ELSEVIER

Peptide-based protease inhibitors from plants

Roland Hellinger and Christian W. Gruber

Center for Pharmacology and Physiology, Medical University of Vienna, Schwarzspanierstr. 17, 1090 Vienna, Austria

Proteases have an important role in homeostasis, and dysregulation of protease function can lead to pathogenesis. Therefore, proteases are promising drug targets in cancer, inflammation, and neurodegenerative disease research. Although there are well-established pharmaceuticals on the market, drug development for proteases is challenging. This is often caused by the limited selectivity of currently available lead compounds. Proteinaceous plant protease inhibitors are a diverse family of (poly)peptides that are important to maintain physiological homeostasis and to serve the innate defense machinery of the plant. In this review, we provide an overview of the diversity of plant peptide- and protein-based protease inhibitors (PIs), provide examples of such compounds that target human proteases, and discuss opportunities for these molecules in protease drug discovery and development.

Introduction

Plants have evolved diverse mechanisms to fend off pathogens, such as viruses, bacteria, or herbivores. These defense strategies often comprise plant chemicals, including secondary metabolite small molecules and gene-derived peptides [1,2]. Most of the reported plant-derived PIs (here, this abbreviation applies solely to protein- or peptide-based protease inhibitors) are organic molecules as well as amino acids or derivatives thereof (reviewed in detail elsewhere [1–4]). However, peptide- and protein-based PIs are abundantly expressed in plants: for instance, PIs in the seeds of legumes account for ~1–10 % of the total soluble protein content, which exemplifies their importance for plant physiology and defense [5]. Plant PIs fulfill two major functions: (i) they prohibit proteolytic degradation of storage proteins in seeds or kernels to enable the controlled mobilization of nutrients in the form of amino acids or small peptides; and (ii) they protect the plant against herbivores, leading to starvation of these pests by inhibition of digestive enzymes or hemolysis of the gastrointestinal tract [5,6].

The most comprehensive classification of PIs was conducted by the Sanger Institute, which launched the MEROPS inhibitor database, covering known inhibitors of microbes, plants, and animals. According to MEROPS, there are 82 PI families that have been classified by sequence homology and subdivided into 39 clans based on secondary and tertiary structure similarities as well as inhibitor function [7,8]. For this review we additionally analyzed public web databases (i.e., www.cybase.org; www.dsimb.inserm.fr/KNOTTIN/; and www.uniprot.org/). Primary scientific literature searches of www.ncbi.nlm.nih.gov/pubmed were conducted using search terms, such as 'plant protease inhibitor' or the PI family names, and no restriction to publication date was set. Overall, there are reportedly >6700 plant-derived proteinaceous PIs, which can be classified by structural similarity or sequence homology into at least 12 distinct families. Few of these have been investigated at the protein level and many PIs have been identified by homology via *in silico* analysis of nucleic acid sequences. Plant PIs are a heterogenous group including proteins (>15 kDa) such as serpins, phytocystatins, Kunitz inhibitors, as well as peptides (<15 kDa), such as Bowman-Birk inhibitors (BBIs), α-amylase-trypsin, mustard-type, potato type-I, potato type-II inhibitors, potato metallocarboxypeptidase inhibitors (MCPIs), squash, and cyclotide inhibitors (Fig. 1a). Overall, 104 plant families have been identified to express PIs, with differences in the diversity (Fig. 1b).

In this review, we summarize the 12 common types of plant PI and provide an overview of their structural characteristics

Corresponding author: Gruber, C.W. (christian.w.gruber@meduniwien.ac.at)

https://doi.org/10.1016/j.drudis.2019.05.026

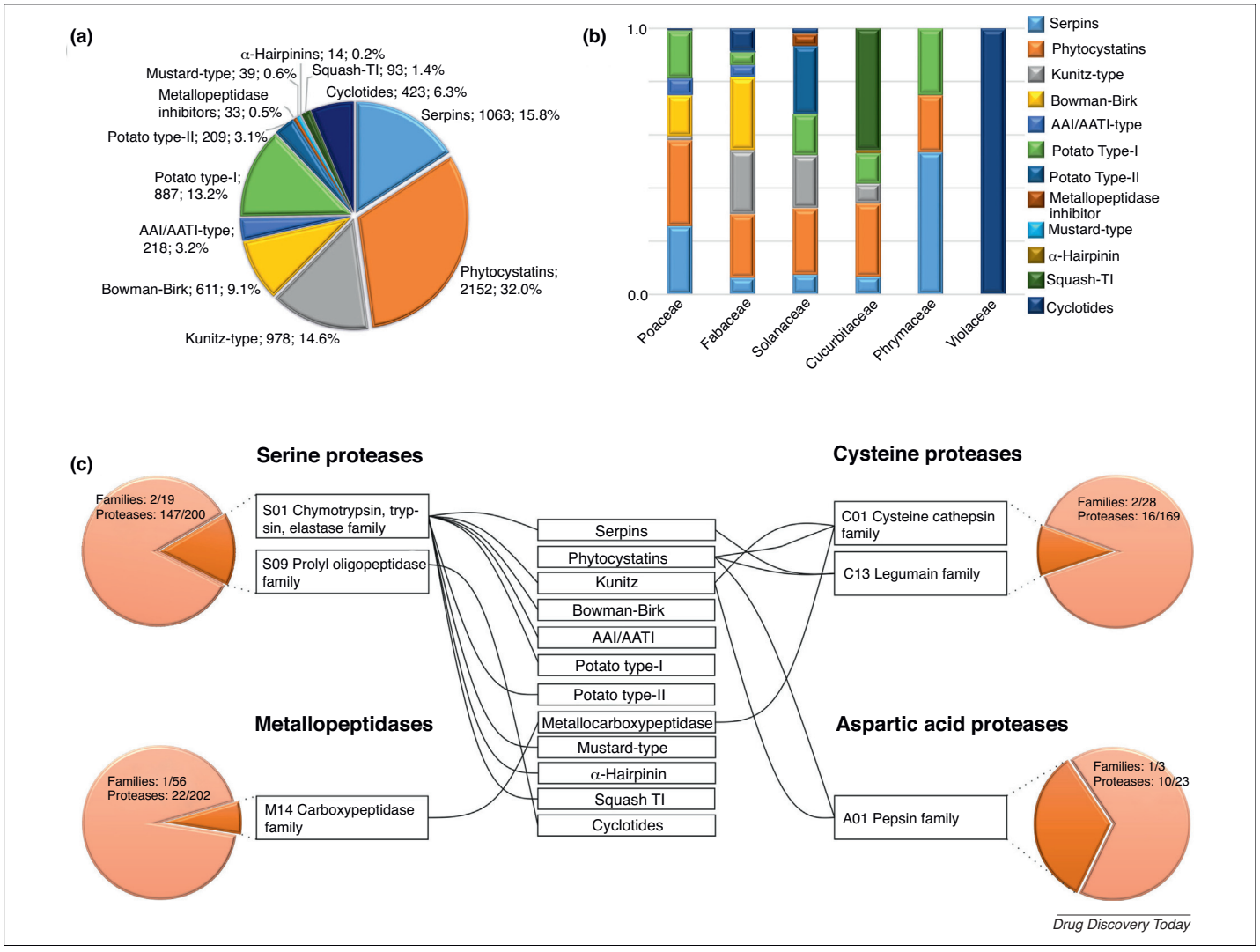

FIGURE 1

Overview of the diversity and distribution of plant protease inhibitors (PIs). **(a)** In total, 6720 inhibitors were identified; the percentage and number of inhibitors for the 12 main plant PI families are illustrated. Proteinaceous plant PIs account for almost half of PIs and peptide-like PIs result in a higher structural diversity. Overall, 81 plant families have been reported to express at least one PI. Importantly, four plant families account for ~66% of all known plant PIs to date: grasses (Poaceae 2139 PIs; 32% of all known PIs), legumes (Fabaceae 933 PIs, 14%), night shade (Solanaceae 732, 10%), and cabbage plants (Brassicaceae 632 PIs, 9%). The most important 22 plant families account for ~93.4% of all recorded plant PIs (these are in addition to those mentioned before: Violaceae 237 PIs, 3.5%; Saliaceae 218, 3.3%; Cucurbitaceae 192, 2.9%; Rosaceae 191, 2.9%; Malvaceae 138, 2.1%; Rubiaceae 120, 1.8%; Rutaceae 117, 1.7%; Asteraceae 116, 1.7%; Euphorbiaceae 93, 1.4%; Prymaceae 71, 1.1%; Amaranthaceae 50, 0.7 %; Moraceae 46, 0.7%; Lythraceae 45, 0.7%; Myrtaceae 44, 0.7%; Vitaceae 38, 0.6%; Musaceae 36, 0.5%; Nelumbonaceae 36, 0.5%; and Cannabaceae 35, 0.5%). The remaining 59 plant families contain ≤33 (~0.5%) inhibitors per plant, which in total accounts for approximately 6.6% of all recorded plant PIs. **(b)** For six plant families, an overview of the (normalized) distribution for the 12 inhibitor classes is shown. Many plant families are a rich source for multiple inhibitor types [e.g., Fabaceae and Solanaceae (seven different classes of PIs) or Poaceae and Cucurbitaceae (six different classes of PIs)]; other plant families exclusively express one to three plant PI types, for instance, violet plants (Violaceae), which are thought to contain only cyclotides, or the lopseed family (Phrymaceae), which contain phytostatins, serpin, and potato-type I inhibitors. Interestingly, Violaceae (cyclotides are the only inhibitors reported so far for this peptide class), Solanaceae (the prime source for potato type-II), and Cucurbitaceae (the major source for squash-TI) show a preference for a single to a few compound classes. **(c)** Plant PIs have been characterized for inhibition of prototypic proteases, including trypsin-, chymotrypsin-, elastase-, aspartic acid-like, and blood coagulation proteases. Representative inhibitors for each plant PI class were used to elucidate the previously tested inhibitory spectrum of the compound class, as indicated by connecting lines. Where applicable, inhibitory activity toward a nonhuman protease was converted to the human protease homolog in this illustration. In total, 200 human serine proteases comprise 19 families with numerous subfamilies, of which the S01 and S09 serine protease families are targeted by plant PIs; 28 families with a total of 169 cysteine proteases are known, of which plant PIs reportedly inhibit enzymes belonging to the C01 and C13 families. The metallocarboxypeptidase family M14 was reported as a protease target for plant PIs (56 families and 202 proteins). The A01 family of aspartic acid proteases is inhibited by plant PIs (three families, including 23 proteases). Human threonin proteases have not been reported as molecular targets for plant-derived PIs. Abbreviations: AAI/AATI, α-amylase and bifunctional trypsin inhibitors.

regarding their protease inhibitory motifs. We also highlight the value of plant peptide PIs as interesting bioactive molecules for drug discovery. We describe plant-derived PIs under investigation as anticancer therapies, in immunopharmacology, as well was targeting proteases involved in neurodegenerative diseases. Finally, we discuss the opportunities for plant-derived PIs in PI drug discovery and development.

Diversity of plant protease inhibitors

Plants produce a range of PIs, including polyphenols, terpenes, flavones, saponins, alkaloids, tannins, amino acids, di- and tripeptides, and derivatives thereof as well as plant peptide- or protein-based PIs [2]. The functionality of the proteinaceous plant PIs usually requires well-structured domains; therefore, most of these molecules incorporate stabilizing motifs [9,10]. Larger proteinogenic plant PIs comprise several domains, including one or several inhibiting units. By contrast, there are many plant PIs with a molecular weight typical for peptides (i.e., 5–100 amino acids). They are commonly referred to as ribosomally synthesized and post-translationally modified peptides [11]. Peptide plant PIs are often molecules with a single inhibiting domain (Fig. 2). These domains include: (i) secondary protein structural elements, which are generally less susceptible to proteolytic degradation compared with unstructured domains (i.e., α-helices or β-sheets); (ii) post-translational modifications (i.e., pyroglutamate to protect the susceptible N-terminus from aminopeptidases); (iii) cysteine-stabilizing motifs (e.g., cystine-knot); or (iv) cyclization (e.g., cyclotides), which protects the peptide termini from carboxypeptidases as well as aminopeptidases [5,8,10,12,13]. Here, we discuss the unique structural diversity of 12 plant PI families.

Serpins

Serpins are found in many higher plants, but rarely in chlorophytes. They are large PIs of ~50 kDa with a common scaffold constructed by three β-sheets and up to nine α-helices [8,10,14] (Fig. 2a). Serpins are referred to as 'suicide inhibitors', because they form a covalent complex with serine or cysteine proteases. The reactive center loop of the inhibitor docks into the catalytic site of the protease. Upon cleavage of the scissile bond (P1-P1'), an acyl ester intermediate is formed between the inhibitor and the γ-oxygen atom of a serine residue of the protease. As a consequence, the serpin reactive center loop adopts a 'superstable' β-sheet conformation, which leads to an irreversible distortion of the protease structure [15]. The main function of serpins is the deactivation of endogenous proteases. Hence, these peptides mainly fulfill a regulatory role in plant homeostasis, but are also described to fend off insects and pathogens by inhibiting their digestive proteases [6].

Phytocystatins

Phytocystatins are plant PIs ranging in molecular weight from 10 kDa to 23 kDa [6,16]. The phytocystatin structure comprises antiparallel β-sheets wrapped around a central α-helix. The inhibitory domain includes two hairpin loops with conserved motifs; these are a central Gln-Xaa-Val-Xaa-Gly (Xaa can be any amino acid) and a Pro/Leu-Trp in the C-terminal region, respectively [16]. They inhibit cysteine proteases in a noncatalytically competent manner (i.e., these inhibitors still block access to the catalytic site, but they do not bind to the protease in a strictly substrate-like manner) [17]. Both hairpin motifs of the inhibitor interact with the active sites of proteases. The N-terminal conserved region does not bind to the active site cleft, but is essential for the tight interaction as well as the specificity of these inhibitors for cysteine proteases (Fig. 2b). Phytocystatins are endogenous regulators of plant proteases and also function *ex planta* to abolish cysteine proteases activity in feeding enemies.

Kunitz-type protease inhibitors

Kunitz-type inhibitors comprise one to several inhibitory domains. Plant Kunitz-type inhibitors generally range in size from 18 kDa to 24 kDa, but a few smaller peptides (~8 kDa) have been also described in plants and animals [18]. The characteristic Kunitz motif is dominated by a universal β-trefoil structure, which is formed by up to 12 antiparallel β-strands organized in a barrel-like domain with up to three disulfide bonds connecting this subdomain [10] (Fig. 2c). They were originally discovered in soybean as potent serine PIs (e.g., of the S01 chymotrypsin superfamily) and for aspartic acid protease inhibition (i.e., human cathepsin D) [6]. Kunitz PI from barely ('Hordeum-type') exhibit α-amylase inhibitory activity. Kunitz inhibitors function as competitive inhibitors and bind to proteases in a classical 'key-to-lock' fashion. In plants, Kunitz-type inhibitors are known as regulators of physiological homeostasis and also function as inhibitors of pathogen proteases.

Bowman-Birk inhibitors

BBIs range in molecular weight from 5 kDa to16 kDa and contain one or two inhibitory domains. An antiparallel β-sheet domain contains two inhibiting loops on opposite sides of the molecule. Seven disulfide bonds function as stabilizing elements for BBI proteins (Fig. 2d). The two-domain structure of the inhibitor can form a stochiometric 1:1:1 complex with two distinct protease molecules [6,10]. They function in a competitive protease inhibition manner following the standard mechanism for substrate-like binding to the catalytic center of the protease. BBI expression and biosynthesis are reportedly strongly inducible by pathogen invasion [19–21].

The sunflower trypsin inhibitor (TI) from *Helianthus annuus* is the smallest known BBI peptide (14-mer). It is a head-to-tail cyclized peptide, stabilized by a single disulfide bond and an intramolecular hydrogen bond network between the antiparallel β-strands [10,22]. The cyclic backbone of the peptide is formed by asparagine endopeptidase, also known as legumain [23]. Given the plasticity and straightforward synthesis of this scaffold, it has been explored as a model peptide for drug design studies [22–24].

α-Amylase inhibitors and bifunctional trypsin inhibitors

In plants, six proteinaceous α-amylase inhibitor (AAI) subtypes are reported: the γ-thionin-like, chloroform-methanol (CM) extractable proteins, knottin-like, Kunitz-type, thaumatin-like, and legume lectin-like inhibitors [25]. Bifunctional α-amylase/TIs (AATI) are 10–15 kDa in size and have been discovered in important starch-containing crop plants, such as barley, amaranth, wheat, and maize (Fig. 2e). These molecules comprise several cystine-spanning loops and several α-helical segments. Their inhibiting loops adopt canonical substrate-like confirmations and, therefore, these peptides function as competitive PIs. α-Amylase inhibitors from the Amaranthaceae and Apocynaceae plant families are small cystine-knot peptides of ~32 residues with antiviral activity [25,26]. α-Amylase inhibitors have a physiological role in the regulation of endogenous α-amylase activity. In addition, they often have high specificity for insect amylases and, hence, are important in herbivore defense [25,27].

Potato-type inhibitors

Potato-type PIs are not exclusive to solanaceous herbs (as the name would suggest) but also occur in other plants [6,28]. Potato type-I

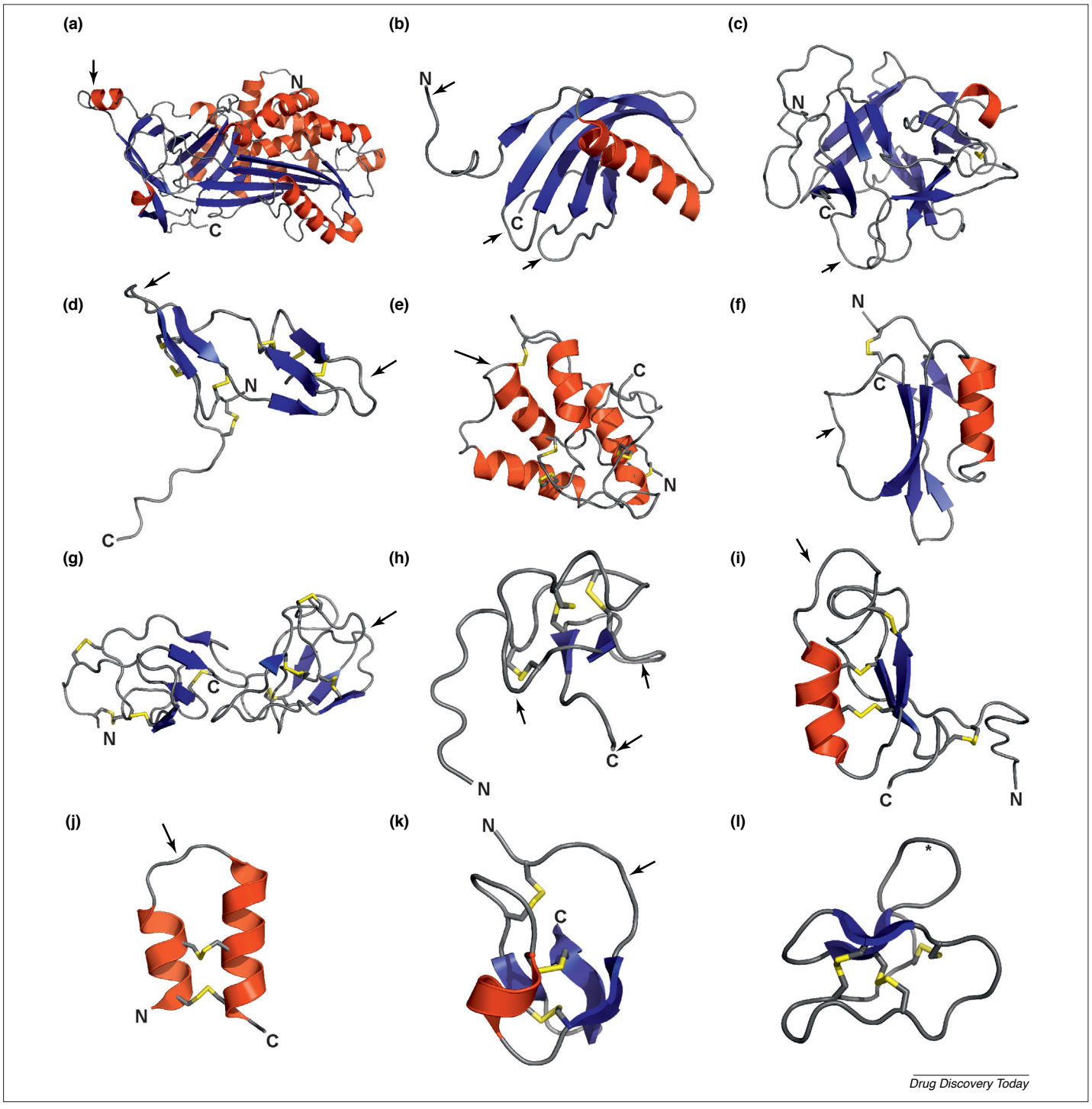

FIGURE 2

Representative structural models of plant protease inhibitors (PIs). **(a)** At-serpin1 from *Arabidopsis thaliana* (PDB code: 3LE2) [63], **(b)** oryzacystatin-I from *Oryza sativa* (1EQK) [68], **(c)** Kunitz inhibitor from *Cicer arietinum* (5XOZ) [117], **(d)** soybean BBI from *Glycine max* (5J4Q) [118], **(e)** bifunctional α-amylase/trypsin inhibitor corn Hageman factor from *Zea mays* (1BEA) [89], **(f)** CMTI-V from *Cucurbita maxima* (1TIN) [93], **(g)** TI-II from *Solanum lycopersicum* (1PJU) [98], **(h)** potato CPI from *Solanum tuberosum* (1H20) [119], **(i)** mustard-TI from *A. thaliana* (1JXC) [120], (j) VcTI (2PLX) from *Veronica hederifolia* [107], **(k)** EETI-II from *Ecballium elaterium* (2IT7) [121], and **(l)** kalata B1 from *Oldenlandia affinis* (1NB1) [122]. Helices are shown in blue, β-sheets in red, disulfide bonds are indicated in yellow in stick representation, and connecting amino acid loops in gray. N- and C-termini are labeled, and the asterisk in (l) indicates the cyclization point of cyclotides. The inhibiting loops or other important sites for protease inhibition are marked with an arrow and explained further in the main text.

inhibitors contain a single polypeptide domain of ~7–9 kDa [10,28]. They consist of four parallel and antiparallel β-sheets, a single α-helix, and a hydrophobic core, which is located opposite to the reactive loop. Potato type-I inhibitors lack any, or contain one disulfide bond. They are classical canonical inhibitors that bind to proteases in a substrate-like manner (Fig. 2f). Potato type-II inhibitors are reported to have two inhibiting domains that are able to adopt a ternary complex simultaneously with two proteases. They have a molecular weight of ~20 kDa (Fig. 2g). Potato type-II inhibitors are generally molecules with little secondary

structure, apart from three antiparallel β-sheets, several turns, and four stabilizing disulfide bonds in each of the two domains. However, when bound to the target protease, they undergo a conformational switch to a more rigid form, which enables tight binding to the active site of the protease (classical competitive inhibition mode).

Metallocarboxypeptidase inhibitors

MCPIs from potato are ~3–4 kDa. Their structure does not comprise typical elements, such as β-sheets or α-helices, but instead they contain several turns, smaller strands, and short 3_{10}-helix type segments. MCPIs are glFig. 2obular peptides comprising a T-knot motif, which is a cystine-knot in the growth-factor configuration (h) [6,10,13]. Potato MCPIs bind to the protease in a substrate-like manner, although the inhibitory segment is located on the C-terminal tail rather than in a stabilized loop or another secondary structural motif. After cleavage of the P1-P1′ bond in the potato carboxypeptidase inhibitor, the newly formed C-tail coordinates with the Zn^{2+} of the carboxypeptidase active site to block the catalytic site of the protease [6]. However, secondary interactions outside of the protease binding cleft contribute to the strength and the specificity of the PI inhibition [10,28].

Mustard-type trypsin inhibitors

Protease inhibitors from Brassicaceae are reported as mustard-type TIs. They are ~7-kDa large polypeptides, which strongly inhibit proteases from the chymotrypsin superfamily and their expression is induced upon wounding of the plant. A few sequences have been reported comprising a structural fold of a single α-helix, a single antiparallel β-sheet with two β-strands in a α-hairpin conformation, stabilized by four disulfide bonds. The orientation of the reactive site loop putatively indicated a noncanonical mode of action for mustard-type TIs (Fig. 2i) [6,29].

α-Hairpinin inhibitors

The α-hairpinin inhibitors are single-strand polypeptides of 3–5 kDa. Only a few members have been described in plants and they comprise two antiparallel oriented α-helices (helix-loop-helix motif) connected by two disulfide bonds with the specific cystine motif $C^1X_3C^2X_nC^3X_3C^4$ (Fig. 2j) [27]. Their N- and C-tails appear disordered. They are reported to inhibit trypsin proteases via a classical substrate-like mode of action. The binding regions of plant PIs usually are located at the end of a β-strand, but the interacting Arg in the prototypic TI peptide VhTI from *Veronica hederifolia* is located within a helix-to-helix motif [10,27].

Squash trypsin inhibitors and cyclic Momordica-type trypsin inhibitors

Squash TIs, also known as Cucurbitaceae-type peptides, are 3–5 kDa in size. They belong to the knottin peptide family, and contain the characteristic structural feature of three interconnected disulfide bonds in a knot fold topology. They typically also contain α-helical and β-sheet secondary elements and their N-termini are often protected with pyroglutamate. Squash TIs are potent trypsin, chymotrypsin, and elastase protease inhibitors occupying the active site of the protease (Fig. 2k). A few N- to C-terminally cyclized TI peptides have been isolated from *Momordica cochichinensis* (MCoTI) [30]. They are thought to be a product of an ancestral gene precursor homologous to that reported for squash TIs [12]. Momordica-type peptides are well-characterized model peptides. Their stabilizing scaffolds have found application in bioactive epitope grafting studies to develop drug leads for pharmaceutical targets [31,32].

Cyclic cysteine-rich peptides

Referred as to 'cyclotides', this family of peptides contains a cyclic cystine-knot motif. Cyclotides (as well as *Momordica*-type inhibitors) are processed *in planta* by asparagine endopeptidase to form the head-to-tail connected backbone [12].[1] Two antiparallel β-sheets in the core of the molecule and an optional α-helical segment supplement the cystine-knot (Fig. 2l). Cyclotides have been isolated from plants from the coffee (Rubiaceae), violet (Violaceae), legume (Fabaceae), potato (Solanaceae), and grass (Poaceae) families [33]. Interestingly, cyclotides are expressed in all plant tissues, including root, stems, and leaves. However, their reported concentration in seeds is low, which is atypical compared with other plant PIs [5,10]. Cyclotides are reported to be resistant to degradation by endo- and exoproteases, but lack the ability to inhibit trypsin or chymotrypsin activity [31,32]. Few cyclotides are known to inhibit post-proline cleaving endoprotease activity, but their mode of action has not yet been determined [34,35].

Following this overview of the structural characteristics and endogenous function of plant PIs, we continue to describe the importance of proteases as therapeutic drug targets.

Human proteases as drug targets

Approximately 1.7 % of the human genome encode proteases (referred to as hydrolases of peptide bonds) and they are the largest family of enzymes in humans, with at least 588 members [36,37]. Five protease families are distinguished in humans (brackets show the number of reported proteases): aspartic acid (21), cysteine (164), serine (184), threonine (27), and metallo- (192) proteases. These enzymes have a significant role in cellular signaling, for instance in neuropeptide and peptide hormone maturation as well as in protein turnover. Malfunctional protease activity can lead to disease and, thus, proteases are common drug targets [37,38]. Several validated protease targets are irreplaceable in human disease therapy: for instance, renin and angiotensin converting enzyme (hypertension), blood coagulation proteases, γ- and β-secretases (Alzheimer's disease), dipeptidyl peptidase-IV (hyperglycemia, type 2 diabetes mellitus, and cardiovascular disorders), HIV aspartyl protease (HIV), cathepsin K (osteoporosis), and matrix metalloproteases (cancer) [38–40]. We reviewed the literature and databases and identified several prototypic plant-derived peptide PIs that have been characterized to target human proteases and modulate their functionality. Table 1 provides a list of these peptides.

Representative examples and applications for plant PIs in autoimmune disorders, cancer, and neurodegenerative diseases

[1] Cyclized trypsin inhibitor peptides from *Momordica cochichinensis* are often referred to as a subfamily of cyclotides. They share the cyclic cystine-knot motif; however, MCoTI genes have higher sequence homology with squash-derived TIs and, hence, are classified as squash TIs in this review.

TABLE 1

Representative prototypic plant PIs of 12 PI families.

Name	PI class	Plant species	Plant family	Protease	Pharmacology[a]	PDB code	References
At-serpin1	Serpin	*Arabidopsis thaliana*	Brassicaceae	Trypsin	n.d.	3LE2	[63]
				Papain-like protease RD21	n.d.	(Fig. 2a)	
HorvuZx (BSZx)[b]	Serpin	*Hordeum vulgare*	Poaceae	Trypsin	k_a 3.9×10^6 M^{-1} s^{-1}	n.d.	[64,65]
				Chymotrypsin	k_a 9.4×10^5 M^{-1} s^{-1}		
				Factor Xa	k_a 6.9×10^3 M^{-1} s^{-1}		
				Cathepsin G	k_a 3.9×10^6 M^{-1} s^{-1}		
				Thrombin	k_a 8.0×10^3 M^{-1} s^{-1}		
				Plasma kallikrein	k_a 1.3×10^5 M^{-1} s^{-1}		
				Factor XIIa	weak		
				m-Plasminogen activator	weak		
				Leukocyte elastase	k_a 1.3×10^2 M^{-1} s^{-1}		
CmPS-1	Serpin	*Cucurbita maxima*	Cucurbitaceae	Pancreatic elastase, porcine	k_a 3.5×10^5 M^{-1} s^{-1}	n.d.	[66]
				Cathepsin G			
WSZ1	Serpin	*Triticum aestivum*	Poaceae	Chymotrypsin	k_a 1.3×10^6 M^{-1} s^{-1}	n.d.	[64,67]
				Cathepsin G	k_a 7.6×10^3 M^{-1} s^{-1}		
Oryzacystatin-I	Phytocystatin	*Oryza sativa*	Poaceae	Cathepsin B	K_i 7.9×10^{-8} M	1EQK	[68,69]
				Cathepsin H	K_i 1.0×10^{-6} M	(Fig. 2b)	
				Cathepsin L	K_i 7.3×10^{-10} M		
				Legumain	n.d.		
Oryzacystatin-II	Phytocystatin	*O. sativa*	Poaceae	Cathepsin H	K_i 2.5×10^{-8} M	n.d.	[68,69]
				Cathepsin F	n.d.		
				Legumain	n.d.		
Corn cystatin-I	Phytocystatin	*Zea mays*	Poaceae	Cathepsin H	K_i 5.7×10^{-9} M	n.d.	[70]
				Cathepsin L	K_i 1.7×10^{-8} M		
SQAPI	Phytocystatin	*Cucurbita maxima*	Cucurbitaceae	Pepsin	K_i 5.1×10^{-8} M	2 KXG	[71]
Cruzipain inhibitor (BbCI)	Kunitz	*Bauhinia bauhinioides*	Fabaceae	Leukocyte elastase	K_i 5.3×10^{-9} M	2 GZB	[72]
				Pancreatic elastase	K_i 4.0×10^{-9} M	(Fig. 2c)	
				Cathepsin L	K_i 2.2×10^{-9} M		
PCPI 6.6	Kunitz	*Solanum tuberosum*	Solanaceae	Cathepsin B	K_i 9.0×10^{-8} M	n.d.	[2,73]
				Cathepsin H	K_i 4.0×10^{-7} M		
				Cathepsin L	K_i 6.0×10^{-11} M		
PCPI 8.3	Kunitz	*S. tuberosum*	Solanaceae	Cathepsin L	K_i 7.0×10^{-11} M	3TC2	[74]
				Cathepsin B	K_i 3.2×10^{-7} M		
				Dipeptidyl peptidase I	K_i $>>5 \times 10^{-7}$ M		
				Clostripain	K_i $>>5 \times 10^{-7}$ M		
Kunitz inhibitor CPTI[b]	Kunitz	*Cicer arietinum*	Fabaceae	Trypsin	IC_{50} $<10^{-7}$ M	5XOZ	[75]
SKTI-3[1]	Kunitz	*Glycine max*	Fabaceae	Trypsin	IC_{50} 0.2×10^{-9} M	1AVU, 1AVW	[76,77]
				Chymotrypsin	K_i 7.0×10^{-8} M		
				Plasma kallikrein, human	K_i 5.7×10^{-9} M		
				Plasmin, human	K_i 1.9×10^{-7} M		
				Factor XIIa	K_i 1.4×10^{-6} M		
Enterolobium EcTI	Kunitz	*Enterolobium contortisiliquum*	Fabaceae	Trypsin	K_i 8.0×10^{-10} M	4J2 K, 4J2Y	[77]
				Chymotrypsin	K_i 1.1×10^{-9} M		
				Factor XIIa	K_i 8.2×10^{-8} M		
				Plasma kallikrein, human	K_i 6.1×10^{-9} M		
				Plasmin, human	K_i 9.4×10^{-9} M		
				Neutrophile elastase, human	K_i 5.5×10^{-8} M		

TABLE 1 (*Continued*)

Name	PI class	Plant species	Plant family	Protease	Pharmacology[a]	PDB code	References
Tamarind Kunitz inhibitor	Kunitz	*Tamarindus indica*	Fabaceae	Trypsin Factor Xa	K_i 3.2×10^{-9} M K_i 2.2×10^{-7} M	n.d.	[78]
Potato PCI	Kunitz	*S. tuberosum*	Solanaceae	Cathepsin D	K_i 1.0×10^{-9} M	5DZU	[2,79]
Soybean BBI (Isotype 2–II, 8 kDa)[b]	BBIs	*G. max*	Fabaceae	Trypsin Chymotrypsin Matriptase Mast cell chymase, human Cathepsin G Leukocyte elastase, human Duadenase	K_i ~4 $\times 10^{-9}$ M K_i 2.9×10^{-8} M n.d. K_i 5.0×10^{-10} M K_i 6.4×10^{-6} M K_i 2.3×10^{-9} M K_i 4.0×10^{-10} M	5J4Q (Fig. 2d)	[80–83]
SFTI-I	BBIs	*H. annuus*	Asteraceae	Trypsin Matriptase Kallikrein 5 Kallikrein 7 Kallikrein 14 Cathepsin G Elastase	K_i 1.7×10^{-10} M K_i 2.0×10^{-7} M K_i 1.4×10^{-7} M K_i 7.4×10^{-7} M K_i 2.5×10^{-8} M K_i 7.3×10^{-7} M IC_{50} ~1 $\times 10^{-5}$ M	1SFI, 3P8F	[84–86]
BI-I (seven isotypes I–VI)[b]	BBIs	*Ananas comosus*	Bromeliaceae	Trypsin Chymotrypsin Bromelain Cathepsin L Papain	n.d. n.d. n.d. n.d. n.d.	n.d.	[2,75]
AAI	α-Amylase inhibitors and bifunctional TIs	*Amaranthus hypochondriacus*	Amaranthaceae	Trypsin	n.d.	1QFD, 1HTX (Fig. 2e)	[87,88]
Corn Hageman factor inhibitor	α-Amylase inhibitors and bifunctional TIs	*Z. mays*	Poaceae	Trypsin Factor XIIa	K_i 2.3×10^{-8} M K_i 2.0×10^{-9} M	1BEA	[89,90]
RATI	α-Amylase inhibitors and bifunctional TIs	*Eleusine coracana*	Poaceae	Trypsin	K_i 1.2×10^{-9} M	1B1U	[91]
SOTI-I	α-Amylase inhibitors and bifunctional TIs	*Spinacia oleracea*	Amaranthaceae	β-Trypsin, bovine	K_a 1.8×10^{7} M^{-1}	n.d.	[92]
CMTI-V	Potato type-I	*C. maxima*	Cucurbitaceae	Trypsin Factor XIIa	K_i 1.6×10^{-8} M K_i 4.1×10^{-8} M	1TIN (Fig. 2f)	[93,94]
ATSI	Potato type-I	*Amaranthus caudatus*	Amaranthaceae	Trypsin Chymotrypsin Factor XIIa Cathepsin G Plasmin	K_i 3.4×10^{-10} M K_i 4.1×10^{-10} M K_i 4.4×10^{-7} M K_i 1.2×10^{-7} M K_i 3.8×10^{-8} M	n.d.	[95]
CI-I	Potato type-I	*H. vulgare*	Poaceae	Trypsin Chymotrypsin Neutrophil elastase Subtilisin	n.d. K_i ~3 $\times 10^{-6}$ M K_i 2.0×10^{-9} M K_i 2.0×10^{-10} M	n.d.	[2,96]
BWI-1	Potato type-I	*Fagopyrum esculentum*	Polygonaceae	Trypsin α-Chymotrypsin Cathepsin G	K_i ~1 $\times 10^{-9}$ M K_i ~3 $\times 10^{-7}$ M K_i ~1 $\times 10^{-7}$ M	n.d.	[97]
TI-II	Potato type-II	*Solanum lycopersicum*	Solanaceae	Trypsin Subtilisin Chymotrypsin	K_i 8.0×10^{-8} M K_i 9.0×10^{-9} M K_i 3.0×10^{-9} M	1PJU (Fig. 2g)	[98,99]
PI-2	Potato type-II	*S. tuberosum*	Solanaceae	Trypsin Chymotrypsin	K_i 4.0×10^{-10} M K_i 9.0×10^{-10} M	n.d.	[100]

TABLE 1 (*Continued*)

Name	PI class	Plant species	Plant family	Protease	Pharmacology[a]	PDB code	References
PSI-1.1	Potato type-II	*Capsicum annuum*	Solanaceae	Trypsin	K_i 4.8×10^{-8} M	n.d.	[101]
				Chymotrypsin	K_i 4.7×10^{-8} M		
Potato CPI	MCPI	*S. tuberosum*	Solanaceae	Carboxypeptidase A	K_i $\sim 2 \times 10^{-9}$ M	4CPA, 1H20	[102,103]
				Carboxypeptidase B	K_i $\sim 4 \times 10^{-9}$ M	(Fig. 2h)	
				Carboxypeptidase (mast cell)	n.d		
				Carboxypeptidase E	n.d		
				Carboxypeptidase M	n.d		
				Cathepsin A/L	n.d		
MCPI	MCPI	*S. lycopersicum*	Solanaceae	Carboxypeptidase A	n.d.	n.d.	[2]
				Carboxypeptidase B	n.d.		
Cabbage TI	Mustard-type PI	*Brassica oleracea*	Brassicaceae	Trypsin	IC_{50} 2.0×10^{-7} M	n.d.	[2,104]
				Thrombin	n.d.	(Fig. 2i)	
				Factor Xa	n.d.		
				Factor XIIa	n.d.		
				Plasmin	n.d.		
RTI-I	Mustard-type PI	*Brassica napus*	Brassicaceae	β-Trypsin, bovine	K_D $\sim 2 \times 10^{-10}$ M	n.d.	[105]
				α-Chymotrypsin, bovine	K_D 4.1×10^{-7} M		
MTI-2	Mustard-type PI	*Sinapis alba*	Brassicaceae	β-Trypsin, bovine	K_i 1.0×10^{-11} M	n.d.	[2,29,106]
				α-Chymotrypsin	K_a 2.0×10^{6} M^{-1}		
VhTI	α-hairpinin	*V. hederifolia*	Plantaginaceae	Trypsin	K_i $<1 \times 10^{-9}$ M	2PLX (Fig. 2j)	[107]
EETI-II	Squash PI	*Ecballium elaterium*	Cucurbitaceae	Trypsin	K_a 8.0×10^{11} M^{-1}	1W7Z (Fig. 2k)	[108,109]
MCoTI-II	Squash PI	*M. cochichinensis*	Cucurbitaceae	Trypsin	K_i 2.3×10^{-12} M	4GUX	[61,110,111]
				Kallikrein 2–3	K_i 1.0×10^{-5} M		
				Kallikrein 4	K_i 1.6×10^{-9} M		
				Plasmin	K_i 2.8×10^{-8} M		
				Factor XIIa	K_i 7.5×10^{-7} M		
				β-tryptase	K_i 6.0×10^{-7} M		
				Matriptase	K_i 9.0×10^{-9} M		
CMTI-I	Squash PI	*C. maxima*	Cucurbitaceae	β-Trypsin, bovine	K_D 1.3×10^{-8} M	1PPE	[94,112]
				Hageman factor	K_D 4.1×10^{-8} M		
				Plasmin	K_a 2.0×10^{7} M^{-1}		
				Plasma kallikrein	K_a 9.4×10^{4} M^{-1}		
				Factor Xa	K_a 4.1×10^{3} M^{-1}		
				Factor XIIa	K_a 4.0×10^{7} M^{-1}		
CMTI-III	Squash PI	*C. maxima*	Cucurbitaceae	β-Trypsin, bovine	K_a 6.8×10^{11} M^{-1}	n.d.	[2,113]
				Factor XIIa	IC_{50} 7.0×10^{-9} M		
				Kallikrein	IC_{50} 3.0×10^{-12} M		
CPTI-II	Squash PI	*Cucurbita pepo*	Cucurbitaceae	Plasmin	K_a 1.7×10^{7} M^{-1}	2BTC	[2,112]
				Plasma kallikrein	K_a 3.8×10^{3} M^{-1}		
				Factor Xa	K_a $>10^{3}$ M^{-1}		
				Factor XIIa	K_a 1.3×10^{6} M^{-1}		
ELTI-I	Squash PI	*Echinocystis lobata*	Cucurbitaceae	β-Trypsin, bovine	K_a 6.6×10^{10} M^{-1}	n.d.	[114]
				Cathepsin G	K_a 3.1×10^{12} M^{-1}		
LLDTI-II	Squash PI	*Lagenaria leucantha*	Cucurbitaceae	Trypsin, bovine	K_i 9.6×10^{-11} M	n.d.	[2,115]
				Factor Xa	K_i 4.1×10^{-5} M		
				Factor XIIa	K_i 1.4×10^{-6} M		
				Plasma kallikrein	K_i 2.7×10^{-5} M		

TABLE 1 (*Continued*)

Name	PI class	Plant species	Plant family	Protease	Pharmacology[a]	PDB code	References
LLTDI-III	Squash PI	*L. leucantha*	Cucurbitaceae	Trypsin	K_i 3.0×10^{-10} M	n.d.	[2,115]
				Factor Xa	K_i 1.9×10^{-5} M		
				Factor XIIa	K_i 4.2×10^{-6} M		
				Plasma kallikrein	K_i 2.0×10^{-4} M		
LCTI-II	Squash PI	*Lagenaria cylindrica*	Cucurbitaceae	Factor Xa	K_i 7.8×10^{-4} M	n.d.	[2,115]
				Factor XIIa	K_i 7.5×10^{-8} M		
				Kallikrein	K_i 9.6×10^{-11} M		
MCEI-III	Squash PI	*Momordica charantia*	Cucurbitaceae	Porcine elastase	K_i 4.0×10^{-9} M	n.d.	[116]
Kalata B1	Cyclotide	*Oldenlandia affinis*	Rubiaceae	Prolyl endopeptidase, human	IC_{50} ~6×10^{-6} M	1NB1 (Fig. 2I)	[35]
Psysol-2	Cyclotide	*Psychotria solitudinum*	Rubiaceae	Prolyl endopeptidase, human	IC_{50} 2.5×10^{-5} M	n.d.	[35]

[a] k_a – apparent rate constant [M^{-1} s^{-1}]; K_a – association equilibrium constant [M^{-1}]; K_i – inhibition constant (also referred to as equilibrium dissociation constant, K_D) [M]; n.d. not determined; IC_{50}, concentration required to produce half maximum inhibition of enzyme activity [M].

[b] Kunitz, phytocystatin as well as BB inhibitors are often present as isotypes in plants, and only differ in a few amino acids to each other. Whether an isolated isotype or a mixture of two or more PIs has been used is often not reported.

are described in Box 1. Here, we discuss opportunities for plant PIs as starting points for PI drug discovery and development.

Opportunities for plant protease inhibitors in drug discovery and development

Proteases are important drug targets with many clinical applications. For instance, serine proteases are in the top ten of all protein targets. Nevertheless, there are several challenges associated with the intrinsic druggability of these enzymes. Above all, there is a limited ligand specificity of existing protease-targeting drugs and drug candidates. Many protease drug discovery efforts build on small molecules [1,38,39]. However, these chemical entities with a small surface area and a low number of contact sites often exhibit limited selectivity for structural and functional homologues proteases [41,42]. As an alternative to small molecules, peptides have gained remarkable interest in drug discovery over the past few years. They cover a unique chemical space, which differs from that of small molecules. In particular, macrocyclic peptides are considered promising drug-like molecules because they combine some of the advantages of traditional small-molecule drugs (e.g., conformational restriction) with those of larger biologics (e.g., target specificity) [17,22,24,31]. There are currently >70 peptides approved as drugs, >150 under clinical investigation, and many thousands in preclinical development [36,43,44]. The intrinsically high target selectivity of peptide-derived drugs has been well documented by some approved peptide and/or peptidomimetic PI drugs; for instance, captopril and enalapril (angiotensin-converting enzyme inhibitors), bivalirudin (thrombin inhibitor), and carfilzomib (inhibitor of proteases from the chymotrypsin superfamily in the 20S proteasome).

Potential peptide drugs have been discovered from natural sources, such as venom-derived peptide toxins. However, plant-derived peptide PIs have received little attention from the protease drug development community so far [44], although these molecules cover a large spectrum of modes of action compared with small molecules [38,39]: Most plant PIs exhibit a canonical-competitive inhibition mode via 'substrate-like' binding to the catalytic domain of the targeted protease (e.g., BBIs and Kunitz inhibitors), or they make use of a noncatalytically competitive inhibition (e.g., cystatins or mustard-type PI). Some plant PIs act via a mixed mode, where the primary competitive binding to the active site is supported by a secondary binding event (e.g., MCPIs). Other plant PIs utilize an irreversible inhibition of proteolytic activity (e.g., serpins) [17,39–41]. Importantly, protease exosites, for instance in the blood coagulation enzymes thrombin and factor Xa, are amenable for targeting by natural peptide inhibitors, such as hirudin isolated from leeches [45]. The modulation of protease activity through exosite binding is a promising concept in PI development and could serve as valuable model for the discovery of plant PIs with interesting mechanisms of action [1,46].

In our opinion, a second challenge associated with protease drug development is a reduced motivation to consider the diversity of this drug target family in early drug discovery. The community was built on 'model proteases', such as trypsin, chymotrypsin, pancreatic elastase, or several blood coagulation factors, which became the work horses for the discovery and initial characterization of PIs. However, many protease families,

BOX 1

Plant protease inhibitors target human disease relevant proteases

Immune system and autoimmunity

Proteases are ubiquitously involved in immune system regulation. They can function as processing enzymes of effector molecules (e.g., tumor necrosis factor-α, interferon-γ, linterleukin-1 (IL-1), IL-18, IL-33, or peptide hormones) as well as surface recognition motifs for cell–cell communication (i.e., in antigen presentation or as CD10 or CD26). The dysregulation of plasmin, thrombin, thymus-specific serine protease, or chymase protease can be associated with the pathogenesis of autoimmune disorders [21]. In two studies using the experimental autoimmune encephalomyelitis (EAE) model of multiple sclerosis, the soybean BBI significantly reduced progression of disease severity and ameliorated demyelination and inflammatory markers after intraperitoneal or peroral application [123]. An IL-10-dependent mode of action was identified [20,123]. The observed therapeutic effect is promising, because the compound reduced the infiltration of encephalomyelitic cells into the spinal cord and the central nervous system, as well as the inflammatory response by Th_1 and Th_{17} effector T cells [21]. Another plant PI, the cyclotide analog [T20 K]kalata B1, exhibited immunosuppressive effects in the EAE model in an IL-2-dependent mechanism [35,57,124].

Cancer

Proteases are relevant drug targets for cancer therapy, because tumor progression, cancer cell migration, and metastasis formation are influenced by proteolytic processes. Plasma and cell surface proteases, for instance, matrix metalloproteases, cathepsin B, kallikreins, hepsin, and matriptase, have been established as drug targets for plant PI inhibitors in antitumor therapies, because these proteins are accessible via systemic delivery of PI drugs [125]. Soybean BBI was reported as nanomolar inhibitor for several proteases: matriptase, human neutrophil elastase, chymase, cathepsin G, or duadenase [126]. In HT-29 colon cancer and MCF-7 breast cancer cell lines, the BBI interrupted cell proliferation, causing cell cycle arrest and apoptosis [56]. The soybean BBI concentrate (a mixture containing several BBI molecules, usually defined by its chymotrypsin-inhibiting units) was also reported as a cancer chemopreventive agent for neoplastic polyps it was applied in clinical trials for oral hyperplasia and oral leukoplakia as well as preclinically for prostate cancer detection [9,127,128]. Furthermore, *Macrotylmoa axillare*-derived peptides suppressed colorectal neuroplasia in the 1,2-dimethylhydrazine (DMH) mouse model [129]; protease inhibitors from *Phaseolus acutifolius* and *Phaseolus vulgaris*, *Medicago scutellate*, and *Pisum sativum* modulate matrix metalloprotease activity (i.e., MMP-2 and -9), resulting in vitro in reduced cancer cell invasion [9,39,56].

Neurodegenerative diseases

Neurodegenerative diseases are progressive illnesses with a growing incidence in Western populations. The development and cause of disease are often not fully understood; however, proteases are involved in the turnover and processing of malfunctioned proteins [38–40,47,130]. For example, secretases process amyloid-β precursor protein in Alzheimer's disease and several proteases, such as prolyl oligopeptidase (POP) or asparagine endopeptidase, are involved in the maturation of α-synuclein (aSyn) in Parkinson's disease and other Lewy pathologies [17,130–133]. The plant PIs kalata B1 and psysol-2 from the cyclotide family have been reported as the first nature-derived kalata-type inhibitors for the human serine endopeptidase POP. The peptides ameliorate postproline cleavage activity in vitro, but do not inhibit trypsin or chymotrypsin activity [35]. Whether cyclotides are inhibitors of serine proteases with Pro-specific catalytic activity has not yet been investigated.

including clinically relevant proteases, have received too little attention and there is a lack of established proteases for preclinical inhibitor drug-screening studies [38,47]. This focus on a few prototypic proteases restricts the discovery of novel inhibitors (and drug leads) and has hampered PI drug development previously. Therefore, most human proteases are still under- or even unexplored as drug targets. For instance, only single to a few serine proteases of two subfamilies are reportedly targeted by at least one plant PI, but the other 53 proteases (classified in 17 subfamilies) remain unexplored regarding plant PI discovery. Similarly, cysteine proteases (two of 28), aspartic acid proteases (one of three), and metalloproteases (one of 56 subfamilies) appear even less researched to date. Given that the latter protease family is among the largest and despite their medical value as drug targets in immunological, pulmonary, or cardiac disease (e.g., matrix metalloproteases), there are few peptide and/or protein-derived inhibitors targeting these proteases, which provides an opportunity for future plant PI drug discovery (Fig. 1c, Table 1) [48].

Importantly, only 1–5% of plant species of the top-ten PI-expressing plant families have been investigated and reported to express at least one of these molecules to date (all species of the assigned category reported in www.theplantlist.org/ have been considered). Thus, we are only beginning to understand the natural diversity of plant PIs. An advantage for future plant PI discovery projects is that they are readily available for *in silico* analysis and, therefore, the identification of promising drug candidates can be guided by genome or transcriptome mining (i.e., 1KP or 10KP consortia) [35,49–51]. In addition, because many structures of prototypic plant PI as well as of disease-relevant protease drug targets have been solved, structure- and ligand-based drug design or similar computational approaches could be applied to screen for likely inhibitors in combinatorial peptide libraries from plants or to provide a basis for the chemical optimization of identified molecules [35,50,52,53].

Plant PIs could provide functionalized scaffolds, which address key features for therapeutic peptide development (i.e., stability, delivery, and specificity) [11,43,54]. Plant PI structures convey high stability towards proteolysis, heat, or acidic conditions, which offers advantages compared with linear peptides [34]. BBIs (e.g., soybean BBI), knottins (e.g., MCoTI-II), or cyclotides (e.g., kalata B1) were successfully tested for therapeutic applications via oral delivery [55–58]. These results are encouraging and support the use of cysteine-stabilized peptide PIs for the development of orally active therapeutic drugs [24,32,57]. Furthermore, peptide therapeutics are also considered 'safe', because they are metabolized and degraded to nontoxic amino acids and smaller peptide fragments. Moreover, plant PI-derived drug candidates offer high target selectivity, which would reduce the risk of off-target adverse effects [22,24,31]. However, peptides also have certain limitations such as low bioavailability, possible immunogenic reactions, and relatively high synthetic costs compared with small organic compounds.

Importantly, plant peptide PIs (e.g., squash TIs, mustard TIs, potato-type inhibitors, α-hairpinins, and cyclotides) are amenable to solid-phase peptide synthesis, which is essential in the design of lead compounds [24,32], and could present an opportunity to use plant-derived peptide PIs as a starting point for drug engineering [31]. For instance, the MCoTI-II or sunflower trypsin inhibitor-I peptides are able to incorporate nonendogenous protease-inhibiting motifs into their stabilized framework [59,60]. These new chimeric molecules conserve the potency and selectivity of the binding epitope and utilize the exceptional stability provided by the template scaffold to yield promising drug candidates [22,24]. As an example, the important cancer target kallikrein 4, a serine protease, has been targeted by an engineered inhibitor derived from the MCoTI-II squash-TI with an inhibitory affinity of $K_i = 0.1$ nM and a selectivity of >100 000-fold for other related kallikreins [61].

Concluding remarks

To conclude, there is a long road ahead until native or 'engineered' plant-derived peptide PIs can be developed as pharmaceuticals, but we predict that, within the next years, the growing interest of the pharmaceutical community in peptide scaffolds for drug design and novel plant PIs will offer opportunities for these peptides to enter clinical development [24,31,44]. Nevertheless, the major challenge of peptide pharmaceuticals also holds true for plant-derived PI drug candidates: their low oral bioavailability limits their clinical use, and it will be imperative to find solutions for improvement [62].

Acknowledgement

Work on peptides in the laboratory of C.W.G. was supported by the Austrian Science Fund FWF (I3243).

References

1 Atanasov, A.G. *et al.* (2015) Discovery and resupply of pharmacologically active plant-derived natural products: a review. *Biotechnol. Adv.* 33, 1582–1614
2 Polya, G.M. (2003) Protein and non-protein protease inhibitors from plants. *Stud. Nat. Prod. Chem.* 29, 567–641
3 Kaysser, L. (2019) Built to bind: biosynthetic strategies for the formation of small–molecule protease inhibitors. *Nat. Prod. Rep.* . http://dx.doi.org/10.1039/c8np00095f
4 Zorzi, A. *et al.* (2017) Cyclic peptide therapeutics: past, present and future. *Curr. Opin. Chem. Biol.* 38, 24–29
5 Ryan, C.A. (1990) Protase inhibitors in plants: genes for improving defenses against insects and pathogens. *Annu. Rev. Phytopathol.* 28, 425–449
6 Gomes, M.T. *et al.* (2011) Plant proteinases and inhibitors: an overview of biological function and pharmacological activity. *Curr. Protein Pept. Sci.* 12, 417–436
7 Rawlings, N.D. *et al.* (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in. (2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46, D624–D632
8 Rawlings, N.D. (2010) Peptidase inhibitors in the MEROPS database. *Biochimie* 92, 1463–1483
9 Srikanth, S. and Chen, Z. (2016) Plant protease inhibitors in therapeutics–focus on cancer therapy. *Front. Pharmacol.* 7, 470
10 Bateman, K.S. and James, M.N. (2011) Plant protein proteinase inhibitors: structure and mechanism of inhibition. *Curr. Protein Pept. Sci.* 12, 340–347
11 Arnison, P.G. *et al.* (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* 30, 108–160
12 Mylne, J.S. *et al.* (2012) Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis. *Plant Cell* 24, 2765–2778
13 Volpicella, M. *et al.* (2011) Cystatins, serpins and other families of protease inhibitors in plants. *Curr. Protein Pept. Sci.* 12, 386–398
14 Gettins, P.G.W. (2002) Serpin structure, mechanism, and function. *Chem. Rev.* 102, 4751–4803
15 Huntington, J.A. (2011) Serpin structure, function and dysfunction. *J. Thromb. Haemost.* 9 (Suppl. 1), 26–34
16 Benchabane, M. *et al.* (2010) Plant cystatins. *Biochimie* 92, 1657–1666
17 Farady, C.J. and Craik, C.S. (2010) Mechanisms of macromolecular protease inhibitors. *ChemBioChem* 11, 2341–2346
18 Bendre, A.D. *et al.* (2018) Analysis of Kunitz inhibitors from plants for comprehensive structural and functional insights. *Int. J. Biol. Macromol.* 113, 933–943
19 Armstrong, W.B. *et al.* (2013) Bowman Birk inhibitor concentrate and oral leukoplakia: a randomized phase IIb trial. *Cancer Prev. Res.* 6, 410–418
20 Dai, H. *et al.* (2011) Bowman-Birk Inhibitor attenuates experimental autoimmune encephalomyelitis by delaying infiltration of inflammatory cells into the CNS. *Immunol. Res.* 51, 145–152
21 Safavi, F. and Rostami, A. (2012) Role of serine proteases in inflammation: Bowman-Birk protease inhibitor (BBI) as a potential therapy for autoimmune diseases. *Exp. Mol. Pathol.* 93, 428–433
22 Craik, D.J. (2009) Circling the enemy: cyclic proteins in plant defence. *Trends Plant Sci.* 14, 328–335
23 Elliott, A.G. *et al.* (2014) Evolutionary origins of a bioactive peptide buried within Preproalbumin. *Plant Cell* 26, 981–995
24 Craik, D.J. *et al.* (2013) The future of peptide-based drugs. *Chem. Biol. Drug Des.* 81, 136–147
25 Svensson, B. *et al.* (2004) Proteinaceous alpha-amylase inhibitors. *Biochim. Biophys. Acta* 1696, 145–156
26 Nguyen, P.Q. *et al.* (2015) Antiviral cystine knot alpha-amylase inhibitors from *Alstonia scholaris*. *J. Biol. Chem.* 290, 31138–31150
27 Tam, J.P. *et al.* (2015) Antimicrobial peptides from plants. *Pharmaceuticals* 8, 711–757
28 Turra, D. and Lorito, M. (2011) Potato type I and II proteinase inhibitors: modulating plant physiology and host resistance. *Curr. Protein Pept. Sci.* 12, 374–385
29 Volpicella, M. *et al.* (2000) Characterization of recombinant mustard trypsin inhibitor 2 (MTI2) expressed in Pichia pastoris. *FEBS Lett.* 468, 137–141
30 Avrutina, O. *et al.* (2005) Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTI-II. *Biol. Chem.* 386, 1301–1306
31 Craik, D.J. *et al.* (2018) Ribosomally-synthesised cyclic peptides from plants as drug leads and pharmaceutical scaffolds. *Bioorg. Med. Chem.* 26, 2727–2737
32 Wang, C.K. and Craik, D.J. (2018) Designing macrocyclic disulfide-rich peptides for biotechnological applications. *Nat. Chem. Biol.* 14, 417–427
33 Gruber, C.W. (2010) Global cyclotide adventure: a journey dedicated to the discovery of circular peptides from flowering plants. *Biopolymers* 94, 565–572
34 Colgrave, M.L. and Craik, D.J. (2004) Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. *Biochemistry* 43, 5965–5975
35 Hellinger, R. *et al.* (2015) Inhibition of human prolyl oligopeptidase activity by the cyclotide psysol 2 isolated from *Psychotria solitudinum*. *J. Nat. Prod.* 78, 1073–1082
36 Rask-Andersen, M. *et al.* (2011) Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discov.* 10, 579–590
37 Perez-Silva, J.G. *et al.* (2016) The Degradome database: expanding roles of mammalian proteases in life and disease. *Nucleic Acids Res.* 44, D351–D355
38 Drag, M. and Salvesen, G.S. (2010) Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discov.* 9, 690–701
39 Turk, B. (2006) Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov.* 5, 785–799
40 Turk, B. *et al.* (2012) Protease signalling: the cutting edge. *EMBO J.* 31, 1630–1643
41 Santos, R. *et al.* (2017) A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* 16, 19–34
42 Vazquez-Lombardi, R. *et al.* (2015) Challenges and opportunities for non-antibody scaffold drugs. *Drug Discov. Today* 20, 1271–1283
43 Gruber, C.W. *et al.* (2010) Ligand-based peptide design and combinatorial peptide libraries to target G protein-coupled receptors. *Curr. Pharm. Des.* 16, 3071–3088
44 Lau, J.L. and Dunn, M.K. (2018) Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg. Med. Chem.* 26, 2700–2707

45 Krowarsch, D. *et al.* (2003) Canonical protein inhibitors of serine proteases. *Cell Mol. Life Sci.* 60, 2427–2444
46 Bock, P.E. *et al.* (2007) Exosites in the substrate specificity of blood coagulation reactions. *J. Thromb. Haemost.* 5 (Suppl. 1), 81–94
47 Lopez-Otin, C. and Matrisian, L.M. (2007) Emerging roles of proteases in tumour suppression. *Nat. Rev. Cancer* 7, 800–808
48 Visse, R. and Nagase, H. (2003) Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. *Circ. Res.* 92, 827–839
49 Koehbach, J. *et al.* (2013) Cyclotide discovery in Gentianales revisited–identification and characterization of cyclic cystine-knot peptides and their phylogenetic distribution in Rubiaceae plants. *Biopolymers* 100, 438–452
50 Hellinger, R. *et al.* (2015) Peptidomics of circular cysteine-rich plant peptides: analysis of the diversity of cyclotides from viola tricolor by transcriptome and proteome mining. *J. Proteome Res.* 14, 4851–4862
51 Cheng, S. *et al.* (2018) 10KP: a phylodiverse genome sequencing plan. *Gigascience* 7, 1–9
52 Silverstein, K.A. *et al.* (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J.* 51, 262–280
53 Serra, A. *et al.* (2016) A high-throughput peptidomic strategy to decipher the molecular diversity of cyclic cysteine-rich peptides. *Sci. Rep.* 6, 23005
54 Rask-Andersen, M. *et al.* (2014) The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu. Rev. Pharmacol. Toxicol.* 54, 9–26
55 Lichtenstein, G.R. *et al.* (2008) Bowman-Birk inhibitor concentrate: a novel therapeutic agent for patients with active ulcerative colitis. *Dig. Dis. Sci.* 53, 175–180
56 Clemente, A. and Arques Mdel, C. (2014) Bowman-Birk inhibitors from legumes as colorectal chemopreventive agents. *World J. Gastroenterol.* 20, 10305–10315
57 Thell, K. *et al.* (2016) Oral activity of a nature-derived cyclic peptide for the treatment of multiple sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3960–3965
58 Wong, C.T. *et al.* (2012) Orally active peptidic bradykinin B1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment. *Angew. Chem. Int. Ed. Engl.* 51, 5620–5624
59 Swedberg, J.E. *et al.* (2019) Highly potent and selective plasmin inhibitors based on the sunflower trypsin inhibiTOR-1 scaffold attenuate fibrinolysis in plasma. *J. Med. Chem.* 62, 552–560
60 de Veer, S.J. *et al.* (2016) Exploring the active site binding specificity of kallikrein-related peptidase 5 (KLK5) guides the design of new peptide substrates and inhibitors. *Biol. Chem.* 397, 1237–1249
61 Swedberg, J.E. *et al.* (2018) Potent, selective, and cell-penetrating inhibitors of kallikrein-related peptidase 4 based on the cyclic peptide MCoTI-II. *ACS Med. Chem. Lett.* 9, 1258–1262
62 Rader, A.F.B. *et al.* (2018) Orally active peptides: is there a magic bullet? *Angew. Chem. Int. Ed. Engl.* 57, 14414–14438
63 Lampl, N. *et al.* (2010) Arabidopsis AtSerpin1, crystal structure and *in vivo* interaction with its target protease RESPONSIVE TO DESICCATION-21 (RD21). *J. Biol. Chem.* 285, 13550–13560
64 Dahl, S.W. *et al.* (1996) Heterologous expression of three plant serpins with distinct inhibitory specificities. *J. Biol. Chem.* 271, 25083–25088
65 Dahl, S.W. *et al.* (1996) Inhibition of coagulation factors by recombinant barley serpin BSZx. *FEBS Lett.* 394, 165–168
66 Yoo, B.C. *et al.* (2000) Characterization of *Cucurbita maxima* phloem serpin-1 (CmPS-1). A developmentally regulated elastase inhibitor. *J. Biol. Chem.* 275, 35122–35128
67 Ostergaard, H. *et al.* (2000) Inhibitory serpins from wheat grain with reactive centers resembling glutamine-rich repeats of prolamin storage proteins. Cloning and characterization of five major molecular forms. *J. Biol. Chem.* 275, 33272–33279
68 Nagata, K. *et al.* (2000) Three-dimensional solution structure of oryzacystatin-I, a cysteine proteinase inhibitor of the rice, *Oryza sativa* L. *japonica*. *Biochemistry* 39, 14753–14760
69 Valadares, N.F. *et al.* (2010) Molecular determinants of improved cathepsin B inhibition by new cystatins obtained by DNA shuffling. *BMC Struct. Biol.* 10, 30
70 Abe, M. *et al.* (1994) Corn cystatin I expressed in *Escherichia coli*: investigation of its inhibitory profile and occurrence in corn kernels. *J. Biochem.* 116, 488–492
71 Headey, S.J. *et al.* (2010) Solution structure of the squash aspartic acid proteinase inhibitor (SQAPI) and mutational analysis of pepsin inhibition. *J. Biol. Chem.* 285, 27019–27025
72 Hansen, D. *et al.* (2007) Crystal structure of a novel cysteinless plant Kunitz-type protease inhibitor. *Biochem. Biophys. Res. Commun.* 360, 735–740
73 Brzin, J. *et al.* (1988) Inhibitors of cysteine proteinases from potato. *Biol. Chem. Hoppe Seyler* 369 (Suppl), 233–238
74 Rowan, A.D. *et al.* (1990) Inhibition of cysteine proteinases by a protein inhibitor from potato. *FEBS Lett.* 269, 328–330
75 Srinivasan, A. *et al.* (2005) A Kunitz trypsin inhibitor from chickpea (*Cicer arietinum* L.) that exerts anti-metabolic effect on podborer (*Helicoverpa armigera*) larvae. *Plant Mol. Biol.* 57, 359–374
76 Song, H.K. and Suh, S.W. (1998) Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J. Mol. Biol.* 275, 347–363
77 *et al.* (2013) Crystal structures of a plant trypsin inhibitor from Enterolobium contortisiliquum (EcTI) and of its complex with bovine trypsin. *PLoS One* 8 (4), e62252. http://dx.doi.org/10.1371/journal.pone.0062252 Published 2013 Apr 23
78 Patil, D.N. *et al.* (2012) Structural basis for dual inhibitory role of tamarind Kunitz inhibitor (TKI) against factor Xa and trypsin. *FEBS J.* 279, 4547–4564
79 Guo, J. *et al.* (2015) Structure of a Kunitz-type potato cathepsin D inhibitor. *J. Struct. Biol.* 192, 554–560
80 Odani, S. and Ikenaka, T. (1977) Studies on soybean trypsin inhibitors. X. Isolation and partial characterization of four soybean double-headed proteinase inhibitors. *J. Biochem.* 82, 1513–1522
81 Seidl, D.S. and Liener, I.E. (1972) Isolation and properties of complexes of the Bowman-Birk soybean inhibitor with trypsin and chymotrypsin. *J. Biol. Chem.* 247, 3533–3538
82 Ware, J.H. *et al.* (1997) Soybean Bowman-Birk protease inhibitor is a highly effective inhibitor of human mast cell chymase. *Arch. Biochem. Biophys.* 344, 133–138
83 Gladysheva, I.P. *et al.* (2002) Study on interaction between duodenase protease with dual specificity and inhibitors of bowman-birk family. *Protein Pept. Lett.* 9, 139–144
84 Quimbar, P. *et al.* (2013) High-affinity cyclic peptide matriptase inhibitors. *J. Biol. Chem.* 288, 13885–13896
85 de Veer, S.J. *et al.* (2015) Engineered protease inhibitors based on sunflower trypsin inhibitor-1 (SFTI-1) provide insights into the role of sequence and conformation in Laskowski mechanism inhibition. *Biochem. J.* 469, 243–253
86 Swedberg, J.E. *et al.* (2017) Design of potent and selective cathepsin G inhibitors based on the sunflower trypsin inhibitor-1 scaffold. *J. Med. Chem.* 60, 658–667
87 Chagolla-Lopez, A. *et al.* (1994) A novel alpha-amylase inhibitor from amaranth (*Amaranthus hypocondriacus*) seeds. *J. Biol. Chem.* 269, 23675–23680
88 Martins, J.C. *et al.* (2001) Solution structure of the main alpha-amylase inhibitor from amaranth seeds. *Eur. J. Biochem.* 268, 2379–2389
89 Behnke, C.A. *et al.* (1998) Structural determinants of the bifunctional corn Hageman factor inhibitor: x-ray crystal structure at 1.9)5 A resolution. *Biochemistry* 37, 15277–15288
90 Chong, G.L. and Reeck, G.R. (1987) Interaction of trypsin, beta-factor XIIa, and plasma kallikrein with a trypsin inhibitor isolated from barley seeds: a comparison with the corn inhibitor of activated Hageman factor. *Thromb. Res.* 48, 211–221
91 Maskos, K. *et al.* (1996) RBI, a one-domain alpha-amylase/trypsin inhibitor with completely independent binding sites. *FEBS Lett.* 397, 11–16
92 Kowalska, J. *et al.* (2007) Trypsin inhibitors from the garden four o'clock (Mirabilis jalapa) and spinach (*Spinacia oleracea*) seeds: isolation, characterization and chemical synthesis. *Phytochemistry* 68, 1487–1496
93 Cai, M. *et al.* (1995) Three-dimensional solution structure of *Cucurbita maxima* trypsin inhibitor-V determined by NMR spectroscopy. *Biochemistry* 34, 5201–5211
94 Krishnamoorthi, R. *et al.* (1990) A new protein inhibitor of trypsin and activated Hageman factor from pumpkin (*Cucurbita maxima*) seeds. *FEBS Lett.* 273, 163–167
95 Hejgaard, J. *et al.* (1994) Primary structure and specificity of the major serine proteinase inhibitor of amaranth (*Amaranthus caudatus* L.) seeds. *Biochim. Biophys. Acta* 1204, 68–74
96 Greagg, M.A. *et al.* (1994) Expression and kinetic characterization of barley chymotrypsin inhibitors 1a and 1b. *Biochim. Biophys. Acta* 1222, 179–186
97 Dunaevsky, Y.E. *et al.* (1997) The anionic protease inhibitor BWI-1 from buckwheat seeds. Kinetic properties and possible biological role. *Physiol. Plant* 101, 483–488
98 Barrette-Ng, I.H. *et al.* (2003) Unbound form of tomato inhibitor-II reveals interdomain flexibility and conformational variability in the reactive site loops. *J. Biol. Chem.* 278, 31391–31400
99 Barrette-Ng, I.H. *et al.* (2003) Structural basis of inhibition revealed by a 1:2 complex of the two-headed tomato inhibitor-II and subtilisin Carlsberg. *J. Biol. Chem.* 278, 24062–24071
100 Beekwilder, J. *et al.* (2000) Characterization of potato proteinase inhibitor II reactive site mutants. *Eur. J. Biochem.* 267, 1975–1984
101 Antcheva, N. *et al.* (1996) Primary structure and specificity of a serine proteinase inhibitor from paprika (*Capsicum annuum*) seeds. *Biochim. Biophys. Acta* 1298, 95–101
102 Rees, D.C. and Lipscomb, W.N. (1982) Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 A resolution. *J. Mol. Biol.* 160, 475–498

103 Hass, G.M. *et al.* (1981) Specificity of the carboxypeptidase inhibitor from potatoes. *Plant Physiol.* 67, 754–758

104 Carter, T.H. *et al.* (1990) Cabbage seed protease inhibitor: a slow, tight-binding inhibitor of trypsin with activity toward thrombin, activated Stuart factor (factor Xa), activated Hageman factor (factor XIIa), and plasmin. *Blood* 75, 108–115

105 Ceciliani, F. *et al.* (1994) Purification, inhibitory properties, amino acid sequence and identification of the reactive site of a new serine proteinase inhibitor from oilrape (*Brassica napus*) seed. *FEBS Lett.* 342, 221–224

106 Ruoppolo, M. *et al.* (2000) Characterization of five new low-molecular-mass trypsin inhibitors from white mustard (*Sinapis alba* L.) seed. *Eur. J. Biochem.* 267, 6486–6492

107 Conners, R. *et al.* (2007) An unusual helix-turn-helix protease inhibitory motif in a novel trypsin inhibitor from seeds of Veronica (*Veronica hederifolia* L.). *J. Biol. Chem.* 282, 27760–27768

108 Kratzner, R. *et al.* (2005) Structure of *Ecballium elaterium* trypsin inhibitor II (EETI-II): a rigid molecular scaffold. *Acta Crystallog. D: Struct. Biol.* 61, 1255–1262

109 Favel, A. *et al.* (1989) Protease inhibitors from *Ecballium elaterium* seeds. *Int. J. Pept. Protein Res.* 33, 202–208

110 Thongyoo, P. *et al.* (2009) Potent inhibitors of beta-tryptase and human leukocyte elastase based on the MCoTI-II scaffold. *J. Med. Chem.* 52, 6197–6200

111 Gray, K. *et al.* (2014) Potent and specific inhibition of the biological activity of the type-II transmembrane serine protease matriptase by the cyclic microprotein MCoTI-II. *Thromb. Haemost.* 112, 402–411

112 Grzesiak, A. *et al.* (2000) Inhibition of serine proteinases from human blood clotting system by squash inhibitor mutants. *Biochim. Biophys. Acta* 1478, 318–324

113 Wieczorek, M. *et al.* (1985) The squash family of serine proteinase inhibitors. Amino acid sequences and association equilibrium constants of inhibitors from squash, summer squash, zucchini, and cucumber seeds. *Biochem. Biophys. Res. Commun.* 126, 646–652

114 Stachowiak, D. *et al.* (1996) Isolation and amino-acid sequence of two inhibitors of serine proteinases, members of the squash inhibitor family, from *Echinocystis lobata* seeds. *Acta Biochim. Pol.* 43, 507–513

115 Matsuo, M. *et al.* (1992) Trypsin inhibitors from bottle gourd (*Lagenaria leucantha* Rusby var. Depressa Makino) seeds. Purification and amino acid sequences. *Biochim. Biophys. Acta* 1120, 187–192

116 Hamato, N. *et al.* (1995) Trypsin and elastase inhibitors from bitter gourd (*Momordica charantia* LINN.) seeds: purification, amino acid sequences, and inhibitory activities of four new inhibitors. *J. Biochem.* 117, 432–437

117 Bendre, A.D. *et al.* (2018) Structural insights into the unique inhibitory mechanism of Kunitz type trypsin inhibitor from Ci*cer arietinum* L. *J. Biomol. Struct. Dyn.* 37, 2669–2677

118 Tornøe, C. *et al.* (2017) Divergent protein synthesis of Bowman–Birk protease inhibitors, their hydrodynamic behavior and co-crystallization with α-chymotrypsin. *Synlett* 28, 1901–1996

119 Gonzalez, C. *et al.* (2003) Structure and dynamics of the potato carboxypeptidase inhibitor by 1H and 15N NMR. *Proteins* 50, 410–422

120 Zhao, Q. *et al.* (2002) NMR solution structure of ATTp, an *Arabidopsis thaliana* trypsin inhibitor. *Biochemistry* 41, 12284–12296

121 Heitz, A. *et al.* (2008) Knottin cyclization: impact on structure and dynamics. *BMC Struct. Biol.* 8, 54

122 Rosengren, K.J. *et al.* (2003) Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. *J. Biol. Chem.* 278, 8606–8616

123 Touil, T. *et al.* (2008) Bowman-Birk inhibitor suppresses autoimmune inflammation and neuronal loss in a mouse model of multiple sclerosis. *J. Neurol. Sci.* 271, 191–202

124 Grundemann, C. *et al.* (2019) T20K: an immunomodulatory cyclotide on its way to the clinic. *Int. J. Peptide Res. Ther.* 25, 9–13

125 Koblinski, J.E. *et al.* (2000) Unraveling the role of proteases in cancer. *Clin. Chim. Acta* 291, 113–135

126 Losso, J.N. (2008) The biochemical and functional food properties of the Bowman-Birk inhibitor. *Crit. Rev. Food Sci. Nutr.* 48, 94–118

127 Clark, R.J. *et al.* (2012) Cyclization of conotoxins to improve their biopharmaceutical properties. *Toxicon* 59, 446–455

128 Clark, R.J. *et al.* (2005) Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13767–13772

129 Wan, X.S. *et al.* (1999) Treatment with soybean-derived Bowman Birk inhibitor increases serum prostate-specific antigen concentration while suppressing growth of human prostate cancer xenografts in nude mice. *Prostate* 41, 243–252

130 Zhang, Z. *et al.* (2017) Asparagine endopeptidase cleaves alpha-synuclein and mediates pathologic activities in Parkinson's disease. *Nat. Struct. Mol. Biol.* 24, 632–642

131 Mannisto, P.T. and Garcia-Horsman, J.A. (2017) Mechanism of action of prolyl oligopeptidase (PREP) in degenerative brain diseases: has peptidase activity only a modulatory role on the interactions of PREP with proteins? *Front. Aging Neurosci.* 9, 27

132 Myohanen, T.T. *et al.* (2012) A prolyl oligopeptidase inhibitor, KYP-2047, reduces alpha-synuclein protein levels and aggregates in cellular and animal models of Parkinson's disease. *Br. J. Pharmacol.* 166, 1097–1113

133 Waumans, Y. *et al.* (2015) The dipeptidyl peptidase family, prolyl oligopeptidase, and prolyl carboxypeptidase in the immune system and inflammatory disease, including atherosclerosis. *Front. Immunol.* 6, 387