, the formation of the cyclopeptide alkaloid bonds between the Tyr-O-Leu-C $\beta$  in selanine A and between the Tyr-O-Tyr-C $\beta$  was reconstituted *in vitro* from purified SkrBURP-1xVLFYPSY in the presence of copper and characterized by bottom-up proteomics of the core peptide, exopeptidase reaction, and comparison to desmethylselanine A and B analytes from SkrBURP tobacco expression experiment.<sup>18</sup> This result showed that cyclopeptide alkaloid ether-bonds can be formed by fused burpitide cyclases. In an automated genome mining study, mono- and bicyclic cyclopeptide alkaloids analytes matching core peptides ILLYPSY and FLLYPY in SkrBURP could be detected in extracts of *S. kraussiana* and as



corresponding analogs in transient tobacco extracts of SkrBURP expression.<sup>21</sup> These observations indicated that burpitide cyclases such as SkrBURP enable the formation of both bi- and monocyclic cyclopeptide alkaloid diversification from a single fused BURP-domain gene (Fig. 10A).

Some of the most prolific producers of cyclopeptide alkaloids are in the buckthorn family (Rhamnaceae) with prominent source plants being New Jersey Tea (*Ceanothus americanus*) and jujube (*Ziziphus jujuba*). A recent study<sup>20</sup> identified repetitive proteins encoded in transcripts of *C. americanus* that had short sequence motifs matching known cyclopeptide alkaloid structures from *C. americanus*. Homologs of these putative cyclopeptide alkaloids were further identified in the genome of *Z. jujuba* and matched jujube cyclopeptide alkaloid structures. Notably, they were co-localized with stand-alone BURP-domain genes in the jujube genome. This strongly suggested the presence of a split burpitide pathway wherein the precursor peptide and burpitide cyclase are translated as two separate polypeptides instead of a fused one. Multiple copies of similar split cyclopeptide alkaloid precursor peptides were found in the *Coffea arabica* genome. One of these hypothetical precursor peptides contained three copies of the core peptide FLWGY flanked by recognition sequences and was located adjacent to a putative burpitide cyclase gene. This observation guided the isolation of a new cyclopeptide alkaloid arabipeptin A. *In vitro* reconstitution of the corresponding burpitide cyclase ArbB2 with the split precursor peptide ArBA2 resulted in the detection of a mass shift matching the Tyr-O-Leu-C $\beta$  macrocyclization observed in arabipeptin A and thus showed that cyclopeptide alkaloids can be biosynthesized through a split burpitide pathway (Fig. 10B).

Based on the reconstitution of multiple pathways, precursor peptides for cyclopeptide alkaloids typically contain multiple core sequences and are cyclized by both split and fused burpitide cyclases. However, the downstream biosynthetic enzymes involved in the N- and C-terminal modification of cyclopeptide alkaloids are mainly unknown. A cyclopeptide alkaloid-guided genome-wide-association-study in *Z. jujuba* identified a putative N-methyltransferase for the N-terminal dimethylation of sanjoinine A co-localized with a BURP-domain gene and belongs to an unknown plant SAM-dependent methyltransferase clade.<sup>341</sup> Methylation activity in a cyclopeptide alkaloid analyte was proposed based on transient gene expression of the methyltransferase in white mature jujube fruits and detection of sanjoinine A formation compared to an empty vector control. The enzymatic basis for the formation of the C-terminal hydroxystyrylamine observed in many cyclopeptide alkaloids remains to be elucidated.

**3.6.3.6 Bioactivity.** Several cyclopeptide alkaloids have been studied for their anxiolytic and analgesic effects. Sanjoinine A is a bioactive ingredient of jujube used in herbal medicine to treat insomnia and further displays GABA-adrenergic activity to reduce insomnia.<sup>342</sup> Studies in mouse models demonstrated that sanjoinine A also has anti-anxiety activity, likely mediated again through the GABA<sub>A</sub> receptor.<sup>343</sup> Studies with adouetine X<sup>344</sup> and multiple cyclopeptide alkaloids from *Ziziphus oxyphylla*<sup>345</sup> showed analgesic activity in multiple mouse pain

models. Subsequent *in vitro* assays with adouetine X suggest that analgesic activity may be due to inhibition of Ca<sup>2+</sup>-ATPase and Na<sup>+</sup>/K<sup>+</sup> ATPase. A similar inhibition was observed with sanjoinine A and sanjoinine F.<sup>346</sup>

Another jujube cyclopeptide alkaloid, jubanine H, was recently reported to have antiviral activity against a pig coronavirus and low mammalian cell cytotoxicity by *in vitro* viral infection assays.<sup>333</sup> Numerous cyclopeptide alkaloids have been investigated for antiplasmodial activity and demonstrated single digit  $\mu$ M IC<sub>50</sub> values, several such as nummularine R, spinanine B, and adouetine X.<sup>347,348</sup> The cyclopeptide alkaloid vignatic acid A from mung bean has insecticidal activity against the azuki bean weevil (*Callosobruchus chinensis*)<sup>349</sup> and it has been hypothesized to act as a resistance-conferring product of a mung bean beetle resistance gene. One of the hypothesized resistance genes is a SkrBURP homolog called resistant-specific protein 1(4) with core peptides matching vignatic acid A.<sup>349</sup> It was shown that the presence of resistant-specific protein 1(4) does not fully confer resistance to *C. chinensis* suggesting that other genes might be involved in the resistance mechanism.<sup>350</sup>

**3.6.4 Stephanotic acid-type peptides (Trp-C-to-C).** Moroidin is a bicyclic peptide which was originally isolated from Australian stinging nettle *Dendrocnide moroides* and recently was characterized as a burpitide.<sup>19</sup> The N-terminal macrocycle of moroidin is also found in the burpitide stephanotic acid, which was first isolated from *Stephanotis floribunda*<sup>351</sup> and then biosynthetically characterized as a burpitide from *Cercis canadensis*.<sup>18</sup> Given a structural overlap of moroidin and stephanotic acid, we include them as stephanotic acid-type burpitides and recommend the following nomenclature below for this RiPP class.

**3.6.4.1 Macrocyclic bond.** Stephanotic acid-type peptides are defined by a C(sp<sup>3</sup>)-C(sp<sup>2</sup>)-crosslinks between a carbon of a tryptophan-indole and a carbon of another amino acid.

**3.6.4.2 Structure.** Moroidins and stephanotic acids are side-chain-to-side-chain-macrocyclic peptides. Moroidin has a bicyclic core structure with a class-defining crosslink between the Leu2-C $\beta$  and the Trp5-indole-C6 (stephanotic acid crosslink) and an additional crosslink between the Trp5-indole-C2 and a His8-imidazole-N1. The Leu-C $\beta$  has *S*-stereochemistry as determined by X-ray crystallography.<sup>352,353</sup> The Trp-His-bond is the only example to date of two crosslinked aromatic rings in a burpitide. Moroidin and ten moroidin analogs have been reported from *Celosia argentea*, named celogentins A–K.<sup>353–355</sup> Several of the celogentins have extended linear C-termini. Another moroidin analog called moroidin-[QLLVWRSH] was also discovered from *Bauhinia tomentosa*.<sup>19</sup> Three stephanotic acids have been characterized so far with core peptides QLIVW, QLLVW and QLKVW. Both moroidins and stephanotic acids share a L-pyroglutamate at the N-terminus.

**3.6.4.3 Biosynthesis.** Moroidin was characterized as a burpitide derived from a fused burpitide pathway from *Kerria japonica* by *in planta* and *in vitro* reconstitution of the moroidin precursor peptide KjaBURP (Fig. 11A).<sup>19</sup> KjaBURP has four repeats in the N-terminal domain, including three moroidin core peptides (QLLVWRGH) and one moroidin-[QLLVWRAH] core peptide. Transient expression of *KjaBURP* in tobacco



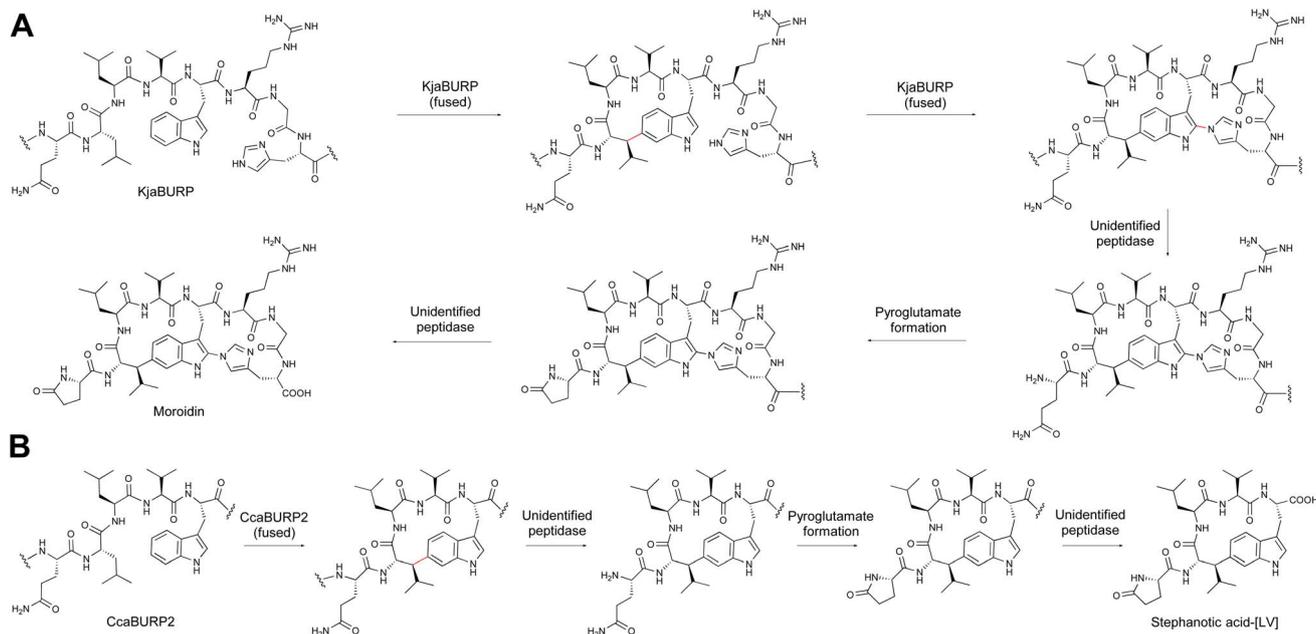


Fig. 11 Biosynthetic proposal for stephanotic acid-type burpitides. (A) Proposed biosynthesis of moroidin in *Kerria japonica*. (B) Proposed biosynthesis of stephanotic acid-[LV] in *Cercis canadensis*.

yielded moroidin and moroidin-[QLLVWRAH] analytes and *in vitro* reconstitution of KjaBURP-[1xQLLVWRGH] with copper and subsequent endo- and exoproteolytic digests also resulted in the formation of a moroidin analyte. *In planta* reconstitution showed that the Leu-Trp-crosslink forms first and the Trp-His-crosslink second. *In planta* and *in vitro* analysis resulted in the detection of stephanotic acid-[LV] and no detection of an analyte corresponding to the second ring. The pyroglutamate in moroidin can form spontaneously from glutamine after N-terminal cleavage of the Gln1. A similar biosynthetic pathway has been characterized for stephanotic acid-[LV] from *Cercis canadensis* burpitide cyclase CcaBURP1 (Fig. 11B).<sup>18</sup>

**3.6.4.4 Bioactivity.** Moroidin was discovered during a study for pain-causing agents from *D. moroides*. While pain-causing activity was reported for purified moroidin,<sup>356</sup> no pharmacological studies exist about the pain-causing activity of moroidins. Moroidin and celogentin C have *in vitro* anti-tubulin polymerization activity and *in vitro* lung adenocarcinoma cell cytotoxicity with reported low micromolar  $IC_{50}$ .<sup>19,352</sup> Stephanotic acid-[LV] has no significant *in vitro* cytotoxicity against a tested lung cancer cell line indicating that the second Trp-His ring is important for cancer cell cytotoxicity.<sup>18</sup> Endogenous functions of moroidins and stephanotic acids in source plants are not established.

**3.6.5 Hibispeptin-type peptides (Tyr-C-to-C).** The type defining molecules, hisispeptins A and B, are produced by the well-known shrub *Hibiscus syriacus* (Fig. 8).<sup>357,358</sup> These molecules are cyclized through a carbon-carbon crosslink between the C-terminal 2-amino-3-(2-hydroxy-5-aminoacetylbenzyl) pentanoic acid (Ahabpa) and the  $\gamma$ -methyl of L-isoleucine. To date, no enzymatic or heterologous reconstitution has been completed for hisispeptins to prove that they are burpitides.

However, precursor peptides that mimic the split precursor peptides of cyclopeptide alkaloids and include cores matching known hisispeptins along with the genomic co-localization of a BURP-domain-containing gene strongly suggests their biosynthetic classification as burpitides.<sup>20</sup>

**3.6.5.1 Macrocyclic bond.** Hisispeptin-type burpitides are characterized by a  $C(sp^3)$ - $C(sp^2)$ -macrocyclic bond between a C-terminal phenolic amino acid, such as tyrosine or a tyrosine-derived residue, and the  $C(sp^3)$ -carbon of another amino acid. This definition also encompasses a few molecules that were originally classified as cyclopeptide alkaloids, such as abyssenine A (Fig. 8).<sup>335</sup>

**3.6.5.2 Structure.** The original hisispeptin A and B peptides are six amino acids in length and the hypothesized core sequences are largely conserved by amino acid sequence as QIPLFY and QIPLLY, respectively.<sup>357,358</sup> In each case, the N-terminal glutamine has been cyclized to pyroglutamate. In both hisispeptin A and B the cyclization occurs between the  $\gamma$ -methyl of L-isoleucine and the *meta* position of the tyrosine derived Ahabpa residue. As stated, we propose to expand the hisispeptin-type burpitides to include all burpitides with a  $C(sp^3)$ - $C(sp^2)$ -macrocyclic bond from a C-terminal phenyl ring-containing amino acid. Natural products such as abyssenine A-C (Fig. 8) and mucronine A-H fall into this classification.<sup>335</sup> They are each five amino acids in length with a terminal *p*-methoxy-styrylamine residue crosslinked between the *meta*-position to the beta-position of the first amino acid. The N-terminus is often methylated and an additional *o*-methoxy modification is often observed on the *p*-methoxy-styrylamine as in mucronine E.<sup>3</sup>

**3.6.5.3 Biosynthesis.** Recent work on *H. syriacus* has identified candidate split precursor peptides for both hisispeptin A and



B.<sup>20</sup> These potential precursor peptide genes for hibispeptin A and B were located adjacent to putative split burpitide cyclases, strongly implying that they are biosynthesized as part of a split burpitide pathway. The adjacent genes for hibispeptin A and B resemble the split burpitide biosynthetic pathway observed in arabipeptin A. Additional precursor peptides were identified containing core peptides that did not match either hibispeptin A or B, such as QVPLVY. Metabolomic analysis of *H. syriacus* root extract indicated a presence of a mass spectrometric feature and MS/MS fragmentation data supporting this new hibispeptin analogue.<sup>20</sup>

**3.6.5.4 Bioactivity.** To date, hibispeptin-type burpitides have minimal noted bioactivity. Hibispeptin A demonstrated weak inhibition of lipid peroxidation<sup>358</sup> and abyssenine A and mucronine E showed weak cytotoxic activity.<sup>359</sup> Antimicrobial tests demonstrated that mucronins E, G, and H and abyssenines A and C have weak antibacterial activity, whereas abyssenine A has some antifungal activity.<sup>335</sup>

## 4. Undefined plant-derived peptides

While the molecules discussed in Section 3 have been established as plant produced RiPP natural products, numerous other peptides have been isolated from plants for which little biosynthetic information is available. Moreover, the authentic producer may also be unclear. In many cases, plants harbor endophytic bacteria or fungi that are associated with a particular species or genus. These microorganisms have been shown to be responsible for the production of many peptidic natural products, further complicating biosynthetic elucidation of a target molecule.<sup>31,360</sup> For example, the astin class of peptides was shown to be biosynthesized by an endophytic fungus.<sup>361</sup> Below we discuss one prevalent class of biosynthetically undefined bioactive plant peptides, the bouvardins.

### 4.1. Bouvardins

Bouvardins are a particularly noteworthy class of cyclic peptides isolated from plants for which both the true producer and

biosynthetic route is unknown. Bouvardin (Fig. 12) was first isolated in 1977 from *Bouvardia ternifolia*.<sup>362</sup> Since then, over 30 analogues have been found in members of the *Rubia* genus including the RA-series,<sup>363,364</sup> the rubiyunnanins,<sup>365,366</sup> and the rubicordins.<sup>367</sup>

**4.1.1 Structure.** Bouvardins are bicyclic peptides that are composed of both an N- to C-amide macrocycle and an ether bond formed from the phenolic oxygen of a tyrosine and the *meta* carbon of an adjacent tyrosine. They are six amino acids in length and are largely comprised of hydrophobic amino acids. Known bouvardins also contain three separate *N*-methylations of the amide. A D-Ala is found at the same position in all known bouvardins, and the  $\beta$ -carbon of the ether forming tyrosine is often hydroxylated. The free phenolic oxygens are often methylated or glycosylated.<sup>3</sup>

**4.1.2 Biosynthesis.** To date, details on the biosynthesis of bouvardins are scarce. Isolation of endophytes from RA-producing plants of the *Rubia* genus identified four stains that produced metabolites consistent with the RA-series of bouvardins by LC-MS/MS.<sup>368</sup> Whether these molecules are produced by plants or an endophyte, the biosynthetic route remains to be elucidated.

**4.1.3 Bioactivity.** The bouvardin class of cyclic peptides has been investigated for antitumor properties. Bouvardin was shown to be active against leukemia and melanoma cells lines.<sup>362</sup> Subsequent mechanistic studies found that bouvardin inhibits protein synthesis by binding to the 80S subunit of the ribosome,<sup>369</sup> likely stabilizing the interaction between the 80S subunit and elongation factor 2.<sup>370</sup> The RA-series of analogues are also cytotoxic. RA-IV, RA-V, and RA-VII were the first members of the RA-series isolated and each showed promising antitumor activity in mouse studies including leukemia, melanomas, carcinomas, and solid tumors.<sup>371</sup> Later studies found that RA-V<sup>372</sup> and RA-XII<sup>373</sup> are also effective against breast cancer cell lines and appears to limit adhesion, motility, and invasion of the cancer cells, possibly through the PI3K/AKT signaling pathway. In the case of RA-V, the cytotoxic activity was linked to the activation of apoptosis through a mitochondrial pathway.<sup>372</sup> In addition, the rubiyunnanins and the rubicordins<sup>367</sup> have some cytotoxic bioactivity as well and members of the bouvardin class of cyclic peptides also demonstrate anti-angiogenesis,<sup>373,374</sup> NO synthesis inhibition,<sup>366</sup> anti-inflammatory,<sup>375</sup> and NF- $\kappa$ B pathway modulation activities.<sup>366,367,375</sup>

## 5. Chemotaxonomy of plant peptides

We sought to understand the chemotaxonomic distribution of plant RiPPs to identify trends and hot spots where specific peptides may be found. To assemble this data, we identified the order responsible for producing every known molecule in the cyclotide, PawS-derived peptide, orbitide, cysteine rich peptide, and burpitide classes of RiPPs. The presence of these peptides was limited to vascularized plants (Tracheophyta). To visualize the distribution of the plant RiPPs in tracheophytes, we collected the orders within Tracheophyta from the NCBI taxonomy database and used a combination of phyloT and the ggtree R package to generate a cladogram.<sup>376</sup> The cladogram was

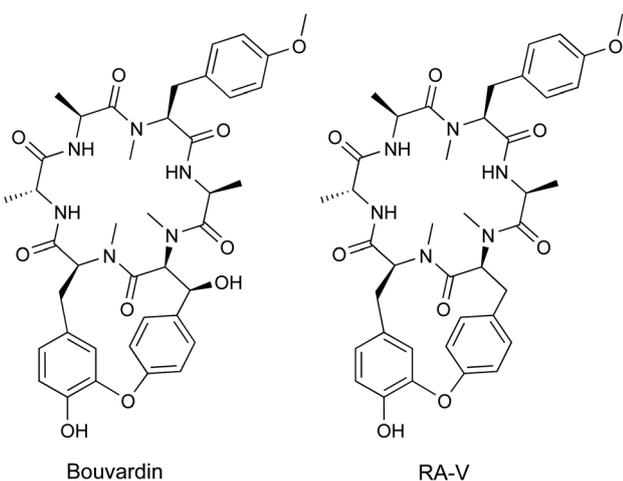


Fig. 12 Bouvardin structures.



then annotated with the RiPP classes data. Our cladogram does not include RALF peptides.

Fig. 13A shows the presence of the different RiPP classes found in plants. The results indicate that eudicots are the most prolific producers of RiPP natural products in plants, with representatives of all five types present. In particular, the rosid clade, including members such as Fabales and Malpighiales, appears to be a particular biosynthetic hot spot. Also of note is the Asterales order which has been shown to biosynthesize all

five classes of plant RiPPs. Outside of eudicots, monocots also contain members that produce RiPPs, but to a lesser extent. The true ferns (Polyopsida) have had no RiPPs from these five classes identified. The cladogram also revealed a differential distribution of each plant RiPP. For example, PawS-derived peptides (PDPs) have only been found in the order Asterales. In contrast, cysteine-rich peptides and burpitides are more widely distributed across nearly all Tracheophyta clades.

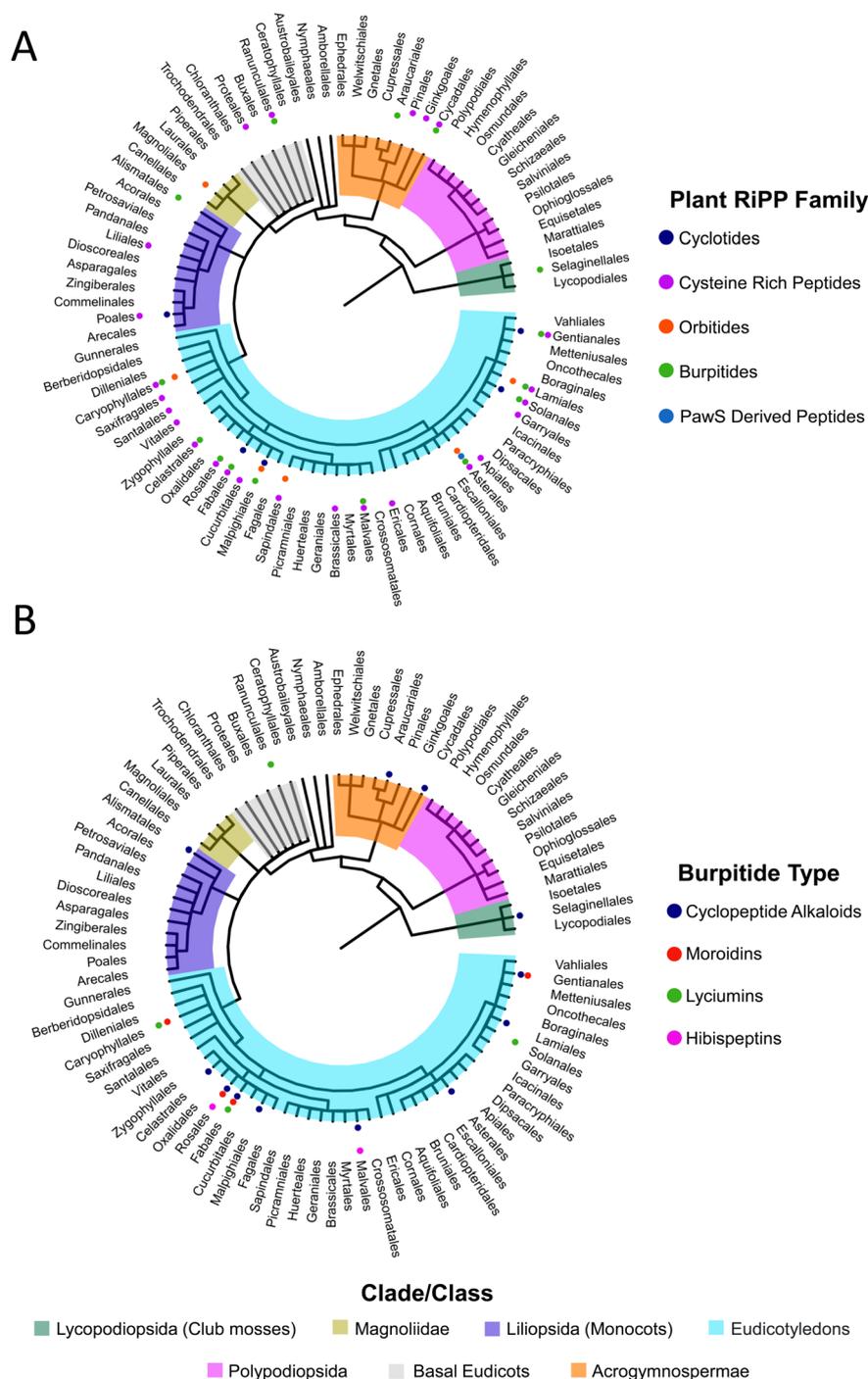


Fig. 13 Chemtaxonomy of Plant RiPPs. Cladograms annotated with the detected presence of plant RiPPs (A) and specific burpitide types (B).



We also generated a second cladogram focused on the distribution of the burpitides (Fig. 13B). Burpitide diversity seems to follow the same trend as all plant RiPP classes because the rosids have the largest diversity of isolated burpitides. Specifically, the Fabales and Rosales both contain three out of the four types of molecules contained within the burpitide class. This cladogram also highlights the fact that cyclopeptide alkaloids are the most widely distributed burpitide discovered to date.

It is important to note that while this figure may not show the authentic distribution of plant RiPPs in nature. As with all isolation experiments, the results can be biased by what specific molecules are being sought, what plant materials are available, and the isolation methodology. The increasing prevalence of plant genomes and transcriptomes is offering a way to target isolation in unexpected producers, such as selanine A in the African clubmoss of the Selaginellales order (see Section 3.6.3).

## 6. Discovery of plant RiPPs

Plant RiPPs have been discovered through bioactivity-guided, structure-guided, and gene-guided approaches that will be reviewed based on representative examples rather than a comprehensive list of discoveries below. A bioactivity-guided approach is less selective for compound classes, whereas structure-guided and gene-guided approaches can be effective in targeted discovery of RiPPs from plants.

### 6.1. Bioactivity-guided discovery

Multiple plant RiPPs have been discovered from source plants through their bioactivity as described in the respective bioactivity sections of Section 3. The choice of source plants is often inspired by application of its extracts in herbal medicine such as kalata B1 as an oxytocic agent, stimulation of pain such as excelsatoin A as Na<sub>v</sub> channel modulator, antimicrobial activity against plant pathogens such as  $\alpha/\beta$ -purothionins, or plant physiological activities such as RALFs.

Challenges of bioactivity-guided discovery can be time-consuming fractionation and rediscovery of known bioactive natural products. To this end, a pipeline called PepSAVI (statistically-guided bioactive peptides prioritized *via* mass spectrometry) was developed to correlate bioactivity of pre-fractionated plant extracts with the presence of peptide mass signals in liquid-chromatography mass spectrometry datasets of the extracts. The masses of candidate bioactive peptides identified by this pipeline enable dereplication by MS1 and tandem MS approaches that was validated by the characterization of cyclotide cyO2 as an antibacterial and cancer cell cytotoxic lead compound from a *Viola odorata* extract fraction library.<sup>377</sup> The assay further enabled discovery of candidate bioactive cyclotides from *V. odorata* by improved tandem mass spectrometry-based sequencing of target peptides *via* ultraviolet photodissociation (UVPD) tandem mass spectrometry.<sup>378</sup> Pipelines such as PepSAVI that connect high-throughput bioactivity assay data with mass spectrometry can rapidly dereplicate and identify new bioactive plant RiPPs.<sup>379</sup> MS-based approaches

such as molecular networking<sup>380</sup> or Dereplicator+<sup>381</sup> can enable rapid dereplication<sup>382</sup> and discovery<sup>383</sup> of bioactive lead compounds *via* spectral library comparison, for example within the Global Natural Product Molecular Networking platform,<sup>384</sup> and they can further assess structural novelty by spectral library comparison<sup>385</sup> or *de novo* structure elucidation.<sup>386,387</sup>

### 6.2. Structure-guided discovery

Many plant RiPPs were discovered based on physicochemical properties during natural product isolation or their structural features revealed by analytical chemistry techniques such as mass spectrometry (MS) as an initial chemotyping step.

Heveins were isolated from a centrifugation fraction including lutoids, vacuoles in latex-producing cells of rubber tree,<sup>388</sup> and cliotides were discovered from heat stable extract fractions as compounds with high heat stability, a well characterized feature of cyclotides and CRPs.<sup>389</sup> Besides physicochemical features, plant RiPPs have been effectively characterized *via* mass spectrometry approaches. Structural information can be derived from a peptide analyte *via* tandem mass spectrometry experiments; however it can be challenging as many plant RiPPs have macrocyclic PTMs, which often prevents sequencing of a given peptide. Therefore, cysteine-rich peptides usually require disulfide bond reduction and *S*-alkylation before tandem mass spectrometry analysis to characterize the complete core peptide sequence, analogous to proteomic protocols for MS-based protein sequencing.<sup>390–392</sup> The characteristic mass shift of iodoacetamide-alkylation at reduced cysteines (+57.0215 *m/z*) has been used for discovery of disulfide-containing plant RiPPs in plant extracts.<sup>40,65,71,173,175,184,252,389</sup> A plant extract can be analyzed for the presence of analytes in the mass range of CRPs and cyclotides (2–6 kDa), treated for disulfide reduction and cysteine alkylation, then analyzed for alkylation-specific mass shifts in candidate peptides by comparison to untreated samples. For cyclotides, an additional step of proteolysis is usually required after disulfide reduction and cysteine alkylation to linearize the head-to-tail-macrocyclic prior to tandem MS analysis. Proteolysis often involves GluC due to the presence of a conserved Glu in cyclotide core peptides<sup>22</sup> or trypsin which can yield multiple tryptic core peptide species.<sup>97</sup> Despite the efficacy of chemical and proteolytic processing of cyclotides and CRPs prior to tandem MS sequencing, the main disadvantage of this sample preparation is higher sample requirement, which can exclude low abundance peptides from discovery. Recently, ultraviolet photodissociation (UVPD) tandem mass spectrometry has been introduced to yield more fragmentation and therefore peptide sequencing data from cyclotides and CRPs in underivatized plant extract samples.<sup>378</sup> In addition, collision-induced dissociation, which is the predominant tandem MS method in metabolomics, was further developed for MS fingerprinting *via* short sequence tags enabling cyclotide dereplication and prioritization for discovery.<sup>393</sup> For burpitides, tandem mass spectrometry (MS/MS) can be applied for characterization of structural features to reveal the presence of a peptidic analyte and specific amino acids within the peptide sequence.<sup>18,394</sup> In



addition, tandem mass spectral data can be used in comparison to spectra of known plant peptides in custom or public databases for peptide classification.<sup>384,395–397</sup> All aspects of peptide identification *via* MS/MS can inform novelty of a putative plant peptide and guide further experiments. Predicted amino acids and the size of the target analyte can enable the connection of a core peptide in a BURP-domain precursor peptide to the target spectrum. This approach has been applied effectively for the discovery of lyciumin-type and stephanotic acid-type burpitides, and cyclopeptide alkaloids. It can also be applied without a precursor peptide sequence as exemplified by the discovery of the glycosylated cyclopeptide alkaloid encephanine.<sup>18</sup> The connection of tandem mass spectra of burpitides to their corresponding precursor peptides sequences has been implemented in the RiPP discovery platform HypoRiPPAtlas, which enables prediction of RiPP structures from candidate core peptide sequences in BURP-domain-containing precursor peptides. Subsequently, an algorithm called Dereplicator+<sup>381</sup> is used to generate theoretical tandem mass spectra of the predicted RiPP structures and compare them to experimental tandem MS data for RiPP discovery. The HypoRiPPAtlas was able to identify several cyclopeptide alkaloids including cyclopeptide alkaloid elaeagnin (Fig. 8) and lyciumins as a proof-of-concept.<sup>21</sup> Orbitides can be characterized by tandem MS in a similar fashion by comparison of tandem mass spectra to predicted precursor peptides identified in plant transcriptomes or by *de novo* sequencing.<sup>398</sup> While mass spectrometry provides new plant RiPP candidates in structure-guided discovery approaches, NMR or crystallography are usually applied on scaled peptide samples to determine new structures of candidate peptides.<sup>88,90</sup> With growing plant metabolomic databases, mass spectrometry-guided approaches will further drive plant RiPP discovery.

### 6.3. Gene-guided discovery

The characterization of precursor peptide genes (Table 1) and the growth of plant genetic resources<sup>25,399–401</sup> has enabled the discovery of plant RiPPs through bioinformatic prediction of peptide chemotypes from plant genes.

Cyclotides were discovered in *Petunia* plants by searching for cyclotide precursor peptides in the EST database of the NCBI.<sup>182</sup> Many CRPs have been predicted in plant transcriptomes by bioinformatic analyses, for instance, identification of a new thionin from *Papaver somniferum*.<sup>190,402</sup> Similarly, genome mining and transcriptome analysis for fused burpitide precursor peptides has led to the discovery of new lyciumins, cyclopeptide alkaloids, and moroidins<sup>17,19</sup> and split burpitide precursor peptides that resulted in discovery of new cyclopeptide alkaloids and hibispeptins.<sup>20</sup> The characterization of a precursor peptide in any gene-guided discovery experiment yields a candidate core peptide sequence which can facilitate chemical characterization, *e.g.* by providing the long sequence information of cyclotides and CRPs. In addition, precursor peptides sequences can provide information about their processing through the secretory pathway *via* signal peptide prediction with SignalP.<sup>403</sup> A challenge for gene-guided discovery of plant RiPPs is the identification of new

precursor peptides and class-defining PTM enzymes in genomes with less biosynthetic gene clustering than in bacteria. A general approach to overcome this challenge is to search for core peptide sequences in predicted ORFs of peptide source plants<sup>4,5,7,12–14,17,52</sup> or identify candidate precursor peptides based on genomic co-localization with scaffold-generating RiPP PTM genes.<sup>20</sup> Finally, tissue-specific paired transcriptomics and metabolomics has proven to be an essential approach for elucidating pathways for which biosynthetic enzymes are not co-localized.<sup>404</sup>

## 7. Future directions

Plant peptides offer exciting research questions in terms of their discovery and biosynthesis and opportunities for further biotechnological and medicinal applications. Like microbial RiPP classes, the field of plant RiPPs is experiencing a gold rush in peptide discovery fueled by increasing genetic resources such as the 1KP database,<sup>400</sup> the 10KP project,<sup>25</sup> Phytozome<sup>401</sup> and metabolomic databases such as GNPS.<sup>384</sup> While many RiPPs have been predicted based on bioinformatic studies, an important direction is the integration of omics-datasets for automated plant RiPP discovery as it has been organized for microbial natural products<sup>405</sup> and chemical characterization of predicted RiPPs. In addition, discovery of plant RiPPs will depend on identification of new precursor peptides such as fused and split burpitide precursor peptides. Given the localization of some RiPP core peptides within larger plant proteins, this identification might require larger theoretical databases to match candidate peptide spectral data to core peptides within proteins encoded in a plant transcriptome or genome. Regarding sequencing data, it will be important to further improve *de novo* assembly for generation of repetitive precursor genes. Regarding mass spectrometry, tandem mass spectrometry experiments could further improve to generate more core peptide sequence data for *de novo* or gene-aided structure elucidation. Remaining biosynthetic questions of plant RiPPs include elucidating the origin of known plant-sourced peptides such as bouvardins and whether non-ribosomal peptide biosynthesis exists within the plant kingdom. In terms of biosynthesis, the several processing proteases of head-to-tail-cyclic peptides are still unknown. For burpitides, structural and mechanistic insight into macrocyclizations by burpitide cyclases is lacking and major biosynthetic questions regarding processing of both split and fused burpitide precursor are unanswered including reactions involved in proteolysis and terminal core-peptide modifications. With the combination of rapidly developing genomic and biosynthetic insights, the future of plant RiPPs holds great promise of the discovery of new bioactive molecules, enzyme reactions, and engineering platforms.

## 8. Conflicts of interest

The authors declare no conflicts of interest.

## 9. Acknowledgments

This work was partially supported by the US National Institutes of Health (R35GM147439 to J. R. C., F32GM146395 to L. S. M.,



R35GM146934 to R. D. K). We thank Sanath K. Kandy for generating a graphical abstract of this review.

## 10. References

- P. G. Arnison, *et al.*, Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature, *Nat. Prod. Rep.*, 2013, **30**, 108–160.
- M. Montalbán-López, *et al.*, New developments in RiPP discovery, enzymology and engineering, *Nat. Prod. Rep.*, 2021, **38**, 130–239.
- N.-H. Tan and J. Zhou, Plant cyclopeptides, *Chem. Rev.*, 2006, **106**, 840–895.
- C. Jennings, J. West, C. Waine, D. Craik and M. Anderson, Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis*, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10614–10619.
- J. S. Mylne, *et al.*, Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis, *Plant Cell*, 2012, **24**, 2765–2778.
- A. S. Jayasena, *et al.*, Stepwise evolution of a buried inhibitor peptide over 45 my, *Mol. Biol. Evol.*, 2017, **34**, 1505–1516.
- J. A. Condie, *et al.*, The biosynthesis of Caryophyllaceae-like cyclic peptides in *Saponaria vaccaria* L. from DNA-encoded precursors, *Plant J.*, 2011, **67**, 682–690.
- Z. Song, C. Burbridge, D. J. Schneider, T. F. Sharbel and M. J. T. Reaney, The flax genome reveals orbitide diversity, *BMC Genomics*, 2022, **23**, 534.
- Y. Matsubayashi and Y. Sakagami, Phytosulfokine, sulfated peptides that induce the proliferation of single mesophyll cells of *Asparagus officinalis* L, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 7623–7627.
- J. C. Fletcher, U. Brand, M. P. Running, R. Simon and E. M. Meyerowitz, Signaling of cell fate decisions by CLAVATA3 in *Arabidopsis* shoot meristems, *Science*, 1999, **283**, 1911–1914.
- G. Pearce, D. S. Moura, J. Stratmann and C. A. Ryan, Production of multiple plant hormones from a single polyprotein precursor, *Nature*, 2001, **411**, 817–820.
- J. Villanueva, *et al.*, Characterization of the wound-induced metallocarboxypeptidase inhibitor from potato1, *FEBS Lett.*, 1998, **440**, 175–182.
- M. F. De Bolle, *et al.*, Cloning and characterization of a cDNA encoding an antimicrobial chitin-binding protein from amaranth, *Amaranthus caudatus*, *Plant Mol. Biol.*, 1993, **22**, 1187–1190.
- I. Broekaert, H. I. Lee, A. Kush, N. H. Chua and N. Raikhel, Wound-induced accumulation of mRNA containing a hevein sequence in laticifers of rubber tree (*Hevea brasiliensis*), *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 7633–7637.
- S. G. Kini, K. H. Wong, W. L. Tan, T. Xiao and J. P. Tam, Morintides: cargo-free chitin-binding peptides from *Moringa oleifera*, *BMC Plant Biol.*, 2017, **17**, 68.
- K. P. B. Van den Bergh, *et al.*, Synergistic antifungal activity of two chitin-binding proteins from spindle tree (*Euonymus europaeus* L.), *Planta*, 2004, **219**, 221–232.
- R. D. Kersten and J.-K. Weng, Gene-guided discovery and engineering of branched cyclic peptides in plants, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E10961–E10969.
- D. N. Chigumba, *et al.*, Discovery and biosynthesis of cyclic plant peptides via autocatalytic cyclases, *Nat. Chem. Biol.*, 2022, **18**, 18–28.
- R. D. Kersten, *et al.*, Gene-Guided Discovery and Ribosomal Biosynthesis of Moroidin Peptides, *J. Am. Chem. Soc.*, 2022, **144**, 7686–7692.
- S. T. Lima, *et al.*, A widely distributed biosynthetic cassette is responsible for diverse plant side chain cross-linked cyclopeptides, *Angew Chem. Int. Ed. Engl.*, 2023, **62**, e202218082.
- Y.-Y. Lee, *et al.*, HypoRiPPAtlas as an Atlas of hypothetical natural products for mass spectrometry database search, *Nat. Commun.*, 2023, **14**, 4219.
- A. G. Poth, M. L. Colgrave, R. E. Lyons, N. L. Daly and D. J. Craik, Discovery of an unusual biosynthetic origin for circular proteins in legumes, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 10127–10132.
- G. K. Oguis, E. K. Gilding, M. A. Jackson and D. J. Craik, Butterfly pea (*Clitoria ternatea*), a cyclotide-bearing plant with applications in agriculture and medicine, *Front. Plant Sci.*, 2019, **10**, 645.
- C. Gründemann, K. G. Stenberg and C. W. Gruber, T20K: An immunomodulatory cyclotide on its way to the clinic, *Int. J. Pept. Res. Ther.*, 2019, **25**, 9–13.
- S. Cheng, *et al.*, 10KP: A phylodiverse genome sequencing plan, *Gigascience*, 2018, **7**, 1–9.
- C. K. L. Wang, Q. Kaas, L. Chiche and D. J. Craik, CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering, *Nucleic Acids Res.*, 2008, **36**, D206–D210.
- R. Hammami, J. Ben Hamida, G. Vergoten and I. Fliss, PhytAMP: a database dedicated to antimicrobial plant peptides, *Nucleic Acids Res.*, 2009, **37**, D963–D968.
- J. Piel, *et al.*, Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 16222–16227.
- M. D. Tianero, J. N. Balaich and M. S. Donia, Localized production of defence chemicals by intracellular symbionts of *Haliclona* sponges, *Nat. Microbiol.*, 2019, **4**, 1149–1159.
- Y. Xu, *et al.*, Bacterial biosynthesis and maturation of the didemnin anti-cancer agents, *J. Am. Chem. Soc.*, 2012, **134**, 8625–8632.
- M. Crüsemann, *et al.*, Heterologous Expression, Biosynthetic Studies, and Ecological Function of the Selective Gq-Signaling Inhibitor FR900359, *Angew. Chem., Int. Ed.*, 2018, **57**, 836–840.
- A. Carlier, *et al.*, The genome analysis of *Candidatus Burkholderia crenata* reveals that secondary metabolism may be a key function of the *Ardisia crenata* leaf nodule symbiosis, *Environ. Microbiol.*, 2016, **18**, 2507–2522.
- C. Hermes, G. M. König and M. Crüsemann, The chromodepsins - chemistry, biology and biosynthesis of



- a selective Gq inhibitor natural product family, *Nat. Prod. Rep.*, 2021, **38**, 2276–2292.
- 34 C. Hermes, *et al.*, Thioesterase-mediated side chain transesterification generates potent Gq signaling inhibitor FR900359, *Nat. Commun.*, 2021, **12**, 144.
- 35 J. G. Schlegel, *et al.*, Macrocyclic Gq Protein Inhibitors FR900359 and/or YM-254890-Fit for Translation?, *ACS Pharmacol. Transl. Sci.*, 2021, **4**, 888–897.
- 36 R. Tschesche, M. Elgamal, G. A. Miana and G. Eckhardt, Alkaloids from rhamnaceae—XXVI, *Tetrahedron*, 1975, **31**, 2944–2947.
- 37 A. Segura, M. Moreno, F. Madueño, A. Molina and F. García-Olmedo, Snakin-1, a peptide from potato that is active against plant pathogens, *Mol. Plant Microbe Interact.*, 1999, **12**, 16–23.
- 38 G. M. Rubin and Y. Ding, Recent advances in the biosynthesis of RiPPs from multicore-containing precursor peptides, *J. Ind. Microbiol. Biotechnol.*, 2020, **47**, 659–674.
- 39 M. F. Fisher, J. Zhang, O. Berkowitz, J. Whelan and J. S. Mylne, Cyclic peptides in seed of *Annona muricata* are ribosomally synthesized, *J. Nat. Prod.*, 2020, **83**, 1167–1173.
- 40 Y. Shen, *et al.*, Potentides: New cysteine-rich peptides with unusual disulfide connectivity from *Potentilla anserina*, *Chembiochem*, 2019, **20**, 1995–2004.
- 41 D. Y. Ryazantsev, *et al.*, A novel hairpin-like antimicrobial peptide from barnyard grass (*Echinochloa crusgalli* L.) seeds: Structure-functional and molecular-genetics characterization, *Biochimie*, 2014, **99**, 63–70.
- 42 R. H. Taylor, *et al.*, A novel family of small cysteine-rich antimicrobial peptides from seed of *Impatiens balsaminalis* derived from a single precursor protein, *J. Biol. Chem.*, 1997, **272**, 24480–24487.
- 43 W. Ding, *et al.*, Biosynthetic investigation of phomopsins reveals a widespread pathway for ribosomal natural products in Ascomycetes, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 3521–3526.
- 44 Y. Ye, *et al.*, Heterologous production of asperipin-2a: proposal for sequential oxidative macrocyclization by a fungi-specific DUF3328 oxidase, *Org. Biomol. Chem.*, 2018, **17**, 39–43.
- 45 N. Nagano, *et al.*, Class of cyclic ribosomal peptide synthetic genes in filamentous fungi, *Fungal Genet. Biol.*, 2016, **86**, 58–70.
- 46 M. Umemura, *et al.*, Characterization of the biosynthetic gene cluster for the ribosomally synthesized cyclic peptide ustiloxin B in *Aspergillus flavus*, *Fungal Genet. Biol.*, 2014, **68**, 23–30.
- 47 A. S. Imani, A. R. Lee, N. Vishwanathan, F. de Waal and M. F. Freeman, Diverse protein architectures and  $\alpha$ -N-methylation patterns define split borosin RiPP biosynthetic gene clusters, *ACS Chem. Biol.*, 2022, **17**, 908–917.
- 48 M. R. Quijano, *et al.*, Distinct autocatalytic  $\alpha$ -N-methylating precursors expand the borosin RiPP family of peptide natural products, *J. Am. Chem. Soc.*, 2019, **141**, 9637–9644.
- 49 E. W. Schmidt, *et al.*, Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7315–7320.
- 50 M. S. Donia, *et al.*, Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians, *Nat. Chem. Biol.*, 2006, **2**, 729–735.
- 51 N. Ziemert, K. Ishida, A. Liaimer, C. Hertweck and E. Dittmann, Ribosomal synthesis of tricyclic depsipeptides in bloom-forming Cyanobacteria, *Angew. Chem.*, 2008, **120**, 7870–7873.
- 52 J. S. Mylne, *et al.*, Albumins and their processing machinery are hijacked for cyclic peptides in sunflower, *Nat. Chem. Biol.*, 2011, **7**, 257–259.
- 53 B. Franke, J. S. Mylne and K. J. Rosengren, Buried treasure: biosynthesis, structures and applications of cyclic peptides hidden in seed storage albumins, *Nat. Prod. Rep.*, 2018, **35**, 137–146.
- 54 R.-H. Huang, *et al.*, Two novel antifungal peptides distinct with a five-disulfide motif from the bark of *Eucommia ulmoides* Oliv, *FEBS Lett.*, 2002, **521**, 87–90.
- 55 A. M. James, *et al.*, Evidence for Ancient Origins of Bowman-Birk Inhibitors from *Selaginella moellendorffii*, *Plant Cell*, 2017, **29**, 461–473.
- 56 C. T. Walsh, S. Garneau-Tsodikova and G. J. Gatto Jr, Protein posttranslational modifications: the chemistry of proteome diversifications, *Angew. Chem. Int. Ed. Engl.*, 2005, **44**, 7342–7372.
- 57 T. L. Aboye, *et al.*, Interlocking disulfides in circular proteins: Toward efficient oxidative folding of cyclotides, *Antioxid. Redox Signaling*, 2011, **14**, 77–86.
- 58 M. Čemažar, S. Zahariev, S. Pongor and P. J. Hore, Oxidative folding of *Amaranthus*  $\alpha$ -amylase inhibitor, *J. Biol. Chem.*, 2004, **279**, 16697–16705.
- 59 C. W. Gruber, *et al.*, A novel plant protein-disulfide isomerase involved in the oxidative folding of cystine knot defense proteins, *J. Biol. Chem.*, 2007, **282**, 20435–20446.
- 60 S. Ramm, *et al.*, A self-sacrificing N-methyltransferase is the precursor of the fungal natural product omphalotin, *Angew. Chem. Int. Ed. Engl.*, 2017, **56**, 9994–9997.
- 61 N. S. van der Velden, *et al.*, Autocatalytic backbone N-methylation in a family of ribosomal peptide natural products, *Nat. Chem. Biol.*, 2017, **13**, 833–835.
- 62 D. Gruis, J. Schulze and R. Jung, Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases, *Plant Cell*, 2004, **16**, 270–290.
- 63 T. Shimada, *et al.*, Vacuolar processing enzymes are essential for proper processing of seed storage proteins in *Arabidopsis thaliana*, *J. Biol. Chem.*, 2003, **278**, 32292–32299.
- 64 T. Paysan-Lafosse, *et al.*, InterPro in 2022, *Nucleic Acids Res.*, 2023, **51**, D418–D427.
- 65 J. P. Tam, *et al.*, Ginsentides: Cysteine and Glycine-rich peptides from the ginseng family with unusual disulfide connectivity, *Sci. Rep.*, 2018, **8**, 16201.



- 66 Y. A. Andreev, *et al.*, Genes encoding hevein-like defense peptides in wheat: Distribution, evolution, and role in stress response, *Biochimie*, 2012, **94**, 1009–1016.
- 67 F. J. Colilla, A. Rocher and E. Mendez, gamma-Purothionins: amino acid sequence of two polypeptides of a new family of thionins from wheat endosperm, *FEBS Lett.*, 1990, **270**, 191–194.
- 68 S. B. Ghag, U. K. S. Shekhawat and T. R. Ganapathi, Petunia floral defensins with unique prodomains as novel candidates for development of Fusarium wilt resistance in transgenic banana plants, *PLoS One*, 2012, **7**, e39557.
- 69 G. Schrader-Fischer and K. Apel, Organ-specific expression of highly divergent thionin variants that are distinct from the seed-specific crambin in the crucifer *Crambe abyssinica*, *Mol. Gen. Genet.*, 1994, **245**, 380–389.
- 70 A. Castagnaro, C. Marana, P. Carbonero and F. Garcia-Olmedo, CDNA cloning and nucleotide sequences of [alpha]1 and [alpha]2 thionins from hexaploid wheat endosperm, *Plant Physiol.*, 1994, **106**, 1221–1222.
- 71 G. Kumari, *et al.*, Cysteine-rich peptide family with unusual disulfide connectivity from *Jasminum sambac*, *J. Nat. Prod.*, 2015, **78**, 2791–2799.
- 72 B. Alunni, *et al.*, Genomic organization and evolutionary insights on *GRP* and *NCR* genes, two large nodule-specific gene families in *Medicago truncatula*, *Mol. Plant Microbe Interact.*, 2007, **20**, 1138–1148.
- 73 G. Pearce, D. S. Moura, J. Stratmann and C. A. Ryan Jr, RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 12843–12847.
- 74 G. K. T. Nguyen, *et al.*, Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis, *Nat. Chem. Biol.*, 2014, **10**, 732–738.
- 75 K. S. Harris, *et al.*, Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase, *Nat. Commun.*, 2015, **6**, 10199.
- 76 F. B. H. Rehm, *et al.*, Papain-like cysteine proteases prepare plant cyclic peptide precursors for cyclization, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 7831–7836.
- 77 C. J. S. Barber, *et al.*, The two-step biosynthesis of cyclic peptides from linear precursors in a member of the plant family Caryophyllaceae involves cyclization by a serine protease-like enzyme, *J. Biol. Chem.*, 2013, **288**, 12500–12510.
- 78 R. Komori, Y. Amano, M. Ogawa-Ohnishi and Y. Matsubayashi, Identification of tyrosylprotein sulfotransferase in *Arabidopsis*, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 15067–15072.
- 79 M. Ogawa-Ohnishi, W. Matsushita and Y. Matsubayashi, Identification of three hydroxyproline O-arabinosyltransferases in *Arabidopsis thaliana*, *Nat. Chem. Biol.*, 2013, **9**, 726–730.
- 80 L. Gran, An oxytocic principle found in *Oldenlandia affinis* DC, *Medd. Nor. Farm. Selsk.*, 1970, **12**, 80.
- 81 L. Gran, Oxytocic principles of *Oldenlandia affinis*, *Lloydia*, 1973, **36**, 174–178.
- 82 L. Grain, Isolation of oxytocic peptides from *Oldenlandia affinis* by solvent extraction of tetraphenylborate complexes and chromatography on sephadex LH-20, *Lloydia*, 1973, **36**, 207–208.
- 83 K. Sletten and L. Gran, Some molecular properties of kalatapeptide B-1. A uterotonic polypeptide isolated from *Oldenlandia affinis* DC, *Medd. Nor. Farm. Selsk.*, 1973, **7**, 69–82.
- 84 J. P. Mulvenna, C. Wang and D. J. Craik, CyBase: a database of cyclic protein sequence and structure, *Nucleic Acids Res.*, 2006, **34**, D192–D194.
- 85 J. Weidmann and D. J. Craik, Discovery, structure, function, and applications of cyclotides: circular proteins from plants, *J. Exp. Bot.*, 2016, **67**, 4801–4812.
- 86 S. J. Veer, M. W. Kan and D. J. Craik, Cyclotides: from structure to function, *Chem. Rev.*, 2019, **119**, 12375–12421.
- 87 D. J. Craik, N. L. Daly, T. Bond and C. Waine, Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif, *J. Mol. Biol.*, 1999, **294**, 1327–1336.
- 88 O. Saether, *et al.*, Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1, *Biochemistry*, 1995, **34**, 4147–4158.
- 89 J. P. Tam, Y. A. Lu, J. L. Yang and K. W. Chiu, An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 8913–8918.
- 90 D. C. Rees and W. N. Lipscomb, Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution, *J. Mol. Biol.*, 1982, **160**, 475–498.
- 91 M. L. Colgrave and D. J. Craik, Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot, *Biochemistry*, 2004, **43**, 5965–5975.
- 92 A. S. Ravipati, *et al.*, Lysine-rich cyclotides: A new subclass of circular knotted proteins from Violaceae, *ACS Chem. Biol.*, 2015, **10**, 2491–2500.
- 93 J. L. Dutton, *et al.*, Conserved structural and sequence elements implicated in the processing of gene-encoded circular proteins, *J. Biol. Chem.*, 2004, **279**, 46858–46867.
- 94 J. P. Mulvenna, L. Sando and D. J. Craik, Processing of a 22 kDa precursor protein to produce the circular protein tricyclon A, *Structure*, 2005, **13**, 691–701.
- 95 S. M. Simonsen, *et al.*, A continent of plant defense peptide diversity: cyclotides in Australian *Hybanthus* (Violaceae), *Plant Cell*, 2005, **17**, 3176–3189.
- 96 A. G. Poth, *et al.*, Discovery of cyclotides in the fabaceae plant family provides new insights into the cyclization, evolution, and distribution of circular proteins, *ACS Chem. Biol.*, 2011, **6**, 345–355.
- 97 G. K. T. Nguyen, *et al.*, Discovery and characterization of novel cyclotides originated from chimeric precursors consisting of albumin-1 chain a and cyclotide domains in the Fabaceae family, *J. Biol. Chem.*, 2011, **286**, 24275–24287.
- 98 R. Yang, *et al.*, Engineering a catalytically efficient recombinant protein ligase, *J. Am. Chem. Soc.*, 2017, **139**, 5351–5358.



- 99 M. Jackson, *et al.*, Molecular basis for the production of cyclic peptides by the plant asparaginyl endopeptidases, *Nat. Commun.*, 2018, 1–12.
- 100 J. Du, *et al.*, A bifunctional asparaginyl endopeptidase efficiently catalyzes both cleavage and cyclization of cyclic trypsin inhibitors, *Nat. Commun.*, 2020, **11**, 1575.
- 101 B. L. Barbeta, A. T. Marshall, A. D. Gillon, D. J. Craik and M. A. Anderson, Plant cyclotides disrupt epithelial cells in the midgut of lepidopteran larvae, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 1221–1225.
- 102 M. F. S. Pinto, *et al.*, Identification and structural characterization of novel cyclotide with activity against an insect pest of sugar cane, *J. Biol. Chem.*, 2012, **287**, 134–147.
- 103 M. R. R. Plan, I. Saska, A. G. Cagauan and D. J. Craik, Backbone cyclised peptides from plants show molluscicidal activity against the rice pest *Pomacea canaliculata* (golden apple snail), *J. Agric. Food Chem.*, 2008, **56**, 5237–5241.
- 104 M. L. Colgrave, *et al.*, Cyclotides: natural, circular plant peptides that possess significant activity against gastrointestinal nematode parasites of sheep, *Biochemistry*, 2008, **47**, 5581–5589.
- 105 M. L. Colgrave, *et al.*, Anthelmintic activity of cyclotides: In vitro studies with canine and human hookworms, *Acta Trop.*, 2009, **109**, 163–166.
- 106 P. Lindholm, *et al.*, Cyclotides: a novel type of cytotoxic agents, *Mol. Cancer Ther.*, 2002, **1**, 365–369.
- 107 E. Svängård, *et al.*, Cytotoxic cyclotides from *Viola tricolor*, *J. Nat. Prod.*, 2004, **67**, 144–147.
- 108 A. Herrmann, *et al.*, The alpine violet, *Viola biflora*, is a rich source of cyclotides with potent cytotoxicity, *Phytochemistry*, 2008, **69**, 939–952.
- 109 S. L. Gerlach, R. Burman, L. Bohlin, D. Mondal and U. Göransson, Isolation, characterization, and bioactivity of cyclotides from the Micronesian plant *Psychotria leptothyrsa*, *J. Nat. Prod.*, 2010, **73**, 1207–1213.
- 110 M. Y. Yeshak, R. Burman, K. Asres and U. Göransson, Cyclotides from an extreme habitat: characterization of cyclic peptides from *Viola abyssinica* of the Ethiopian highlands, *J. Nat. Prod.*, 2011, **74**, 727–731.
- 111 Z. Sen, X. K. Zhan, J. Jing, Z. Yi and Z. Wanqi, Chemosensitizing activities of cyclotides from *Clitoria ternatea* in paclitaxel-resistant lung cancer cells, *Oncol. Lett.*, 2013, **5**, 641–644.
- 112 M. A. Esmaeili, *et al.*, *Viola* plant cyclotide vigno 5 induces mitochondria-mediated apoptosis via cytochrome C release and caspases activation in cervical cancer cells, *Fitoterapia*, 2016, **109**, 162–168.
- 113 M. E. F. Pinto, *et al.*, Inhibition of breast cancer cell migration by cyclotides isolated from *Pombalia calceolaria*, *J. Nat. Prod.*, 2018, **81**, 1203–1208.
- 114 X. Ding, D. Bai and J. Qian, Novel cyclotides from *Hedyotis biflora* inhibit proliferation and migration of pancreatic cancer cell in vitro and in vivo, *Med. Chem. Res.*, 2014, **23**, 1406–1413.
- 115 M. Pránting, C. Lööv, R. Burman, U. Göransson and D. I. Andersson, The cyclotide cycloviolacin O2 from *Viola odorata* has potent bactericidal activity against Gram-negative bacteria, *J. Antimicrob. Chemother.*, 2010, **65**, 1964–1971.
- 116 M. Zarrabi, R. Dalirfardouei, Z. Sepehrzade and R. K. Kermanshahi, Comparison of the antimicrobial effects of semipurified cyclotides from Iranian *Viola odorata* against some of plant and human pathogenic bacteria, *J. Appl. Microbiol.*, 2013, **115**, 367–375.
- 117 M. Narayani, A. Chadha and S. Srivastava, Callus and cell suspension culture of *Viola odorata* as in vitro production platforms of known and novel cyclotides, *Plant Cell, Tissue Organ Cult.*, 2017, **130**, 289–299.
- 118 A. A. Strömstedt, S. Park, R. Burman and U. Göransson, Bactericidal activity of cyclotides where phosphatidylethanolamine-lipid selectivity determines antimicrobial spectra, *Biochim. Biophys. Acta, Biomembr.*, 2017, **1859**, 1986–2000.
- 119 K. R. Gustafson, *et al.*, Circulins A and B. Novel human immunodeficiency virus (HIV)-inhibitory macrocyclic peptides from the tropical tree *Chassalia parvifolia*, *J. Am. Chem. Soc.*, 1994, **116**, 9337–9338.
- 120 J.-F. Hernandez, *et al.*, Squash trypsin inhibitors from *Momordica cochinchinensis* exhibit an atypical macrocyclic structure, *Biochemistry*, 2000, **39**, 5722–5730.
- 121 H. P. Kaufmann and A. Tobschirbel, Über ein Oligopeptid aus Leinsamen, *Chem. Ber.*, 1959, **92**, 2805–2809.
- 122 M. F. Fisher, *et al.*, The genetic origin of evolidine, the first cyclopeptide discovered in plants, and related orbitides, *J. Biol. Chem.*, 2020, **295**, 14510–14521.
- 123 M. F. Fisher, *et al.*, A family of small, cyclic peptides buried in preproalbumin since the Eocene epoch, *Plant Direct*, 2018, **2**, e00042.
- 124 A. G. Elliott, *et al.*, Evolutionary origins of a bioactive peptide buried within Preproalbumin, *Plant Cell*, 2014, **26**, 981–995.
- 125 D. P. Okinyo-Owiti, L. Young, P.-G. G. Burnett and M. J. T. Reaney, New flaxseed orbitides: Detection, sequencing, and <sup>15</sup>N incorporation, *Biopolymers*, 2014, **102**, 168–175.
- 126 W. R. Belknap, *et al.*, A family of small cyclic amphipathic peptides (SCAMPs) genes in citrus, *BMC Genomics*, 2015, **16**, 303.
- 127 M. F. Fisher, C. D. Payne, K. J. Rosengren and J. S. Mylne, An orbitide from *Ratibida columnifera* seed containing 16 amino acid residues, *J. Nat. Prod.*, 2019, **82**, 2152–2158.
- 128 Y.-R. Zhao, *et al.*, Cyclopeptides from *Stellaria yunnanensis*, *Phytochemistry*, 1995, **40**, 1453–1456.
- 129 B. Picur, M. Lisowski and I. Z. Siemion, A new cyclolinopeptide containing nonproteinaceous amino acid-N-methyl-4-aminoproline, *Lett. Pept. Sci.*, 1998, **5**, 183–187.
- 130 J. R. Chekan, P. Estrada, P. S. Covelto and S. K. Nair, Characterization of the macrocyclase involved in the biosynthesis of RiPP cyclic peptides in plants, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 6551–6556.
- 131 H. Ludewig, *et al.*, Characterization of the fast and promiscuous macrocyclase from plant PCY1 enables the



- use of simple substrates, *ACS Chem. Biol.*, 2018, **13**, 801–811.
- 132 B. Gui, *et al.*, Identification and quantification of cyclolinopeptides in five flaxseed cultivars, *J. Agric. Food Chem.*, 2012, **60**, 8571–8579.
- 133 P.-G. G. Burnett, P. D. Jadhav, D. P. Okinyo-Owiti, A. G. Poth and M. J. T. Reaney, Glycine-containing flaxseed orbitides, *J. Nat. Prod.*, 2015, **78**, 681–688.
- 134 Z. Wieczorek, B. Bengtsson, J. Trojnar and I. Z. Siemion, Immunosuppressive activity of cyclolinopeptide A, *Pept. Res.*, 1991, **4**, 275–283.
- 135 I. Z. Siemion, A. Pędyczak, J. Trojnar, M. Zimecki and Z. Wieczorek, Immunosuppressive activity of antamanide and some of its analogues, *Peptides*, 1992, **13**, 1233–1237.
- 136 T. J. Gaymes, M. Cebrat, I. Z. Siemion and J. E. Kay, Cyclolinopeptide A (CLA) mediates its immunosuppressive activity through cyclophilin-dependent calcineurin inactivation, *FEBS Lett.*, 1997, **418**, 224–227.
- 137 M. Cebrat, Z. Wieczorek and I. Z. Siemion, Immunosuppressive activity of hymenistatin I, *Peptides*, 1996, **17**, 191–196.
- 138 H. Morita, *et al.*, Solution state conformation of an immunosuppressive cyclic dodecapeptide, cycloleonurinin, *Tetrahedron*, 1997, **53**, 7469–7478.
- 139 H. Morita, Y. S. Yun, K. Takeya, H. Itokawa and K. Yamada, Segetalins B, C and D, three new cyclic peptides from *Vaccaria segetalis*, *Tetrahedron*, 1995, **51**, 6003–6014.
- 140 H. Itokawa, Y. Yun, H. Morita, K. Takeya and K. Yamada, Estrogen-like activity of cyclic peptides from *Vaccaria segetalis* extracts<sup>1</sup>, *Planta Med.*, 1995, **61**, 561–562.
- 141 Y. S. Yun, H. Morita, K. Takeya and H. Itokawa, Cyclic peptides from higher plants. 34. Segetalins G and H, structures and estrogen-like activity of cyclic pentapeptides from *Vaccaria segetalis*, *J. Nat. Prod.*, 1997, **60**, 216–218.
- 142 B. Poojary and S. L. Belagali, Synthesis, characterization and biological evaluation of cyclic peptides: Viscumamide, yunnanin A and evolidine, *ChemInform*, 2006, **37**, 1313–1320.
- 143 P.-H. Chuang, *et al.*, Cyclopeptides with Anti-inflammatory Activity from Seeds of *Annona montana*, *J. Nat. Prod.*, 2008, **71**, 1365–1370.
- 144 X.-G. Zou, *et al.*, Flaxseed orbitides, linusorbs, inhibit LPS-induced THP-1 macrophage inflammation, *RSC Adv.*, 2020, **10**, 22622–22630.
- 145 A. Wele, Y. Zhang, L. Dubost, J.-L. Pousset and B. Bodo, Cyclic peptides from the seeds of *Annona glauca* and *A. cherimola*, *Chem. Pharm. Bull.*, 2006, **54**, 690–692.
- 146 H. Morita, T. Kayashita, M. Shimomura, K. Takeya and H. Itokawa, Cyclic peptides from higher plants. 24. Yunnanin C, a novel cyclic heptapeptide from *Stellaria yunnanensis*, *J. Nat. Prod.*, 1996, **59**, 280–282.
- 147 X.-G. Zou, *et al.*, iCellular uptake of [1–9-N $\alpha$ C]-linusorb B2 and [1–9-N $\alpha$ C]-linusorb B3 isolated from flaxseed, and their antitumor activities in human gastric SGC-7901 cells, *J. Funct. Foods*, 2018, **48**, 692–703.
- 148 X.-G. Zou, *et al.*, [1–9-N $\alpha$ C]-linusorb B2 and [1–9-N $\alpha$ C]-linusorb B3 isolated from flaxseed induce G1 cell cycle arrest on SGC-7901 cells by modulating the AKT/JNK signaling pathway, *J. Funct. Foods*, 2019, **52**, 332–339.
- 149 N. Y. Sung, *et al.*, The anti-cancer effect of linusorb B3 from flaxseed oil through the promotion of apoptosis, inhibition of actin polymerization, and suppression of Src activity in glioblastoma cells, *Molecules*, 2020, **25**, 5881.
- 150 H. Morita, *et al.*, Structure of a new cyclic nonapeptide, segetalin F, and vasorelaxant activity of segetalins from *Vaccaria segetalis*, *Bioorg. Med. Chem. Lett.*, 2006, **16**, 4458–4461.
- 151 M. E. F. Pinto, *et al.*, Ribifolin, an orbitide from *Jatropha ribifolia*, and its potential antimalarial activity, *J. Nat. Prod.*, 2015, **78**, 374–380.
- 152 S. Lockett, *et al.*, High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds, *J. Mol. Biol.*, 1999, **290**, 525–533.
- 153 A. S. Jayasena, *et al.*, Next generation sequencing and de novo transcriptomics to study gene evolution, *Plant Methods*, 2014, **10**, 34.
- 154 S. J. Veer, A. M. White and D. J. Craik, Sunflower trypsin inhibitor-1 (SFTI-1): Sowing seeds in the fields of chemistry and biology, *Angew Chem. Int. Ed. Engl.*, 2021, **60**, 8050–8071.
- 155 A. Gitlin-Domagalska, A. Maciejewska and D. Dębowski, Bowman-Birk inhibitors: Insights into family of multifunctional proteins and peptides with potential therapeutical applications, *Pharmaceuticals*, 2020, **13**, 421.
- 156 B. Franke, *et al.*, Two proteins for the price of one: Structural studies of the dual-destiny protein preproalbumin with sunflower trypsin inhibitor-1, *J. Biol. Chem.*, 2017, **292**, 12398–12411.
- 157 P. R. Shewry and M. J. Pandya, The 2S albumin storage proteins, in *Seed Proteins*, Springer, Netherlands, 1999, pp. 563–586.
- 158 J. S. Mylne, I. Hara-Nishimura and K. J. Rosengren, Seed storage albumins: biosynthesis, trafficking and structures, *Funct. Plant Biol.*, 2014, **41**, 671.
- 159 K. Bernath-Levin, *et al.*, Peptide macrocyclization by a bifunctional endoprotease, *Chem. Biol.*, 2015, **22**, 571–582.
- 160 J. Haywood, *et al.*, Structural basis of ribosomal peptide macrocyclization in plants, *Elife*, 2018, **7**, e32955.
- 161 U. C. Marx, *et al.*, Enzymatic cyclization of a potent bowman-Birk protease inhibitor, sunflower trypsin inhibitor-1, and solution structure of an acyclic precursor peptide, *J. Biol. Chem.*, 2003, **278**, 21782–21789.
- 162 W. F. Broekaert, *et al.*, Antimicrobial peptides from plants, *CRC Crit. Rev. Plant Sci.*, 1997, **16**, 297–323.
- 163 A. A. Slavokhotova, A. A. Shelenkov, Y. A. Andreev and T. I. Odintsova, Hevein-like antimicrobial peptides of plants, *Biochemistry*, 2017, **82**, 1659–1674.
- 164 J. R. Kintzing and J. R. Cochran, Engineered knottin peptides as diagnostics, therapeutics, and drug delivery vehicles, *Curr. Opin. Chem. Biol.*, 2016, **34**, 143–150.



- 165 G. M. Clore, A. M. Gronenborn, M. Nilges and C. A. Ryan, Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics, *Biochemistry*, 1987, **26**, 8012–8023.
- 166 D. Le Nguyen, *et al.*, Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure, *Biochimie*, 1990, **72**, 431–435.
- 167 G. M. Clore, *et al.*, The three-dimensional structure of  $\alpha$ -purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics, *EMBO J.*, 1986, **5**, 2729–2735.
- 168 M. Bruix, *et al.*, Solution structure of  $\gamma$ .1-H and  $\gamma$ .1-P thionins from barley and wheat endosperm determined by proton NMR: a structural motif common to toxic arthropod proteins, *Biochemistry*, 1993, **32**, 715–724.
- 169 H. Yeung, *et al.*, Radiation damage and racemic protein crystallography reveal the unique structure of the GASA/snakin protein superfamily, *Angew Chem. Int. Ed. Engl.*, 2016, **55**, 7930–7933.
- 170 J. Van Parijs, W. F. Broekaert, I. J. Goldstein and W. J. Peumans, Hevein: an antifungal protein from rubber-tree (*Hevea brasiliensis*) latex, *Planta*, 1991, **183**, 258–264.
- 171 A. Rodríguez-Romero, K. G. Ravichandran and M. Soriano-García, Crystal structure of hevein at 2.8 Å resolution, *FEBS Lett.*, 1991, **291**, 307–309.
- 172 S. B. Nolde, *et al.*, Disulfide-stabilized helical hairpin structure and activity of a novel antifungal peptide EcAMP1 from seeds of barnyard grass (*Echinochloa crus-galli*), *J. Biol. Chem.*, 2011, **286**, 25145–25153.
- 173 K. H. Wong, *et al.*, Ginkgotides: Proline-Rich Hevein-Like Peptides from Gymnosperm *Ginkgo biloba*, *Front. Plant Sci.*, 2016, **7**, 1639.
- 174 G. Kumari, *et al.*, Molecular diversity and function of jasmintides from *Jasminum sambac*, *BMC Plant Biol.*, 2018, **18**, 144.
- 175 W. L. Tan, *et al.*, Lybatides from *Lycium barbarum* contain an unusual cystine-stapled helical peptide scaffold, *Sci. Rep.*, 2017, **7**, 5194.
- 176 N. Isozumi, *et al.*, Structure and antimicrobial activity of NCR169, a nodule-specific cysteine-rich peptide of *Medicago truncatula*, *Sci. Rep.*, 2021, **11**, 9923.
- 177 J.-C. Kader, Lipid-transfer proteins in plants, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 1996, **47**, 627–654.
- 178 P. K. Pallaghy, K. J. Nielsen, D. J. Craik and R. S. Norton, A common structural motif incorporating a cystine knot and a triple-stranded beta-sheet in toxic and inhibitory polypeptides, *Protein Sci.*, 1994, **3**, 1833–1839.
- 179 C. A. Ryan, Inhibition of carboxypeptidase A by a naturally occurring polypeptide from potatoes, *Biochem. Biophys. Res. Commun.*, 1971, **44**, 1265–1270.
- 180 R. Krätzner, *et al.*, Structure of *Ecballium elaterium* trypsin inhibitor II (EETI-II): a rigid molecular scaffold, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2005, **61**, 1255–1262.
- 181 L. Jouvensal, *et al.*, PA1b, an insecticidal protein extracted from pea seeds (*Pisum sativum*):  $^1\text{H}$ - $^2\text{D}$  NMR study and molecular modeling, *Biochemistry*, 2003, **42**, 11915–11923.
- 182 A. G. Poth, *et al.*, Cyclotides associate with leaf vasculature and are the products of a novel precursor in petunia (*Solanaceae*), *J. Biol. Chem.*, 2012, **287**, 27033–27046.
- 183 Y. Watanabe, *et al.*, A peptide that stimulates phosphorylation of the plant insulin-binding protein. Isolation, primary structure and cDNA cloning, *Eur. J. Biochem.*, 1994, **224**, 167–172.
- 184 J. Huang, *et al.*, Astratides: Insulin-Modulating, Insecticidal, and Antifungal Cysteine-Rich Peptides from *Astragalus membranaceus*, *J. Nat. Prod.*, 2019, **82**, 194–204.
- 185 P. Q. T. Nguyen, *et al.*, Antiviral Cystine Knot  $\alpha$ -Amylase Inhibitors from *Alstonia scholaris*, *J. Biol. Chem.*, 2015, **290**, 31138–31150.
- 186 G. K. T. Nguyen, W. H. Lim, P. Q. T. Nguyen and J. P. Tam, Novel cyclotides and uncyclotides with highly shortened precursors from *Chassalia chartacea* and effects of methionine oxidation on bioactivities, *J. Biol. Chem.*, 2012, **287**, 17598–17607.
- 187 G. K. T. Nguyen, *et al.*, Discovery of a linear cyclotide from the bracelet subfamily and its disulfide mapping by top-down mass spectrometry, *J. Biol. Chem.*, 2011, **286**, 44833–44844.
- 188 G. K. T. Nguyen, *et al.*, Discovery of linear cyclotides in monocot plant *Panicum laxum* of Poaceae family provides new insights into evolution and distribution of cyclotides in plants, *J. Biol. Chem.*, 2013, **288**, 3370–3380.
- 189 H. Bohlmann and K. Apel, Thionins, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 1991, **42**, 227–240.
- 190 K. Hông, T. Austerlitz, T. Bohlmann and H. Bohlmann, The thionin family of antimicrobial peptides, *PLoS One*, 2021, **16**, e0254549.
- 191 A. Castagnaro, C. Marañá, P. Carbonero and F. García-Olmedo, Extreme divergence of a novel wheat thionin generated by a mutational burst specifically affecting the mature protein domain of the precursor, *J. Mol. Biol.*, 1992, **224**, 1003–1009.
- 192 A. K. Balls, A crystalline sulphur-protein from wheat, *J. Wash. Acad. Sci.*, 1942, **32**, 132–137.
- 193 U. Rao, B. Stec and M. M. Teeter, Refinement of purothionins reveals solute particles important for lattice formation and toxicity. Part 1:  $\alpha$ -purothionin revisited, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1995, **51**, 904–913.
- 194 C. H. Vanetten, H. C. Nielsen and J. E. Peters, A crystalline polypeptide from the seed of *Crambe abyssinica*, *Phytochemistry*, 1965, **4**, 467–473.
- 195 M. M. Teeter, J. A. Mazer and J. J. L'Italien, Primary structure of the hydrophobic plant protein crambin, *Biochemistry*, 1981, **20**, 5437–5443.
- 196 A. Yamano, N.-H. Heo and M. M. Teeter, Crystal structure of Ser-22/ile-25 form crambin confirms solvent, side chain substate correlations, *J. Biol. Chem.*, 1997, **272**, 9597–9600.
- 197 M. Fujimura, M. Ideguchi, Y. Minami, K. Watanabe and K. Tadera, Purification, characterization, and sequencing



- of novel antimicrobial peptides, *Tu*-AMP 1 and *Tu*-AMP 2, from bulbs of tulip (*Tulipa gesneriana* L.), *Biosci., Biotechnol., Biochem.*, 2004, **68**, 571–577.
- 198 A. G. Milbradt, F. Kerek, L. Moroder and C. Renner, Structural characterization of helletionins from *Helleborus purpurascens*, *Biochemistry*, 2003, **42**, 2404–2411.
- 199 G. Samuelsson, L. Seger and T. Olson, The amino acid sequence of oxidized viscotoxin A3 from the European mistletoe (*Viscum album* L, Loranthaceae), *Acta Chem. Scand.*, 1968, **22**, 2624–2642.
- 200 G. Schrader and K. Apel, Isolation and characterization of cDNAs encoding viscotoxins of mistletoe (*Viscum album*), *Eur. J. Biochem.*, 1991, **198**, 549–553.
- 201 S. T. Mellstrand and G. Samuelsson, Phoratoxin, a Toxic Protein from the Mistletoe *Phoradendron tomentosum* subsp. *macrophyllum* (Loranthaceae). Improvements in the Isolation Procedure and Further Studies on the Properties, *Eur. J. Biochem.*, 1973, **32**, 143–147.
- 202 S.-S. Li, *et al.*, Ligatoxin B, a new cytotoxic protein with a novel helix–turn–helix DNA-binding domain from the mistletoe *Phoradendron liga*, *Biochem. J.*, 2002, **366**, 405–413.
- 203 T. Shafee and M. A. Anderson, A quantitative map of protein sequence space for the cis-defensin superfamily, *Bioinformatics*, 2019, **35**, 743–752.
- 204 D. Ming and G. Hellekant, Brazzein, a new high-potency thermostable sweet protein from *Pentadiplandra brazzeana* B, *FEBS Lett.*, 1994, **355**, 106–108.
- 205 V. S. De-Paula, *et al.*, Evolutionary relationship between defensins in the Poaceae family strengthened by the characterization of new sugarcane defensins, *Plant Mol. Biol.*, 2008, **68**, 321–335.
- 206 R. d. O. Dias and O. L. Franco, Cysteine-stabilized  $\alpha\beta$  defensins: From a common fold to antibacterial activity, *Peptides*, 2015, **72**, 64–72.
- 207 K. Thevissen, *et al.*, DmAMP1, an antifungal plant defensin from dahlia (*Dahlia merckii*), interacts with sphingolipids from *Saccharomyces cerevisiae*, *FEMS Microbiol. Lett.*, 2003, **226**, 169–173.
- 208 F. Fant, W. Vranken, W. Broekaert and F. Borremans, Determination of the three-dimensional solution structure of *Raphanus sativus* Antifungal Protein 1 by 1H NMR 1 Edited by P. E. Wright, *J. Mol. Biol.*, 1998, **279**, 257–270.
- 209 F. R. Terras, *et al.*, Analysis of two novel classes of plant antifungal proteins from radish (*Raphanus sativus* L.) seeds, *J. Biol. Chem.*, 1992, **267**, 15301–15309.
- 210 Q. Zhao, Y. K. Chae and J. L. Markley, NMR solution structure of ATT<sub>p</sub>, an *Arabidopsis thaliana* trypsin inhibitor, *Biochemistry*, 2002, **41**, 12284–12296.
- 211 K. Vriens, *et al.*, The antifungal plant defensin AtPDF2.3 from *Arabidopsis thaliana* blocks potassium channels, *Sci. Rep.*, 2016, **6**, 32121.
- 212 A. de Beer and M. A. Vivier, Four plant defensins from an indigenous South African Brassicaceae species display divergent activities against two test pathogens despite high sequence similarity in the encoding genes, *BMC Res. Notes*, 2011, **4**, 459.
- 213 A. A. Slavokhotova, *et al.*, Isolation, molecular cloning and antimicrobial activity of novel defensins from common chickweed (*Stellaria media* L.) seeds, *Biochimie*, 2011, **93**, 450–456.
- 214 Y.-J. Liu, *et al.*, Solution structure of the plant defensin VrD1 from mung bean and its possible role in insecticidal activity against bruchids, *Proteins*, 2006, **63**, 777–786.
- 215 R. Pinheiro-Aguiar, V. S. G. do Amaral, I. B. Pereira, E. Kurtenbach and F. C. L. Almeida, Nuclear magnetic resonance solution structure of *Pisum sativum* defensin 2 provides evidence for the presence of hydrophobic surface-clusters, *Proteins*, 2020, **88**, 242–246.
- 216 Z. O. Shenkarev, *et al.*, Heterologous expression and solution structure of defensin from lentil *Lens culinaris*, *Biochem. Biophys. Res. Commun.*, 2014, **451**, 252–257.
- 217 U. S. Sagaram, *et al.*, Structural and functional studies of a phosphatidic acid-binding antifungal plant defensin MtDef4: Identification of an RGFRRR motif governing fungal cell entry, *PLoS One*, 2013, **8**, e82485.
- 218 M. S. Almeida, K. M. S. Cabral, E. Kurtenbach, F. C. L. Almeida and A. P. Valente, Solution structure of *Pisum sativum* defensin 1 by high resolution NMR: plant defensins, identical backbone with different mechanisms of action 1 Edited by M. F. Summers, *J. Mol. Biol.*, 2002, **315**, 749–757.
- 219 H. Li, S. L. S. Velivelli and D. M. Shah, Antifungal potency and modes of action of a novel Olive tree defensin against closely related Ascomycete fungal pathogens, *Mol. Plant Microbe Interact.*, 2019, **32**, 1649–1664.
- 220 V. Kovaleva, H. Krynytskyy, I. Gout and R. Gout, Recombinant expression, affinity purification and functional characterization of Scots pine defensin 1, *Appl. Microbiol. Biotechnol.*, 2011, **89**, 1093–1101.
- 221 C. Bloch and M. Richardson, A new family of small (5 kDa) protein inhibitors of insect  $\alpha$ -amylases from seeds or sorghum (*Sorghum bicolor* (L.) Moench) have sequence homologies with wheat  $\gamma$ -purothionins, *FEBS Lett.*, 1991, **279**, 101–104.
- 222 T. I. Odintsova, *et al.*, Seed defensins of barnyard grass *Echinochloa crusgalli* (L.) Beauv, *Biochimie*, 2008, **90**, 1667–1673.
- 223 S. Tantong, *et al.*, Two novel antimicrobial defensins from rice identified by gene coexpression network analyses, *Peptides*, 2016, **84**, 7–16.
- 224 E. A. Rogozhin, *et al.*, Novel antifungal defensins from *Nigella sativa* L. seeds, *Plant Physiol. Biochem.*, 2011, **49**, 131–137.
- 225 F. Fant, W. F. Vranken and F. A. M. Borremans, The three-dimensional solution structure of *Aesculus hippocastanum* antimicrobial protein 1 determined by 1H nuclear magnetic resonance, *Proteins*, 1999, **37**, 388–403.
- 226 A. M. Aerts, *et al.*, The Antifungal Plant Defensin HsAFP1 from *Heuchera sanguinea* Induces Apoptosis in *Candida Albicans*, *Front. Microbiol.*, 2011, **2**, 47.



- 227 F. T. Lay, H. J. Schirra, M. J. Scanlon, M. A. Anderson and D. J. Craik, The three-dimensional solution structure of NaD1, a new floral defensin from *Nicotiana glauca* and its application to a homology model of the crop defense protein alfAFP, *J. Mol. Biol.*, 2003, **325**, 175–188.
- 228 B. J. C. Janssen, H. J. Schirra, F. T. Lay, M. A. Anderson and D. J. Craik, Structure of *petunia hybrida* defensin 1, a novel plant defensin with five disulfide bonds, *Biochemistry*, 2003, **42**, 8214–8222.
- 229 A. de Beer and M. A. Vivier, Vv-AMP1, a ripening induced peptide from *Vitis vinifera* shows strong antifungal activity, *BMC Plant Biol.*, 2008, **8**, 75.
- 230 K. K. Nielsen, J. E. Nielsen, S. M. Madrid and J. D. Mikkelsen, Characterization of a new antifungal chitin-binding peptide from sugar beet leaves, *Plant Physiol.*, 1997, **113**, 83–91.
- 231 S. Loo, *et al.*, Anti-fungal hevein-like peptides biosynthesized from quinoa cleavable holiolectins, *Molecules*, 2021, **26**, 5909.
- 232 S. G. Kini, *et al.*, Studies on the chitin binding property of novel cysteine-rich peptides from *Alternanthera sessilis*, *Biochemistry*, 2015, **54**, 6639–6649.
- 233 J. C. Martins, *et al.*, 1H NMR study of the solution structure of ac-AMP2, a sugar binding antimicrobial protein isolated from *Amaranthus caudatus*, *J. Mol. Biol.*, 1996, **258**, 322–333.
- 234 W. F. Broekaert, *et al.*, Antimicrobial peptides from *Amaranthus caudatus* seeds with sequence homology to the cysteine/glycine-rich domain of chitin-binding proteins, *Biochemistry*, 1992, **31**, 4308–4314.
- 235 A. Lipkin, *et al.*, An antimicrobial peptide Ar-AMP from amaranth (*Amaranthus retroflexus* L.) seeds, *Phytochemistry*, 2005, **66**, 2426–2431.
- 236 A. A. Astafieva, *et al.*, Discovery of novel antimicrobial peptides with unusual cysteine motifs in dandelion *Taraxacum officinale* Wigg. flowers, *Peptides*, 2012, **36**, 266–271.
- 237 S. Loo, A. Kam, T. Xiao and J. P. Tam, Bleogens: Cactus-derived anti-candida cysteine-rich peptides with three different precursor arrangements, *Front. Plant Sci.*, 2017, **8**, 2162.
- 238 E. A. Rogozhin, *et al.*, A novel antifungal peptide from leaves of the weed *Stellaria media* L., *Biochimie*, 2015, **116**, 125–132.
- 239 K. H. Wong, *et al.*, Vaccatides: Antifungal Glutamine-Rich Hevein-Like Peptides from *Vaccaria hispanica*, *Front. Plant Sci.*, 2017, **8**, 1100.
- 240 K. P. B. Van den Bergh, *et al.*, Five disulfide bridges stabilize a hevein-type antimicrobial peptide from the bark of spindle tree (*Euonymus europaeus* L.), *FEBS Lett.*, 2002, **530**, 181–185.
- 241 J. C. Koo, *et al.*, Two hevein homologs isolated from the seed of *Pharbitis nil* L. exhibit potent antifungal activity, *Biochim. Biophys. Acta*, 1998, **1382**, 80–90.
- 242 S. Yokoyama, *et al.*, Purification, characterization, and sequencing of antimicrobial peptides, Cy-AMP1, Cy-AMP2, and Cy-AMP3, from the Cycad (*Cycas revoluta*) seeds, *Peptides*, 2008, **29**, 2110–2117.
- 243 Y. Xiang, R.-H. Huang, X.-Z. Liu, Y. Zhang and D.-C. Wang, Crystal structure of a novel antifungal protein distinct with five disulfide bridges from *Eucommia ulmoides* Oliver at an atomic resolution, *J. Struct. Biol.*, 2004, **148**, 86–97.
- 244 S. Loo, *et al.*, Identification and characterization of roseltide, a knottin-type neutrophil elastase inhibitor derived from *Hibiscus sabdariffa*, *Sci. Rep.*, 2016, **6**, 39401.
- 245 A. Kam, *et al.*, Roseltide rT7 is a disulfide-rich, anionic, and cell-penetrating peptide that inhibits proteasomal degradation, *J. Biol. Chem.*, 2019, **294**, 19604–19615.
- 246 B. P. Cammue, *et al.*, Isolation and characterization of a novel class of plant antimicrobial peptides from *Mirabilis jalapa* L. seeds, *J. Biol. Chem.*, 1992, **267**, 2228–2233.
- 247 G.-H. Gao, *et al.*, Solution structure of PAFP-S: A new knottin-type antifungal peptide from the seeds of *Phytolacca americana*, *Biochemistry*, 2001, **40**, 10973–10978.
- 248 S.-S. Li and P. Claeson, Cys/Gly-rich proteins with a putative single chitin-binding domain from oat (*Avena sativa*) seeds, *Phytochemistry*, 2003, **63**, 249–255.
- 249 T. I. Odintsova, *et al.*, A novel antifungal hevein-type peptide from *Triticum kiharae* seeds with a unique 10-cysteine motif, *FEBS J.*, 2009, **276**, 4266–4275.
- 250 S. Loo, S. V. Tay, A. Kam, W. Lee and J. P. Tam, Holiolectin interdomain linker determines asparaginyl endopeptidase-mediated maturation of antifungal hevein-like peptides in oats, *Front. Plant Sci.*, 2022, **13**, 899740.
- 251 M. Fujimura, Y. Minami, K. Watanabe and K. Tadera, Purification, characterization, and sequencing of a novel type of antimicrobial peptides, *fa*-AMP1 and *fa*-AMP2, from seeds of buckwheat (*Fagopyrum esculentum* Moench.), *Biosci., Biotechnol., Biochem.*, 2003, **67**, 1636–1642.
- 252 K. H. Wong, W. L. Tan, T. Xiao and J. P. Tam,  $\beta$ -Ginkgotides: Hyperdisulfide-constrained peptides from *Ginkgo biloba*, *Sci. Rep.*, 2017, **7**, 6140.
- 253 M. He, *et al.*, Discovery of a cysteine-rich peptide with glycation modification from *Achyranthes bidentata* Blume, *Fitoterapia*, 2022, **163**, 105338.
- 254 J. P. Marcus, J. L. Green, K. C. Goulter and J. M. Manners, A family of antimicrobial peptides is produced by processing of a 7S globulin protein in *Macadamia integrifolia* kernels, *Plant J.*, 1999, **19**, 699–710.
- 255 J. P. Duvick, T. Rood, A. G. Rao and D. R. Marshak, Purification and characterization of a novel antimicrobial peptide from maize (*Zea mays* L.) kernels, *J. Biol. Chem.*, 1992, **267**, 18814–18820.
- 256 L. Campbell and S. R. Turner, A comprehensive analysis of RALF proteins in green plants suggests there are two distinct functional groups, *Front. Plant Sci.*, 2017, **8**, 37.
- 257 B. Svensson, *et al.*, Primary structure of barwin: a barley seed protein closely related to the C-terminal domain of proteins encoded by wound-induced plant genes, *Biochemistry*, 1992, **31**, 8767–8770.



- 258 J. L. Arolas, F. X. Aviles, J.-Y. Chang and S. Ventura, Folding of small disulfide-rich proteins: clarifying the puzzle, *Trends Biochem. Sci.*, 2006, **31**, 292–301.
- 259 G. Venhudová, F. Canals, E. Querol and F. X. Aviles, Mutations in the N- and C-terminal tails of potato carboxypeptidase inhibitor influence its oxidative refolding process at the reshuffling stage, *J. Biol. Chem.*, 2001, **276**, 11683–11690.
- 260 D. Le-Nguyen, A. Heitz, L. Chiche, M. E. Hajji and B. Castro, Characterization and 2D NMR study of the stable [9-21, 15-27] 2 disulfide intermediate in the folding of the 3 disulfide trypsin inhibitor EETI II, *Protein Sci.*, 1993, **2**, 165–174.
- 261 W.-J. He, *et al.*, Novel inhibitor cystine knot peptides from *Momordica charantia*, *PLoS One*, 2013, **8**, e75334.
- 262 S. Laps, *et al.*, Insight on the order of regioselective ultrafast formation of disulfide bonds in (antimicrobial) peptides and miniproteins, *Angew Chem. Int. Ed. Engl.*, 2021, **60**, 24137–24143.
- 263 A. Romero, J. M. Alamillo and F. García-Olmedo, Processing of thionin precursors in barley leaves by a vacuolar proteinase, *Eur. J. Biochem.*, 1997, **243**, 202–208.
- 264 S. Plattner, *et al.*, Isolation and characterization of a thionin proprotein-processing enzyme from barley, *J. Biol. Chem.*, 2015, **290**, 18056–18067.
- 265 U. Reimann-Philipp, G. Schrader, E. Martinoia, V. Barkholt and K. Apel, Intracellular thionins of barley. A second group of leaf thionins closely related to but distinct from cell wall-bound thionins, *J. Biol. Chem.*, 1989, **264**, 8978–8984.
- 266 H. Bohlmann, *et al.*, Leaf-specific thionins of barley—a novel class of cell wall proteins toxic to plant-pathogenic fungi and possibly involved in the defence mechanism of plants, *EMBO J.*, 1988, **7**, 1559–1565.
- 267 J. Quilis, *et al.*, A potato carboxypeptidase inhibitor gene provides pathogen resistance in transgenic rice, *Plant Biotechnol. J.*, 2007, **5**, 537–553.
- 268 A. Chagolla-Lopez, A. Blanco-Labra, A. Patthy, R. Sánchez and S. Pongor, A novel alpha-amylase inhibitor from amaranth (*Amaranthus hypocondriacus*) seeds, *J. Biol. Chem.*, 1994, **269**, 23675–23680.
- 269 S. Lu, *et al.*, Solution structure of the major alpha-amylase inhibitor of the crop plant amaranth, *J. Biol. Chem.*, 1999, **274**, 20473–20478.
- 270 P. J. B. Pereira, *et al.*, Specific inhibition of insect  $\alpha$ -amylases: yellow meal worm  $\alpha$ -amylase in complex with the Amaranth  $\alpha$ -amylase inhibitor at 2.0 Å resolution, *Structure*, 1999, **7**, 1079–1088.
- 271 A. A. Slavokhotova, *et al.*, Novel mode of action of plant defense peptides - hevein-like antimicrobial peptides from wheat inhibit fungal metalloproteases, *FEBS J.*, 2014, **281**, 4754–4764.
- 272 D. E. A. Florack, B. Visser, P. M. Vries, J. W. L. Vuurde and W. J. Stiekema, Analysis of the toxicity of purothionins and hordothionins for plant pathogenic bacteria, *Neth. J. Plant Pathol.*, 1993, **99**, 259–268.
- 273 A. Molina, P. A. Goy, A. Fraile, R. Sánchez-Monge and F. García-Olmedo, Inhibition of bacterial and fungal plant pathogens by thionins of types I and II, *Plant Sci.*, 1993, **92**, 169–177.
- 274 S. Ohtani, T. Okada, H. Yoshizumi and H. Kagamiyama, Complete primary structures of two subunits of purothionin A, a lethal protein for brewer's yeast from wheat flour, *J. Biochem.*, 1977, **82**, 753–767.
- 275 B. Stec, *et al.*, Proposal for molecular mechanism of thionins deduced from physico-chemical studies of plant toxins, *J. Pept. Res.*, 2004, **64**, 210–224.
- 276 A. G. Gao, *et al.*, Fungal pathogen protection in potato by expression of a plant defensin peptide, *Nat. Biotechnol.*, 2000, **18**, 1307–1310.
- 277 S. Vijayan, N. K. Singh, P. Shukla and P. B. Kirti, Defensin (TvD1) from *Tephrosia villosa* exhibited strong anti-insect and anti-fungal activities in transgenic tobacco plants, *J. Pest Sci.*, 2013, **86**, 337–344.
- 278 S. L. S. Velivelli, K. T. Islam, E. Hobson and D. M. Shah, Modes of action of a bi-domain plant defensin MtDef5 against a bacterial pathogen *Xanthomonas campestris*, *Front. Microbiol.*, 2018, **9**, 934.
- 279 P. B. Pelegrini, F. T. Lay, A. M. Murad, M. A. Anderson and O. L. Franco, Novel insights on the mechanism of action of  $\alpha$ -amylase inhibitors from the plant defensin family, *Proteins*, 2008, **73**, 719–729.
- 280 M. Berrocal-Lobo, *et al.*, Snakin-2, an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection, *Plant Physiol.*, 2002, **128**, 951–961.
- 281 V. Herbel and M. Wink, Mode of action and membrane specificity of the antimicrobial peptide snakin-2, *PeerJ*, 2016, **4**, e1987.
- 282 V. Herbel, H. Schäfer and M. Wink, Recombinant production of snakin-2 (an antimicrobial peptide from tomato) in *E. coli* and analysis of its bioactivity, *Molecules*, 2015, **20**, 14889–14901.
- 283 B. Horváth, *et al.*, Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula* dnf7 mutant, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 15232–15237.
- 284 R. Zhang, *et al.*, Nodule-specific cysteine-rich peptide 343 is required for symbiotic nitrogen fixation in *Medicago truncatula*, *Plant Physiol.*, 2023, **193**, 1897–1912.
- 285 B. Alunni and B. Gourion, Terminal bacteroid differentiation in the legume–rhizobium symbiosis: nodule-specific cysteine-rich peptides and beyond, *New Phytol.*, 2016, **211**, 411–417.
- 286 S. Sankari, *et al.*, A haem-sequestering plant peptide promotes iron uptake in symbiotic bacteria, *Nat. Microbiol.*, 2022, **7**, 1453–1465.
- 287 O. S. Lee, *et al.*, Pn-AMPs, the hevein-like proteins from *Pharbitis nil* confers disease resistance against phytopathogenic fungi in tomato, *Lycopersicon esculentum*, *Phytochemistry*, 2003, **62**, 1073–1079.
- 288 J. Choon Koo, *et al.*, *Plant Mol. Biol.*, 2002, **50**, 441–452.
- 289 S. Kanrar, J. C. Venkateswari, P. B. Kirti and V. L. Chopra, Transgenic expression of hevein, the rubber tree lectin, in



- Indian mustard confers protection against *Alternaria brassicae*, *Plant Sci.*, 2002, **162**, 441–448.
- 290 K. Das, K. Datta, S. N. Sarkar and S. K. Datta, Expression of antimicrobial peptide snak-in-1 confers effective protection in rice against sheath blight pathogen, *Rhizoctonia solani*, *Plant Biotechnol. Rep.*, 2021, **15**, 39–54.
- 291 N. Cox, J. R. Kintzing, M. Smith, G. A. Grant and J. R. Cochran, Integrin-targeting knottin peptide-drug conjugates are potent inhibitors of tumor cell proliferation, *Angew Chem. Int. Ed. Engl.*, 2016, **55**, 9894–9897.
- 292 B. Glotzbach, *et al.*, Combinatorial optimization of cystine-knot peptides towards high-affinity inhibitors of human matriptase-1, *PLoS One*, 2013, **8**, e76956.
- 293 E. K. Gilding, *et al.*, Neurotoxic peptides from the venom of the giant Australian stinging tree, *Sci. Adv.*, 2020, **6**, eabb8828.
- 294 S. Jami, *et al.*, Pain-causing stinging nettle toxins target TMEM233 to modulate Nav1.7 function, *Nat. Commun.*, 2023, **14**, 2442.
- 295 Z. Miao, *et al.*, An engineered knottin peptide labeled with 18F for PET imaging of integrin expression, *Bioconjugate Chem.*, 2009, **20**, 2342–2347.
- 296 C. Kaufmann and M. Sauter, Sulfated plant peptide hormones, *J. Exp. Bot.*, 2019, **70**, 4267–4277.
- 297 B. L. Petersen, C. A. MacAlister and P. Ulvskov, Plant protein O-arabinylation, *Front. Plant Sci.*, 2021, **12**, 645219.
- 298 Y. Matsubayashi, Posttranslationally modified small-peptide signals in plants, *Annu. Rev. Plant Biol.*, 2014, **65**, 385–413.
- 299 J. S. Kim, B. W. Jeon and J. Kim, Signaling peptides regulating abiotic stress responses in plants, *Front. Plant Sci.*, 2021, **12**, 704490.
- 300 Y. Matsubayashi, L. Takagi and Y. Sakagami, Phytosulfokine- $\alpha$ , a sulfated pentapeptide, stimulates the proliferation of rice cells by means of specific high- and low-affinity binding sites, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 13357–13362.
- 301 H. Yang, Y. Matsubayashi, K. Nakamura and Y. Sakagami, *Oryza sativa* PSK gene encodes a precursor of phytosulfokine- $\alpha$ , a sulfated peptide growth factor found in plants, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 13560–13565.
- 302 T. Nakayama, *et al.*, A peptide hormone required for Casparian strip diffusion barrier formation in *Arabidopsis* roots, *Science*, 2017, **355**, 284–286.
- 303 V. G. Doblaz, *et al.*, Root diffusion barrier control by a vasculature-derived peptide binding to the SGN3 receptor, *Science*, 2017, **355**, 280–284.
- 304 Y. Matsuzaki, M. Ogawa-Ohnishi, A. Mori and Y. Matsubayashi, Secreted peptide signals required for maintenance of root stem cell niche in *Arabidopsis*, *Science*, 2010, **329**, 1065–1067.
- 305 Y. Amano, H. Tsubouchi, H. Shinohara, M. Ogawa and Y. Matsubayashi, Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 18333–18338.
- 306 Y. Ito, *et al.*, Dodeca-CLE peptides as suppressors of plant stem cell differentiation, *Science*, 2006, **313**, 842–845.
- 307 M. J. Chrispeels, Prolyl hydroxylase in plants, in *Methods in Enzymology*, Academic Press, 1984, vol. 107, pp. 361–369.
- 308 N. Stührwohldt, A. Ehinger, K. Thellmann and A. Schaller, Processing and formation of bioactive CLE40 peptide are controlled by posttranslational proline hydroxylation, *Plant Physiol.*, 2020, **184**, 1573–1584.
- 309 F. Takahashi, *et al.*, A small peptide modulates stomatal control via abscisic acid in long-distance signalling, *Nature*, 2018, **556**, 235–238.
- 310 N. Stührwohldt, E. Bühler, M. Sauter and A. Schaller, Phytosulfokine (PSK) precursor processing by subtilase SBT3.8 and PSK signaling improve drought stress tolerance in *Arabidopsis*, *J. Exp. Bot.*, 2021, **72**, 3427–3440.
- 311 R. Tabata, *et al.*, Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling, *Science*, 2014, **346**, 343–346.
- 312 D. D. Luu, *et al.*, Biosynthesis and secretion of the microbial sulfated peptide RaxX and binding to the rice XA21 immune receptor, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 8525–8534.
- 313 J. Hattori, K. A. Boutilier, M. M. van Lookeren Campagne and B. L. Miki, A conserved BURP domain defines a novel group of plant proteins with unusual primary structures, *Mol. Gen. Genet.*, 1998, **259**, 424–428.
- 314 B. K. Treacy, *et al.*, Bnm1, a Brassica pollen-specific gene, *Plant Mol. Biol.*, 1997, **34**, 603–611.
- 315 L. Van Son, *et al.*, The BURP domain protein AtUSPL1 of *Arabidopsis thaliana* is destined to the protein storage vacuoles and overexpression of the cognate gene distorts seed development, *Plant Mol. Biol.*, 2009, **71**, 319–329.
- 316 K. Yamaguchi-Shinozaki and K. Shinozaki, The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of rd22, a gene responsive to dehydration stress in *Arabidopsis thaliana*, *Mol. Gen. Genet.*, 1993, **238**, 17–25.
- 317 L. Zheng, R. C. Heupel and D. DellaPenna, The beta subunit of tomato fruit polygalacturonase isoenzyme 1: isolation, characterization, and identification of unique structural features, *Plant Cell*, 1992, **4**, 1147–1156.
- 318 L. S. Mydy, D. N. Chigumba and R. D. Kersten, Plant copper metalloenzymes as prospects for new metabolism involving aromatic compounds, *Front. Plant Sci.*, 2021, **12**, 692108.
- 319 M. Ragland and K. M. Soliman, Sali5-4a and Sali3-2. Two genes induced by aluminum in soybean roots, *Plant Physiol.*, 1997, **114**, 395.
- 320 H. Xu, *et al.*, Genome-scale identification of soybean BURP domain-containing genes and their expression under stress treatments, *BMC Plant Biol.*, 2010, **10**, 197.
- 321 H. Xun, *et al.*, Over-expression of GmKR3, a TIR-NBS-LRR type R gene, confers resistance to multiple viruses in soybean, *Plant Mol. Biol.*, 2019, **99**, 95–111.



- 322 T. Chen, *et al.*, Construction of high-density genetic map and identification of a bruchid resistance locus in mung bean (*Vigna radiata* L.), *Front. Genet.*, 2022, **13**, 903267.
- 323 S. Yahara, *et al.*, Cyclic Peptides, Acyclic Diterpene Glycosides and Other Compounds from *Lycium chinense* MILL, *Chem. Pharm. Bull.*, 1993, **41**, 703–709, DOI: [10.1248/cpb.41.703](https://doi.org/10.1248/cpb.41.703).
- 324 H. Morita, N. Yoshida, K. Takeya, H. Itokawa and O. Shirota, Configurational and conformational analyses of a cyclic octapeptide, lyciumin A, from *Lycium chinense* Mill, *Tetrahedron*, 1996, **52**, 2795–2802.
- 325 H. Morita, H. Suzuki and J. Kobayashi, Celogenamide A, a new cyclic peptide from the seeds of *Celosia argentea*, *J. Nat. Prod.*, 2004, **67**, 1628–1630.
- 326 Y. Tang, *et al.*, Expression of a vacuole-localized BURP-domain protein from soybean (SALI3-2) enhances tolerance to cadmium and copper stresses, *PLoS One*, 2014, **9**, e98830.
- 327 L. S. Mydy, *et al.*, An intramolecular macrocyclase in plant ribosomal peptide biosynthesis, *Nat. Chem. Biol.*, 2024, DOI: [10.1038/s41589-024-01552-1](https://doi.org/10.1038/s41589-024-01552-1).
- 328 A. Messerschmidt, 5.11 – Copper Metalloenzymes, in *Comprehensive Natural Products III*, ed. H.-W. Liu and T. P. Begley, Elsevier, 2010, pp. 251–297.
- 329 H. Song, *et al.*, A molecular mechanism for the enzymatic methylation of nitrogen atoms within peptide bonds, *Sci. Adv.*, 2018, **4**, eaat2720.
- 330 R. Tschesche and E. U. Kaufmann, Chapter 4 the cyclopeptide alkaloids, in *Chemistry and Physiology*, Elsevier, 1975, pp. 165–205.
- 331 E. Tuentner, V. Exarchou, S. Apers and L. Pieters, Cyclopeptide alkaloids, *Phytochem. Rev.*, 2017, **16**, 623–637.
- 332 D. C. Gournelis, G. G. Laskaris and R. Verpoorte, Cyclopeptide alkaloids, *Nat. Prod. Rep.*, 1997, **14**, 75–82.
- 333 K. B. Kang, *et al.*, Jubanines F–J, cyclopeptide alkaloids from the roots of *Ziziphus jujuba*, *Phytochemistry*, 2015, **119**, 90–95.
- 334 R. Tschesche and H. Last, Alkaloide aus rhamnaceen, V franganin und frangufolin, zwei weitere peptid-alkaloide aus L, *Tetrahedron Lett.*, 1968, **9**, 2993–2998.
- 335 R. Tschesche, *et al.*, Alkaloide aus Rhamnaceen, XIX1) Mucronin-E,-F,-G und-H sowie Abyssenin-A,-B und-C, weitere 15 gliedrige Cyclopeptidalkaloide, *Justus Liebigs Ann. Chem.*, 1974, 1915–1928.
- 336 P. Retailleau, E. Numbi Wa Ilunga, V. Fontaine, J.-F. Gallard and P. Le Pogam, Clarifying the configuration of pandamine by an extensive spectroscopic reinvestigation of the authentic 1964 sample, *Metabolites*, 2023, **13**, 470.
- 337 H.-Y. Lin, C.-H. Chen, K. C. S. Chen Liu and S.-S. Lee, 14-Membered Cyclopeptides from *Paliurus ramosissimus* and *P. hemsleyanus*, *Helv. Chim. Acta*, 2003, **86**, 127–138.
- 338 S. R. Giacomelli, *et al.*, Cyclic peptide alkaloids from the bark of *Discaria americana*, *Phytochemistry*, 2004, **65**, 933–937.
- 339 J. Han, *et al.*, Cyclopeptide alkaloids from *Ziziphus apetala*, *J. Nat. Prod.*, 2011, **74**, 2571–2575.
- 340 S. Shabani, J. M. White and C. A. Hutton, Total synthesis of the putative structure of asperipin-2a and stereochemical reassignment, *Org. Lett.*, 2020, **22**, 7730–7734.
- 341 Z. Zhang, *et al.*, Jujube metabolome selection determined the edible properties acquired during domestication, *Plant J.*, 2022, **109**, 1116–1133.
- 342 Y. Ma, *et al.*, Sanjoinine A isolated from *Ziziphi spinosi* semen augments pentobarbital-induced sleeping behaviors through the modification of GABA-ergic systems, *Biol. Pharm. Bull.*, 2007, **30**, 1748–1753.
- 343 H. Han, *et al.*, Anxiolytic-like effects of sanjoinine A isolated from *Ziziphi Spinosi* Semen: possible involvement of GABAergic transmission, *Pharmacol., Biochem. Behav.*, 2009, **92**, 206–213.
- 344 G. Trevisan, *et al.*, Antinociceptive effects of 14-membered cyclopeptide alkaloids, *J. Nat. Prod.*, 2009, **72**, 608–612.
- 345 W. A. Kaleem, *et al.*, Antinociceptive activity of cyclopeptide alkaloids isolated from *Ziziphus oxyphylla* Edgew (Rhamnaceae), *Fitoterapia*, 2013, **91**, 154–158.
- 346 K. H. Hwang, Y. N. Han and B. H. Han, Inhibition of calmodulin-dependent Calcium-ATPase and phosphodiesterase by various cyclopeptides and peptide alkaloids from the *Ziziphus* species, *Arch. Pharmacol Res.*, 2001, **24**, 202–206.
- 347 E. Tuentner, *et al.*, Antiplasmodial activity, cytotoxicity and structure-activity relationship study of cyclopeptide alkaloids, *Molecules*, 2017, **22**, 224.
- 348 E. Tuentner, *et al.*, Cyclopeptide alkaloids from *Hymenocardia acida*, *J. Nat. Prod.*, 2016, **79**, 1746–1751.
- 349 F. Sugawara, *et al.*, Insecticidal Peptide from Mungbean: A Resistant Factor against Infestation with Azuki Bean Weevil, *J. Agric. Food Chem.*, 1996, **44**, 3360–3364.
- 350 A. Kaga and M. Ishimoto, Genetic localization of a bruchid resistance gene and its relationship to insecticidal cyclopeptide alkaloids, the vignatic acids, in mungbean (*Vigna radiata* L. Wilczek), *Mol. Gen. Genet.*, 1998, **258**, 378–384.
- 351 K. Yoshikawa, S. Tao and S. Arihara, Stephanotic acid, a novel cyclic pentapeptide from the stem of *Stephanotis floribunda*, *J. Nat. Prod.*, 2000, **63**, 540–542.
- 352 X. Xu, *et al.*, Moroidin, a cyclopeptide from the seeds of *Celosia cristata* that induces apoptosis in A549 human lung cancer cells, *J. Nat. Prod.*, 2022, **85**, 1918–1927.
- 353 H. Suzuki, H. Morita, M. Shiro and J. Kobayashi, Celogentin K, a new cyclic peptide from the seeds of *Celosia argentea* and X-ray structure of moroidin, *Tetrahedron*, 2004, **60**, 2489–2495.
- 354 J. Kobayashi, H. Suzuki, K. Shimbo, K. Takeya and H. Morita, Celogentins A–C, new antimitotic bicyclic peptides from the seeds of *Celosia argentea*, *J. Org. Chem.*, 2001, **66**, 6626–6633.
- 355 H. Suzuki, H. Morita, S. Iwasaki and J. Kobayashi, New antimitotic bicyclic peptides, celogentins D–H, and J, from the seeds of *Celosia argentea*, *Tetrahedron*, 2003, **59**, 5307–5315.
- 356 T.-W. Christina Leung, D. H. Williams, J. C J Barna, S. Foti and B. P. Oelrichs, Structural studies on the peptide



- moroidin from *laportea moroides*, *Tetrahedron*, 1986, **42**, 3333–3348.
- 357 B.-S. Yun, I.-J. Ryoo, I.-K. Lee and I.-D. Yoo, Hibispeptin A, a novel cyclic peptide from *Hibiscus syriacus*, *Tetrahedron Lett.*, 1998, **39**, 993–996.
- 358 B.-S. Yun, I.-J. Ryoo, I.-K. Lee and I.-D. Yoo, Hibispeptin B, a novel cyclic peptide from *Hibiscus syriacus*, *Tetrahedron*, 1998, **54**, 15155–15160.
- 359 M. Toumi, *et al.*, A general route to cyclopeptide alkaloids: Total syntheses and biological evaluation of paliurines E and F, ziziphines N and Q, abyssenine A, mucronine E, and analogues, *Eur. J. Org. Chem.*, 2009, 3368–3386.
- 360 M. A. Abdalla and J. C. Matasyoh, Endophytes as producers of peptides: An overview about the recently discovered peptides from endophytic microbes, *Nat. Prod. Bioprospect.*, 2014, **4**, 257–270.
- 361 T. Schafhauser, *et al.*, Antitumor astins originate from the fungal endophyte *Cyanoderrella asteris* living within the medicinal plant *Aster tataricus*, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 26909–26917.
- 362 S. D. Jolad, *et al.*, Bouvardin and deoxybouvardin, antitumor cyclic hexapeptides from *Bouvardia ternifolia* (Rubiaceae), *J. Am. Chem. Soc.*, 1977, **99**, 8040–8044.
- 363 H. Itokawa, *et al.*, Studies on antitumor cyclic hexapeptides RA obtained from *Rubiae Radix*, Rubiaceae. VI Minor antitumor constituents, *Chem. Pharm. Bull.*, 1986, **34**, 3762–3768.
- 364 H. Itokawa, *et al.*, Studies on the antitumor cyclic hexapeptides obtained from *Rubiae radix*, *Chem. Pharm. Bull.*, 1983, **31**, 1424–1427.
- 365 J.-T. Fan, *et al.*, Rubiyunnanins A and B, two novel cyclic hexapeptides from *Rubia yunnanensis*, *Tetrahedron Lett.*, 2010, **51**, 6810–6813.
- 366 J.-T. Fan, *et al.*, Rubiyunnanins C–H, cytotoxic cyclic hexapeptides from *Rubia yunnanensis* inhibiting nitric oxide production and NF- $\kappa$ B activation, *Bioorg. Med. Chem.*, 2010, **18**, 8226–8234.
- 367 X.-Q. Chen, *et al.*, Rubicordins A–C, new cyclopeptides from *Rubia cordifolia* with cytotoxicity and inhibiting NF- $\kappa$ B signaling pathway, *Tetrahedron*, 2015, **71**, 9673–9678.
- 368 L. Feng, *et al.*, Diversity of cultivable endophytic fungi in two *Rubia* plants and their potential for production of anti-tumour Rubiaceae-type cyclopeptides, *Letts. Appl. Microbiol.*, 2021, **73**, 759–769.
- 369 M. Zalacain, E. Zaera, D. Vázquez and A. Jiménez, The mode of action of the antitumor drug bouvardin, an inhibitor of protein synthesis in eukaryotic cells, *FEBS Lett.*, 1982, **148**, 95–97.
- 370 S. A. Stickel, N. P. Gomes, B. Frederick, D. Raben and T. T. Su, Bouvardin is a radiation modulator with a novel mechanism of action, *Radiat. Res.*, 2015, **184**, 392.
- 371 H. Itokawa, *et al.*, Isolation and antitumor activity of cyclic hexapeptides isolated from *Rubiae radix*, *Chem. Pharm. Bull.*, 1984, **32**, 284–290.
- 372 X.-Y. Fang, *et al.*, Plant cyclopeptide RA-V kills human breast cancer cells by inducing mitochondria-mediated apoptosis through blocking PDK1–AKT interaction, *Toxicol. Appl. Pharmacol.*, 2013, **267**, 95–103.
- 373 H.-W. Leung, *et al.*, RA-XII inhibits tumour growth and metastasis in breast tumour-bearing mice via reducing cell adhesion and invasion and promoting matrix degradation, *Sci. Rep.*, 2015, **5**, 16985.
- 374 G. G. L. Yue, *et al.*, Cyclopeptide RA-V inhibits angiogenesis by down-regulating ERK1/2 phosphorylation in HUVEC and HMEC-1 endothelial cells, *Br. J. Pharmacol.*, 2011, **164**, 1883–1898.
- 375 X. An and F. Shang, RA-XII exerts anti-oxidant and anti-inflammatory activities on lipopolysaccharide-induced acute renal injury by suppressing NF- $\kappa$ B and MAPKs regulated by HO-1/Nrf2 pathway, *Biochem. Biophys. Res. Commun.*, 2018, **495**, 2317–2323.
- 376 G. Yu, D. K. Smith, H. Zhu, Y. Guan and T. T.-Y. Lam, ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data, *Methods Ecol. Evol.*, 2017, **8**, 28–36.
- 377 C. L. Kirkpatrick, *et al.*, The “PepSAVI-MS” pipeline for natural product bioactive peptide discovery, *Anal. Chem.*, 2017, **89**, 1194–1201.
- 378 N. C. Parsley, *et al.*, PepSAVI-MS reveals anticancer and antifungal cycloviolacins in *Viola odorata*, *Phytochemistry*, 2018, **152**, 61–70.
- 379 T. B. Moyer, *et al.*, PepSAVI-MS Reveals a Proline-rich Antimicrobial Peptide in *Amaranthus tricolor*, *J. Nat. Prod.*, 2019, **82**, 2744–2753.
- 380 J. Watrous, *et al.*, Mass spectral molecular networking of living microbial colonies, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E1743–E1752.
- 381 H. Mohimani, *et al.*, Dereplication of microbial metabolites through database search of mass spectra, *Nat. Commun.*, 2018, **9**, 4035.
- 382 J. Y. Yang, *et al.*, Molecular networking as a dereplication strategy, *J. Nat. Prod.*, 2013, **76**, 1686–1699.
- 383 L.-F. Nothias, *et al.*, Bioactivity-based molecular networking for the discovery of drug leads in natural product bioassay-guided fractionation, *J. Nat. Prod.*, 2018, **81**, 758–767.
- 384 M. Wang, *et al.*, Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, *Nat. Biotechnol.*, 2016, **34**, 828–837.
- 385 M. Wang, *et al.*, Mass spectrometry searches using MASST, *Nat. Biotechnol.*, 2020, **38**, 23–26.
- 386 K. Dührkop, *et al.*, SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information, *Nat. Methods*, 2019, **16**, 299–302.
- 387 K. Dührkop, *et al.*, Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra, *Nat. Biotechnol.*, 2020, **39**, 462–471.
- 388 B. L. Archer, The proteins of *Hevea brasiliensis* Latex. 4. Isolation and characterization of crystalline hevein, *Biochem. J.*, 1960, **75**, 236–240.
- 389 G. K. T. Nguyen, *et al.*, Discovery and characterization of novel cyclotides originated from chimeric precursors consisting of albumin-1 chain a and cyclotide domains in the Fabaceae family, *J. Biol. Chem.*, 2011, **286**, 24275–24287.



- 390 D. C. Ireland, R. J. Clark, N. L. Daly and D. J. Craik, Isolation, sequencing, and Structure–Activity relationships of cyclotides, *J. Nat. Prod.*, 2010, **73**, 1610–1622.
- 391 T. B. Moyer, N. C. Parsley, P. W. Sadecki, W. J. Schug and L. M. Hicks, Leveraging orthogonal mass spectrometry based strategies for comprehensive sequencing and characterization of ribosomal antimicrobial peptide natural products, *Nat. Prod. Rep.*, 2021, **38**, 489–509.
- 392 A. Shevchenko, M. Wilm, O. Vorm and M. Mann, Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels, *Anal. Chem.*, 1996, **68**, 850–858.
- 393 N. C. Parsley, O. L. Williams and L. M. Hicks, Exploring the diversity of cysteine-rich natural product peptides via MS/MS fingerprint ions, *J. Am. Soc. Mass Spectrom.*, 2020, **31**, 1833–1843.
- 394 A. K. Jarmusch, *et al.*, A universal language for finding mass spectrometry data patterns, *bioRxiv*, 2022, preprint, DOI: [10.1101/2022.08.06.503000](https://doi.org/10.1101/2022.08.06.503000).
- 395 H. Horai, *et al.*, MassBank: a public repository for sharing mass spectral data for life sciences, *J. Mass Spectrom.*, 2010, **45**, 703–714.
- 396 C. A. Smith, *et al.*, METLIN: a metabolite mass spectral database, *Ther. Drug Monit.*, 2005, **27**, 747–751.
- 397 R. M. Salek, *et al.*, The MetaboLights repository: curation challenges in metabolomics, *Database*, 2013, **2013**, bat029.
- 398 B. Behsaz, *et al.*, De Novo Peptide Sequencing Reveals Many Cyclopeptides in the Human Gut and Other Environments, *Cell Syst.*, 2020, **10**, 99–108.
- 399 N. Matasci, *et al.*, Data access for the 1,000 Plants (1KP) project, *Gigascience*, 2014, **3**, 17.
- 400 One Thousand Plant Transcriptomes Initiative, One thousand plant transcriptomes and the phylogenomics of green plants, *Nature*, 2019, **574**, 679–685.
- 401 D. M. Goodstein, *et al.*, Phytozome: a comparative platform for green plant genomics, *Nucleic Acids Res.*, 2012, **40**, D1178–D1186.
- 402 A. Shelenkov, A. Slavokhotova and T. Odintsova, Predicting antimicrobial and other cysteine-rich peptides in 1267 plant transcriptomes, *Antibiotics*, 2020, **9**, 60.
- 403 F. Teufel, *et al.*, SignalP 6.0 predicts all five types of signal peptides using protein language models, *Nat. Biotechnol.*, 2022, **40**, 1023–1025.
- 404 W. Lau and E. S. Sattely, Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone, *Science*, 2015, **349**, 1224–1228.
- 405 M. A. Schorn, *et al.*, A community resource for paired genomic and metabolomic data mining, *Nat. Chem. Biol.*, 2021, **17**, 363–368.

