

Juggling jobs: roles and mechanisms of multifunctional protease inhibitors in plants

Friederike M. Grosse-Holz, Renier A. L. van der Hoorn

Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, South Parks Road,
Oxford, OX1 3RB, UK

Author of correspondence: Renier A. L. van der Hoorn

renier.vanderhoorn@plants.ox.ac.uk

Tel +44 (0) 1865 275077

Key words (in alphabetical order): plant defence, protease inhibitors, protein structure, regulatory networks, resistance breeding, storage protein protection

Acknowledgements: We thank Thomas Roberts, Daniela Sueldo, Judith Paulus and Jan M. Brauner for critical reading. The authors are supported by the ERC Project 616449 'GreenProteases', University of Oxford and Somerville College, Oxford.

Word counts

Complete text: 8198

summary: 169; introduction: 438; main text: 5556; discussion: 652; box 1: 87; box 2: 121

Summary

Multifunctional protease inhibitors juggle jobs by targeting different enzymes and thereby often controlling more than one biological process. Here, we discuss biological functions, mechanisms and evolution of three types of multifunctional protease inhibitors in plants. The first type are double-headed inhibitors, which feature two inhibitory sites targeting proteases with different specificities (e.g. Bowman-Birk inhibitors) or even different hydrolases (e.g. α -amylase/protease inhibitors preventing both early germination and seed predation). The second type consists of multidomain inhibitors which evolved by intragene duplication and are released by processing (e.g. multicystatins and cyclotides, implicated in tuber dormancy and defence, respectively). The third type consists of promiscuous inhibitory folds which resemble mouse traps that can inhibit different proteases cleaving the bait they offer (e.g. serpins, regulating cell death, and α -macroglobulins). Understanding how multifunctional inhibitors juggle biological jobs increases our knowledge of the connections between networks they regulate. These examples show that multifunctionality evolved independently from a remarkable diversity of molecular mechanisms that can be exploited for crop improvement and provide concepts for protein design.

Introduction

Multifunctional protease inhibitors contradict the old dogma of one gene – one protein – one function (Beadle & Tatum, 1941). Multifunctionality implies that a protein or a structural fold can regulate various partners, resulting in different biological outputs. About 30 years after the influential paper of Beadle and Tatum, the first report of a multifunctional protease inhibitor (Odani & Ikenaka, 1973) challenged their model. A protein from soybean was separated, yielding two parts, one of which inhibited trypsin, the other chymotrypsin (Odani & Ikenaka, 1973). Trypsin and chymotrypsin are closely related Ser proteases, but they differ in the surface loops that determine substrate specificity (Hedstrom *et al.*, 1992) and thus require distinct inhibitory sites. This first bifunctional protease inhibitor belongs to the Bowman-Birk inhibitors (BBI), classified in the MEROPS database as family I12 (see Box 1 for a brief description of the MEROPS system). Only a few years after the discovery that a single protein can inhibit two distinct proteases, it became clear that the Indian staple crop Ragi (*Eleusine coracana*) produces an even more peculiar multifunctional inhibitor (Shivaraj & Pattabiraman, 1981). The Ragi Bifunctional Inhibitor (RBI, family I6) can form a trimeric complex with α -amylase and trypsin, thereby simultaneously inactivating a starch-degrading enzyme and a protease (Shivaraj & Pattabiraman, 1981). Two years later, it turned out

that a protein from barley also targets an α -amylase and a protease (Mundy *et al.*, 1983). However, BASI (Barley α -amylase/Subtilisin Inhibitor, family I3) is structurally unrelated to RBI. The bifunctional BBI, RBI and BASI were the first of many multifunctional plant protease inhibitors to be discovered. In these three cases, multifunctionality is facilitated by two inhibitory sites on a single protein. More recent research has characterized other types of multifunctional protease inhibitors. Some of these contain more than one inhibitory domain (multidomain inhibitors) or utilize the same binding site to inhibit one of various target proteases at a time (promiscuous inhibitors) (Fig. 1). The frequent occurrence of multifunctionality among protease inhibitors has raised many questions about the biological roles of these proteins. Are multifunctional inhibitors regulatory links between the proteases they affect? How has multifunctionality evolved in different inhibitor structures and do they share common “roads to multifunctionality”? How can multifunctional protease inhibitors be exploited for crop improvement and protein design? These questions are addressed in this review. With multifunctional plant protease inhibitors, we bring together a range of proteins and biological processes that are not usually studied jointly. Awareness of a potential hub between proteolytic networks will help to gain a comprehensive overview. Finally, understanding the inherent power of multifunctional protease inhibitors will prove useful to design successful agricultural or biotechnological strategies.

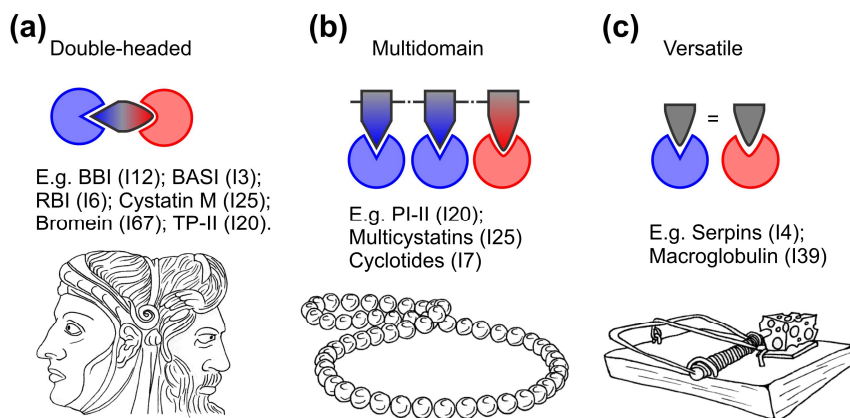


Fig. 1 Three classes of multifunctional protease inhibitors. BBI, Bowman–Birk inhibitor; BASI, barley amylase/subtilisin inhibitor, family I3; RBI, ragi bifunctional inhibitor.

Three types of multifunctionality

Juggling several biological jobs is facilitated in different ways, which we group into three types of inhibitor multifunctionality (Fig. 1). The first type of multifunctional inhibitors are Janus-type inhibitors, double

headed proteins with two inhibitory faces matching distinct targets. The second type are multidomain proteins, which resemble a necklace of pearls (inhibitor domains). Within the pearl necklace type, the I20 inhibitor domains have diversified with regard to specificity, while the circular cyclotide peptides can have both inhibitory and non-inhibitory functions. The repetitive nature of the pearl-necklace type allows multicystatins to crystallize *in vivo*, which seems to facilitate tightly regulated release of the inhibitory cystatin domains when necessary in tuber formation or defence. The third type of multifunctionality is represented by promiscuous inhibitors such as serpins and α -macroglobulins. These structurally different families act as mouse traps that are able to inhibit different proteases cleaving the bait they offer.

We summarize the current scientific knowledge on biological role, inhibitory mechanism and evolution for eight families of multifunctional plant protease inhibitors. A comprehensive overview of all discussed families and the examples mentioned throughout this review is given in Table 1. The inhibitors are first discussed in groups based on their type of multifunctionality. We conclude by discussing all eight multifunctional inhibitor families in the context of common biological roles, evolutionary history and potential applications.

1. Double-headed inhibitors: the Janus-type

With his two faces looking in opposite directions, the ancient Roman God Janus provides a good metaphor for bifunctional inhibitors. Each of the two faces stands for a binding site on which a target enzyme can be inhibited. Surprisingly many protein architectures facilitate multifunctionality by providing two sites with different inhibitory specificity, thus the Janus-type inhibitors include representatives from at least four MEROPS families that occur in plants (Table 1).

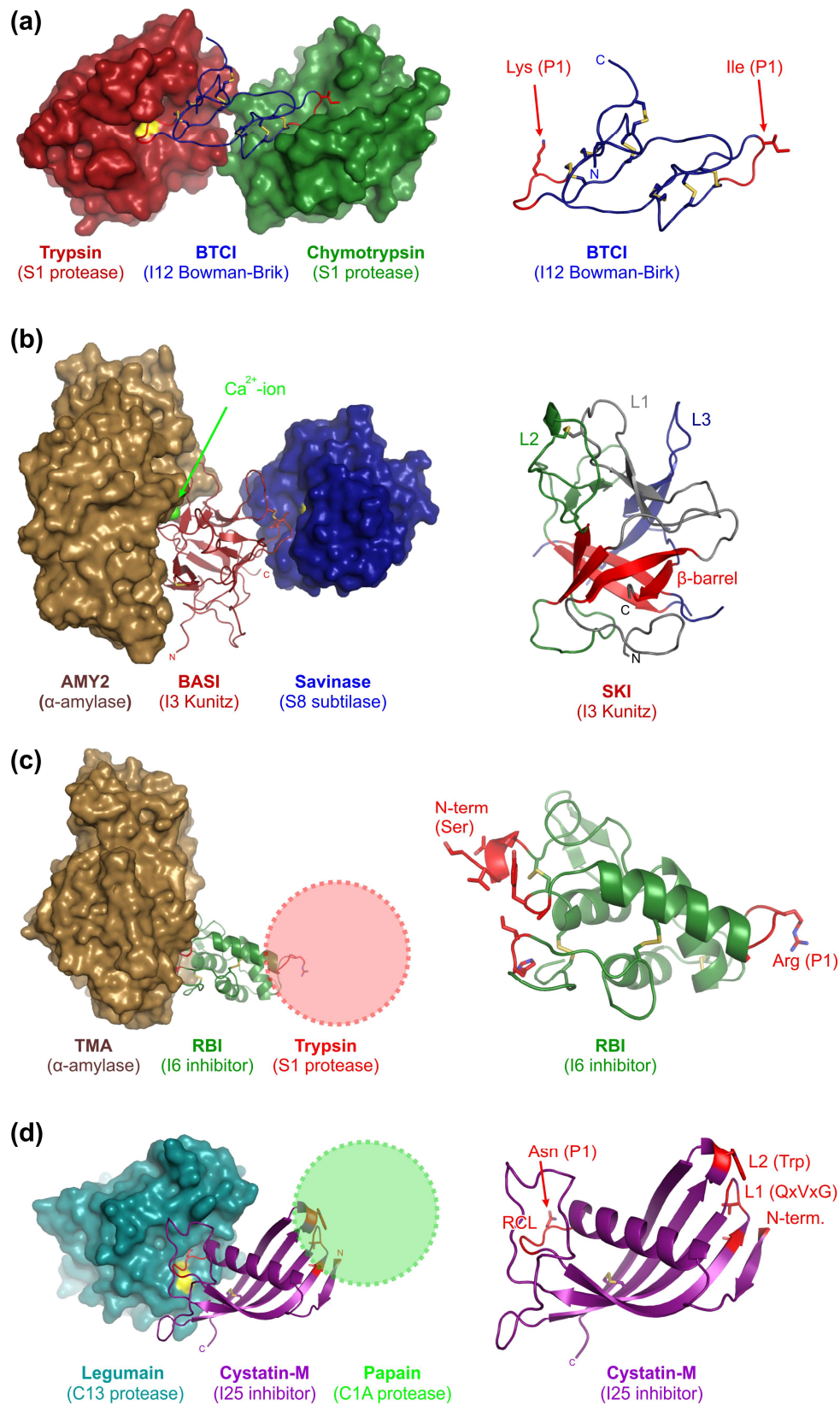


Fig. 2 Four examples of double-headed inhibitors. (a) Left panel, crystal structure (PDB ID3ru4) of *Vigna unguiculata* Bowman–Birk trypsin and chymotrypsin inhibitor (BTCL, family I12; purple) in complex with bovine trypsin (family S1; red) and bovine chymotrypsin (family S1; green). Right panel, BTCL inhibits trypsin and chymotrypsin using loops (red) that contain Lys and Ile residues mimicking the P1 substrate recognition sites of the respective target enzymes. (b) Left panel, crystal structures of *Hordeum vulgare* (barley) α -amylase/subtilisin inhibitor (BASI, family I3; PDB ID 3bx1) in complex with barley α -amylase (AMY2; brown; PDB ID 1ava) and *Bacillus lentus* savinase (blue; family S8; PDB ID 3bx1). A calcium ion (green) acts as molecular glue in the AMY2–BASI interaction. Right panel, the Glycine max (soybean) Kunitz inhibitor (SKI, family I3; PDB ID 1avu) illustrates the overall structure of Kunitz inhibitors: a β -barrel (red) with three extended loops (L1–L3), representing a tree trunk and branches, respectively. (c) Left panel, crystal structure of ragi bifunctional inhibitor (RBI, family I6; PDBID 1tmq; green) from *Eleusine coracana* (ragi or Indian finger millet) with *Tenebrio molitor* (yellow meal worm) α -amylase (TMA; brown; PDB ID 1tmq). The location where a trypsin-like protease could bind is indicated by a red circle. Right panel, inhibition is facilitated by the N-terminal Ser residue of RBI (red) that interacts with the active site of TMA. The loop that can interact with trypsin-like proteases (family S1) contains an Arg residue (red) that mimics the P1 substrate recognition motif for trypsin. (d) Left: crystal structure (PDB ID 4n6o) of human cystatinM (family I25; purple) with human legumain (family C13; cyan). The location where a papain-like protease could bind is indicated by a green circle. Right panel, legumains specifically cleave after Asn (P1 = Asn), and therefore selectivity of the interaction is facilitated by an Asn residue in the reactive centre loop (RCL; red). The three regions of cystatinM that would bind the substrate-binding groove of papain-like Cys proteases are located on the opposing side (red).

1.1 Bowman-Birk inhibitors (I12): losing and gaining multifunctionality

The first Janus-type inhibitor was discovered in soybean flour by Bowman in 1946 (Bowman, 1946). The inhibitor was further purified and characterized by Yehudith Birk, who showed that it inhibits both trypsin and chymotrypsin (Birk, 1961). However, it was not yet clear that this was due to two binding sites. Separation of the protein in two fragments which retained inhibitory activity towards either trypsin or chymotrypsin, respectively, elucidated that the soybean Bowman-Birk inhibitor (BBI, MEROPS family I12, see Box 1 for a brief description of the MEROPS system) has two independent binding sites (Odani & Ikenaka, 1973), making the soybean BBI the first known multifunctional protease inhibitor. BBI encoding genes are also present in *Medicago*, rice and maize, but they seem absent in the model plants *Arabidopsis thaliana* and *Nicotiana benthamiana* (Rawlings *et al.*, 2014). BBIs may have a defensive function, because

BBI gene expression in rice is upregulated in response to wounding and the defence-related phytohormone jasmonic acid (Rakwal *et al.*, 2001; Qu *et al.*, 2003) and overexpression of an endogenous Janus-type BBI in the staple crop rice increases resistance to *Magnaporthe grisea*, a fungal pathogen causing rice blast (Qu *et al.*, 2003). BBIs fold into a core of antiparallel beta-sheets that is crosslinked by multiple disulphide bridges (seven in the case of soybean BBI). The two inhibitory sites are located on protruding loops on opposing ends of the beta-sheet core (Fig. 2a) and function via the Laskowski mechanism (Box 2) (Chen *et al.*, 1992; Voss *et al.*, 1996). Monocot I12 inhibitors contain up to three BBI domains (Qu *et al.*, 2003), thus they should in theory have up to six protease inhibitory sites. However, the monocot BBI domain has lost a disulphide bridge restraining the conformation of one of the inhibitory loops. Having lost the inhibitory activity of one of the two subdomains, monocot BBIs bind only one target protease per BBI domain. Apparently, the loss of the second inhibitory site in monocots was corrected by domain duplication, so that contemporary monocot BBIs can bind multiple proteases (Song *et al.*, 1999; Park *et al.*, 2004). Surprisingly, though BBIs can inhibit trypsin and chymotrypsin, it is still unknown what the natural targets of these seed proteins are. Identification of both endogenous (plant) and exogenous (insect/bacterial/fungal) target proteases might increase our understanding of the apparent evolutionary pressure towards multifunctional plant BBIs.

1.2 Kunitz inhibitors (I3): a very versatile fold

Ser protease inhibitors from the Kunitz (I3) family exist in most higher plants, but seem to be absent from green algae genome sequences according to the current release of the MEROPS data base (Rawlings *et al.*, 2014). The inhibitors were named after Moses Kunitz, who crystallized the first representative from soybean flour (Kunitz, 1945). The I3 family includes Janus-type inhibitors that bind different target enzymes on two reactive sites, e.g. the barley α -amylase/subtilase inhibitor BASI and its rice orthologue, both of which occur in grains (Leah & Mundy, 1989; Yamagata *et al.*, 1998). α -amylase/subtilase inhibitors are believed to regulate germination, as they inhibit endogenous α -amylases (Abdul-Hussain & Paulsen, 1989; Vallée *et al.*, 1998; Nielsen *et al.*, 2003), which mobilize storage carbohydrates during germination (Fincher, 1989). Rice α -amylase/subtilase inhibitors in rice also block α -amylases from insects that feed on the grain starch (Bellincampi *et al.*, 2004). Using its second reactive site, BASI inhibits proteases from the fungal pathogen *Fusarium culmorum* (Pekkarinen *et al.*, 2007), also indicating defensive functions that make α -amylase/subtilase inhibitors interesting candidates for crop improvement. Another member of the Kunitz family, ApKTI (*Adenanthera pavonina* Kunitz type inhibitor), occurs in seeds of the leguminous tree *Adenanthera pavonina* contain. ApKTI can inhibit both trypsin (S1) and papain (C1) simultaneously

(Migliolo *et al.*, 2010) and is active against gut proteases from herbivorous insects, including beetles and moths ((Da Silva *et al.*, 2014), and references therein). Artificial diets containing ApKTI reduce the viability and fertility of these insects, indicating that ApKTI acts in defence against herbivorous insects. ApKTI is a promising candidate for crop improvement, as the multifunctionality of the inhibitor may help to impede insect adaptation (Da Silva *et al.*, 2014). Potato tubers also contain a Janus-type Kunitz type inhibitor, named Potato Serine Protease Inhibitor (PSPI). PSPI can bind simultaneously to both trypsin and chymotrypsin, two S1 Ser proteases with different substrate specificities (Valueva *et al.*, 2000; Meulenbroek *et al.*, 2012). The biological role of PSPI *in planta* is unclear. In principle, PSPI could act in defence or protect storage proteins from endogenous proteases to prevent premature sprouting. Considering that PSPI is one of the most abundant proteins in potato tubers (Meulenbroek *et al.*, 2012), it probably also serves as a storage protein itself. Recently, a new biochemical function was proposed for KTI3 (Kunitz trypsin inhibitor 3), a *Populus deltoides* (poplar) inhibitor from the Kunitz family active against trypsin and chymotrypsin (Major & Constabel, 2008). KTI3 is expressed *in planta* upon exposure to heavy metals and this protein confers heavy metal resistance when expressed in transgenic yeast. Molecular modelling suggests that KTI3 can chelate copper ions, but this remains to be demonstrated experimentally (Guerra *et al.*, 2015). Multifunctionality in Kunitz inhibitors is not always achieved via the double-headed Janus structure. For instance, a Kunitz inhibitor from *Prosopis juliflora*, a South American shrub, can inhibit either trypsin (family S1) or papain (family C1) using overlapping binding sites (Franco *et al.*, 2002). All Kunitz (I3) inhibitors share a tree-like tertiary structure called the β -trefoil fold (Fig. 2b). The “tree” consists of a β -barrel (the trunk) with flexible loops protruding from each side (branches and roots, respectively) (Sweet *et al.*, 1974). The protease binding sites of β -trefoil inhibitors are located in the loops (Azarkan *et al.*, 2011) and inhibition occurs via the Laskowski mechanism (Box 2) (Renko *et al.*, 2012), with the notable exception of the α -amylase/subtilase inhibitors.

BASI has been crystallized in complex with savinase, a subtilase from the soil bacterium *Bacillus lentus* (Micheelsen *et al.*, 2008). Structural and mutational analysis of this complex revealed that the BASI inhibitory mechanism for subtilases differs from the canonical Laskowski mechanism. The protease inhibitory loop of BASI is shorter than usual and pulled out of the protease active site by a disulphide bridge. Thus, BASI cannot be cleaved by the subtilase, which may result in a more stable inhibitory complex. The cysteine residues facilitating this version of the Laskowski mechanism are conserved in the rice and wheat orthologues of BASI, suggesting that they function in the same way (Micheelsen *et al.*, 2008). The inhibitory sites of BASI are known to be independent because protease inhibition still occurs when the inhibitor is saturated with α -amylase (Mundy *et al.*, 1983). Inhibition of barley α -amylase occurs

via a large binding interface and involves a fully hydrated Ca^{2+} ion. This way, BASI sterically hinders the access to the active site of the α -amylase (Fig. 2b) (Vallée *et al.*, 1998). Since affinity of BASI for barley α -amylase increases with increasing pH ((Nielsen *et al.*, 2003) and references therein), it is thought that the protonation state of the amino acid side chains involved in Ca^{2+} binding may be affected by pH changes (Nielsen *et al.*, 2003). During germination, a pH increase might release α -amylase from BASI (Mundy *et al.*, 1983; Vallée *et al.*, 1998; Nielsen *et al.*, 2003).

Length, orientation and amino acid composition of the loops vary between the β -trefoil inhibitors, allowing them to inhibit up to two proteases of different classes (e.g. (Dattagupta *et al.*, 1999; Meulenbroek *et al.*, 2012)). Protease families inhibited by Kunitz inhibitors include S1, S8, C1 and A1 (Azarkan *et al.*, 2011; Rawlings *et al.*, 2014). Although the overall structure is highly conserved, the inhibitory motifs of the Kunitz inhibitors are diversified across the family. The core β -barrel, the conserved element of the Kunitz fold, is a rigid structure that presumably confers stability to the Kunitz inhibitors in harsh environments. The β -barrel can fold via various routes, rendering the folding process of Kunitz proteins relatively immune to point mutations in the loops. Taken together, it appears that the stable β -trefoil fold provided a platform for evolution of a wide range of molecular recognition mechanisms (Azarkan *et al.*, 2011). In plants, this includes the inhibition of amylases and proteases via various mechanisms. In other organisms, the Kunitz fold with its versatile loops is shared by growth factors, Interleukins and DNA binding proteins (Renko *et al.*, 2012).

1.3. RBI (I6): bifunctionality by two versions of substrate mimicry

Ragi bifunctional inhibitor (RBI) was first purified from seeds of the Indian finger millet, locally called Ragi (*Eleusine coracana*) (Shivaraj & Pattabiraman, 1981). Like BASI, RBI inhibits both insect and mammalian α -amylases as well as proteases (Maskos *et al.*, 1996; Strobl *et al.*, 1998). Astonishingly, BASI and RBI are unrelated in terms of sequence, structure and inhibitory mechanisms, thus inhibition of different hydrolase classes using one multifunctional protein must have evolved twice independently. RBI adopts an all- α -fold (Fig. 2c), which is unrelated to any other known protease inhibitor family (Strobl *et al.*, 1995, 1998). The trypsin binding site of RBI forms a loop that can be cleaved by bovine trypsin, but RBI remains inhibitory afterwards (Strobl *et al.*, 1995), confirming that protease inhibition follows the canonical Laskowski mechanism (Box 2) (Maskos *et al.*, 1996). Structural analysis of a complex between RBI and α -amylase from yellow meal worm (*Tenebrio molitor*) revealed that the amylase inhibitory mechanism differs remarkably between I3 and I6 inhibitors. RBI inserts its N-terminus into the amylase active site, almost completely filling the catalytic cleft and directly interacting with the active site residues (Strobl *et*

al., 1998). The inhibitor can be displaced by large substrates (>7 saccharide units), confirming that RBI acts in a competitive, substrate-like manner on α -amylase (Maskos *et al.*, 1996). RBI can form a ternary complex with bovine trypsin and porcine α -amylase (Maskos *et al.*, 1996), confirming that the binding sites for the protease and amylase are independent, as is the case for BASI. The extensive *in vitro* studies on the RBI inhibitory mechanisms have, to our knowledge, not been complemented with studies on the role of this bifunctional inhibitor *in vivo*. For instance, it is unknown whether RBI inhibits endogenous α -amylases as well as pathogen enzymes and how important this is for germination and defence. The structural similarity between RBI and seed storage proteins, the albumins, provokes the speculation that this Janus-type inhibitor evolved from the small, stable albumin fold by acquiring additional inhibitory activities.

1.4 Legumain-inhibiting cystatins (I25A): More than just an extension

Plant representatives of the cystatin family (I25) are called phytocystatins (MEROPS family I25A, see Box 1 for a brief description of the MEROPS system) and occur in a wide range of taxonomic groups, including green algae, mosses, monocots and dicots (Martinez & Diaz, 2008). Most phytocystatins have a molecular weight of 12 – 16 kDa and inhibit Cys proteases of the family C1. Higher plants contain additional C-terminally extended cystatins with a molecular weight of ca 24 kDa (Martinez & Diaz, 2008). The C-terminal extension is a second inhibitory site, specific for C13 proteases, called aspariginyl endopeptidases (AEPs), legumains or vacuolar processing enzymes (VPEs) (Martinez *et al.*, 2007). Presence of both C1 and C13 proteases is necessary *in vitro* for complete degradation of the bean storage protein phaseolin (Zakharov *et al.*, 2004). To prevent germination, C1 and C13 proteases may thus be co-regulated by legumain-inhibiting cystatins. Indeed, overexpression of the C-terminally extended cystatin AtCYS6 in *Arabidopsis* causes reduced Cys protease activity and delayed germination, whereas *atcys6* null mutant seeds exhibit higher Cys protease activity and germinate early (Hwang *et al.*, 2009). It remains unclear, however, whether this requires the C1 or the C13 protease inhibitory activity, or both. Bifunctional cystatins may be efficient regulators of germination, as they are able to prevent both the C1- and the C13-dependent steps of storage protein hydrolysis.

Bifunctional cystatins target different proteases via different mechanisms, namely steric hindrance of C1 proteases and substrate mimicry for C13 proteases. The mechanism for C1 protease inhibition by cystatins is conserved between animals and plants (Rawlings *et al.*, 2014) and is based on a tripartite wedge consisting of the N-terminus and two hairpin loops carrying the conserved QxVxG motif (Fig. 2d). This wedge is inserted into the active site of the C1 protease in a tight, but reversible interaction (Stubbs *et al.*,

1990). C13 protease inhibition by phytocystatins involves substrate mimicry using an exposed Asn residue following an α -helix on the C-terminal extension. Interestingly, animal cystatins can inhibit legumains (AEPs/VPEs) without an extension, as they harbour such an Asn residue next to a helix on the back of the cystatin domain itself (Alvarez-Fernandez *et al.*, 1999). Bifunctional phytocystatins may thus have evolved from an ancient domain duplication, upon which each of the two cystatin domains lost one of the inhibitory activities (Martinez *et al.*, 2007).

2. Multidomain Inhibitors: the pearl necklace

The pearl necklace type inhibitors consist of an array of protease inhibitory domains. Their inhibitory potential is increased through cleavage, which releases the inhibitory domains like pearls from a string. Gene duplication probably gave rise to these multidomain proteins. Notably, a new folding pathway led to reorganization of the inhibitory domains of I20 inhibitors. Cyclization of the multidomain precursor or the released inhibitory peptides occurs for I20 and I7 inhibitors, respectively. The accumulation of structurally similar domains in one protein can also facilitate new regulatory mechanisms, as seen with the multicystatins (I25).

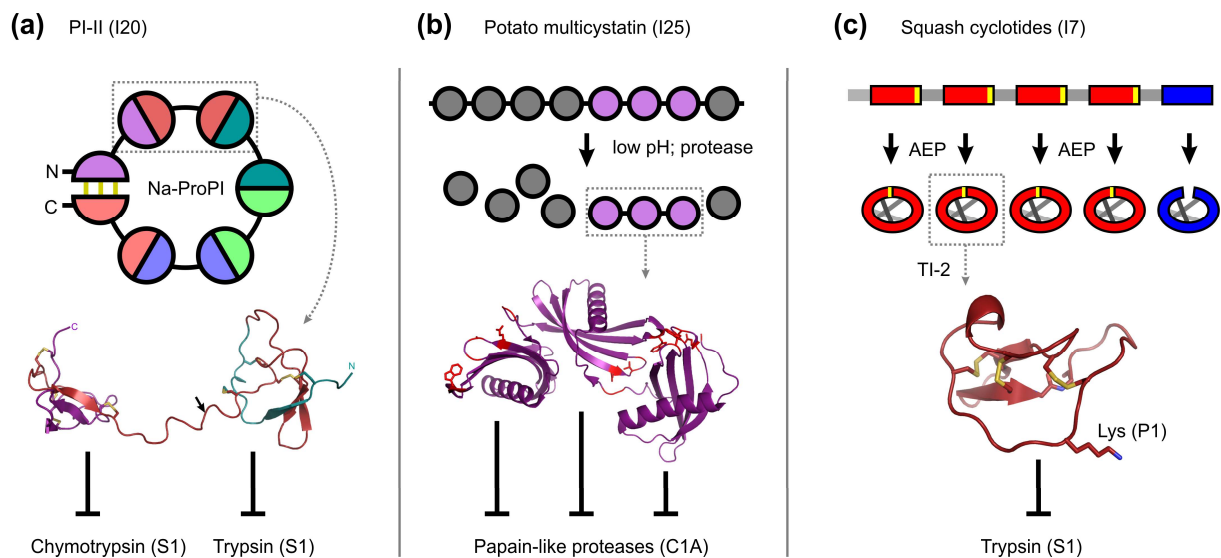


Fig. 3 Folding and processing of multidomain inhibitors. (a) Folding and processing of the PI-II inhibitor of *Nicotiana glauca* (Na-ProPI, family I20). Na-ProPI 19

contains six sequence repeats (highlighted in different colours) that fold into six hybrid inhibitory domains (depicted as pearls). One domain consists of the N- and C-termini linked by three disulphide bridges. The inhibitory domains are released through proteolytic cleavage by Cys proteases (presumably family C13). The nuclear magnetic resonance (NMR) structure of the second and third domains with the cleavable linker between them has been resolved (bottom; PDB ID 1fyb). The cleavage site for release of the two domains, which target chymotrypsin and trypsin (both family S1), respectively, is indicated by a black arrow. (b) Folding and processing of potato multicystatin (PMC, family I25). The precursor consists of eight cystatin domains (depicted as pearls) that are released through proteolytic cleavage by Ser proteases at low pH. Domains 5–7 remain together and have been crystallized (PDB ID 4lzi). All released cystatin domains target papain-like Cys proteases (family C1) using a tripartite wedge inhibitory site (highlighted in red). (c) Folding and processing of squash inhibitors (family I7). The precursor consists of five sequence repeats, each but the last ending in the aspariginyl endopeptidase (AEP)/vacuolar processing enzyme (VPE) cleavage site Asp or Asn, indicated in yellow. Four of the released peptides (red) undergo cyclization by AEP/VPE, while the last repeat (blue) remains acyclic. All five released squash inhibitors share the cystine knot motif with three disulphide bridges (grey). The crystal structure of the macrocyclic knottin MCoTI-II (PDB ID 4gux) shows the disulphide bridges linking the central loop to the periphery (yellow) and the Lys residue recognized by trypsin (family S1).

2.1 Multidomain Potato peptidase inhibitor II (I20): cyclization shuffles domains

Already in the 70s, the Japanese scientist Teruo Iwasaki determined the amino acid sequence of a fragment of potato peptidase inhibitor II (PI II) that inhibited trypsin (S1) and, to a lesser extent chymotrypsin (S1) and subtilisin (S8) (Iwasaki *et al.*, 1976). Similar proteins occur in tomato, where their production is induced by wounding (Graham *et al.*, 1985). Today, it is clear that inhibitors of the I20 family are present in most monocots and dicots including maize, Arabidopsis, poplar and many solanaceous species. Most plants appear to only have one I20 inhibitor gene, whereas tomato and potato contain more homologues (Rawlings *et al.*, 2014). The I20 inhibitors are clearly associated with plant defence. Their expression increases upon wounding (Graham *et al.*, 1985; Kong & Ranganathan, 2008) and they are constitutively made in reproductive organs, which seem well worthy of special protection (Atkinson *et al.*, 1993). Furthermore, I20 inhibitors from pepper inhibit the gut proteases of the cotton bollworm (*Helicoverpa armigera*) *in vitro* (Tamhane *et al.*, 2005; Joshi *et al.*, 2014) and overexpression of different

members of the I20 family confers increased resistance to insect pests in various species (Johnson *et al.*, 1989; Tamhane *et al.*, 2009; Dunse *et al.*, 2010; Joshi *et al.*, 2014), highlighting the potential of I20 inhibitors for crop improvement.

Structure and fold of the I20 inhibitors are prime examples for the fascinating intricacy of plant defence proteins. The *Nicotiana glauca* I20 inhibitor precursor NaProPI is a well-studied example (Kong & Ranganathan, 2008). NaProPI is a 43 kDa protein produced at high levels in the female reproductive tissues of *N. glauca*. The amino acid sequence of NaProPI consists of six homologous sequence repeats flanked by an N-terminal signal peptide and a C-terminal vacuolar targeting signal (Atkinson *et al.*, 1993). Both signal peptides are removed during posttranslational processing and eventually, the six-repeat precursor is cleaved into six individual protease inhibitors of 6 kDa each (Fig. 3a). Two of the inhibitors released from NaProPI inhibit chymotrypsin and four are specific for trypsin (Lee *et al.*, 1999). Each of the released inhibitors contains eight Cys residues forming four disulphide bonds and adopts a compact, stable fold (Nielsen *et al.*, 1995) (Fig. 3a). All I20 inhibitors utilize the Laskowski mechanism (Box 2) (Greenblatt *et al.*, 1989), but their *in vivo* target proteases remain unknown. Five of the 6 kDa inhibitors released from NaProPI are single-chain peptides, but the sixth one consists of two chains held together by three disulphide bridges (Lee *et al.*, 1999). This peculiarity arises from the way in which NaProPI is processed into the 6 kDa mature inhibitors. Interestingly, the underlying cleavage sites are located not between, but within each of the sequence repeats of NaProPI (Heath *et al.*, 1995). After removal of the signal peptides, the NaProPI precursor adopts a cyclic conformation, linking the N-terminus to the C-terminus of the peptide chain via three disulphide bridges. This circle is then cleaved once within each sequence repeat, releasing five single-chain inhibitors as well as one two-chain protein which contains the N- and C-terminal ends of the NaProPI precursor (Lee *et al.*, 1999). NaProPI can be processed *in vitro* using C13 endopeptidases (Heath *et al.*, 1995), suggesting that asparaginyl endopeptidases (AEPs/VPs/legumains, family C13) might release the 6 kDa inhibitors *in vivo*.

The peculiar mechanism of the NaProPI maturation prompts the question how this protein might have evolved. It is very likely that the ancestor of the I20 family consisted of a single sequence repeat folding into a single domain inhibitor. Sequence duplication events then gave rise to multi-repeat proteins that eventually adopted the cyclic fold and processing sites within each sequence repeat. This scenario is endorsed by the finding that a single ancestral NaProPI sequence repeat can fold into a functional inhibitor when it is expressed in bacteria (Scanlon *et al.*, 1999). As soon as more than one repeat is present, however, the artificial ancestral I20 inhibitor adopts a cyclic conformation and is processed within the repeats (Lee *et al.*, 1999). Notably, not all of the I20 inhibitors need to be cleaved in order to be active.

The two-domain tomato inhibitor II has been crystallized in a ternary complex with two subtilisin molecules, showing that both inhibitory sites are accessible simultaneously (Barrette-Ng *et al.*, 2003). Analysis of nonsynonymous/synonymous mutation rates within the whole I20 family revealed that the cysteine scaffold that determines the structure of the cyclic precursors as well as of the released inhibitors is under purifying selection. This underscores the evolutionary pressure towards a cyclic precursor that may be explained by its elevated thermodynamic stability. The active residue of the released inhibitors, however, is under diversifying selection, presumably leading to specificity for different target proteases (Kong & Ranganathan, 2008).

2.2 Cyclotides (I7 and non-inhibitory versions): a very tight knot

Cyclotides are peptides of up to 50 amino acids, stabilized by multiple disulphide bridges in the cystine knot motif. Most cyclotides have a circular backbone and the ones we discuss here are released from multidomain precursors (Fig. 3c). Not all cyclotides are protease inhibitors, and based on their structure and activity, cyclotides are subdivided into three groups: Moebius, bracelet and trypsin inhibitor (Craik *et al.*, 1999; Burman *et al.*, 2014; Mahatmanto, 2015). The Moebius cyclotides are cyclized with a twisted backbone, resulting in the Moebius strip structure, whereas bracelet cyclotides are closed without a twist (Craik *et al.*, 1999). Representatives of the Moebius group were first discovered in *Oldenlandia affinis*, an African plant used in traditional medicine (Gran, 1973). The non-inhibitory cyclotides purified from *O. affinis*, locally known as “Kalata-kalata”, are named kalata B1 to B7 (Jennings *et al.*, 2001). Other cyclotides of the bracelet and Moebius groups are widespread in angiosperms (Mahatmanto, 2015). The trypsin inhibitory cyclotides (MEROPS family I7) occur in the seeds of *Cucurbitaceae* and are therefore called squash inhibitors. The squash inhibitors include both cyclic and acyclic cystine knot peptides (Hernandez *et al.*, 2000).

Given their abundance and strong inhibitory activity against trypsin-like proteases, which are used by mammalian, insect and fungal seed predators, squash inhibitors might be involved in defending seeds against predation (Burman *et al.*, 2014; Mahatmanto, 2015), but a role as storage proteins in seeds has also been envisaged (Mahatmanto *et al.*, 2015). To our knowledge, no protease has yet been identified as a natural target of inhibitory cyclotides.

The kalata (*O. affinis*) cyclotides in the Moebius group impair growth and development of insect larvae, but do not inhibit trypsin, chymotrypsin or α -amylases from insect guts (Jennings *et al.*, 2001). Instead, kalata cyclotides penetrate and thereby destroy membranes of insect gut cells (Barbeta *et al.*, 2008).

The cystine knot motif that stabilizes cyclotides is based on a hairpin of two antiparallel β -strands containing three cysteines linked to the periphery via three disulphide bridges. A third, non-standard β -strand is present in one of the loops, which are otherwise not canonically structured (Fig. 3c) (Saether *et al.*, 1995; Craik *et al.*, 1999; Mylne *et al.*, 2012). Squash inhibitors and kalata cyclotides are processed from repetitive precursors, yielding up to eight mature peptides from each precursor. The released peptides are cyclized via transpeptidation by asparaginyl endopeptidases, also called vacuolar processing enzymes (AEPs/VPEs, family C13) (Saska *et al.*, 2007; Gillon *et al.*, 2008; Mylne *et al.*, 2012). In the case of the squash inhibitors, the last repeat remains acyclic, but nevertheless inhibits trypsin (Mylne *et al.*, 2012). The squash inhibitors are members of the I7 family that shares the Laskowski mechanism of inhibition (see Box 2) (Otlewski & Zbyryt, 1994; Hernandez *et al.*, 2000). The cystine knot and the AEP/VPE-mediated backbone cyclization render cyclotides resilient to boiling or proteolysis, unless the disulphide bridges are reduced prior to the treatment (Craik *et al.*, 1999). Interestingly, cyclization of cystine knot peptides seems to have evolved several times in parallel. The acyclic cystine knot inhibitor in *Momordica*, a subtropical cucurbit genus, is conserved as a single-domain gene in a wide range of *Momordica* species, while the multidomain precursors encoding cyclic cystine knot inhibitors are only found in a subgroup of related species, indicating they may have arisen through recent gene duplication events (Mahatmanto *et al.*, 2015). However, cyclic cystine knot peptides also occur in the Rubiaceae, Violaceae, Fabaceae and Solanaceae. The corresponding precursors differ between species with regard to the number of cyclotides and the amount of non-cyclotide sequence they contain (Mylne *et al.*, 2012). All of these precursors share the recognition sites necessary for cleavage and cyclization by AEPs/VPEs (Hara-Nishimura *et al.*, 1991; Hiraiwa *et al.*, 1999) and a Gly residue needed to form a circle (Mylne *et al.*, 2011). The ability of AEPs/VPEs to cyclize peptides may have acted as an evolutionary channel, lending the selective advantage of increased stability to cystine knot peptides that start with a Gly and end in Asp/Asn (Mylne *et al.*, 2012). Cyclotides are an excellent template for artificial multifunctionalization. Different engineered MCoTI-II (*Momordica cochinchinensis* trypsin inhibitor-II) cyclotides specifically inhibit proteases from different catalytic classes, namely chymotrypsin (family S1, K_i in nM range), subtilisin (family S8, K_i in μ M range), a viral Cys protease (family C3, K_i in μ M range) (Thongyoo *et al.*, 2008), tryptase and leukocyte elastase (both family S1) (Thongyoo *et al.*, 2009).

2.3 Multicystatins (I25A): release at the right moment

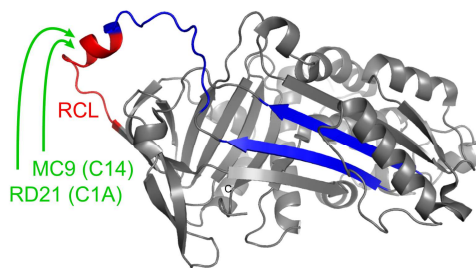
Phytocystatins form a separate subfamily I25A within the cystatin family I25 (Margis *et al.*, 1998; Rawlings *et al.*, 2014). Phytocystatins inhibit papain-like Cys proteases (family C1) and are common among plants from green algae to Arabidopsis, tomato, rice and maize (Rawlings *et al.*, 2014). Some members of the phytocystatin family also inhibit C13 proteases using a C-terminal extension, as discussed in section 1.4. In this paragraph, we focus on the numerous cases where phytocystatins occur as multidomain proteins (Rawlings *et al.*, 2014) consisting of up to eight inhibitory cystatin domains (Walsh & Strickland, 1993). Multidomain cystatins (multicystatins) play various biological roles in both potato and tomato. The potato multicystatin (PMC) occurs in potato tubers and inhibits tuber proteases (Kumar *et al.*, 1999). PMC is produced during early stages of tuber formation, associated with a decrease in proteolytic activity in the tuber, which may facilitate storage protein accumulation. Indeed, accumulation of PMC up to 12% of total soluble tuber protein precedes the accumulation of patatin, the main storage protein in potatoes that makes up 40% of the soluble protein content (Mignery *et al.*, 1988; Pouvreau *et al.*, 2001; Weeda *et al.*, 2009). In ageing tubers, proteolytic activity increases again, coinciding with a drop in the detectable levels of PMC (Kumar *et al.*, 1999). Thus, PMC abundance appears to regulate the process from building up the protein reserves in potato tubers to their mobilisation during ageing and germination. Multicystatin expression in tomato leaves is induced by various elicitors of plant immunity (Siqueira-Júnior *et al.*, 2002; Uppalapati *et al.*, 2005; Girard *et al.*, 2007a) and tomato multicystatin inhibits proteases from insect digestive tracts and impairs the growth of plant pathogenic fungi *in vitro* (Siqueira-Júnior *et al.*, 2002). Similar experiments show that potato multicystatin inhibits the growth of corn rootworm larvae, predators of an important food crop, when added to an artificial diet (Orr *et al.*, 1994). Multicystatins have the intriguing capacity to crystallize in native tissues. Cys protease inhibiting crystals were first observed in potato tubers (Cohn, F., 1859; Rodis & Hoff, 1984), but they also occur in tomato leaves (Akers & Hoff, 1980). Multicystatin crystals localize to the cytosol in potato (Nissen *et al.*, 2009) as well as in tomato (Madureira *et al.*, 2006). The transition between crystalline and soluble states for PMC is now quite well understood. Solubilisation of PMC occurs at mildly acidic pH, exposing the inhibitory domains to target cysteine proteases (Orr *et al.*, 1994). Recent structural analyses show that low pH weakens the interdomain interactions in PMC, explaining this regulatory mechanism (Green *et al.*, 2013). As soon as it is soluble, PMC can be cleaved by serine proteases (Walsh & Strickland, 1993) to release three fragments, which collectively contain eight cystatin domains. Inhibition of Cys proteases by each of the eight domains occurs via the cystatin mechanism (section 1.4). The acidic environment of insect

digestive tracts is thought to activate the inhibitor exactly when and where it is needed to hinder the protein catabolism of the insect pathogen, but not the host plant (Green *et al.*, 2013). Multicystatins presumably evolved from an ancestral, single-domain plant cystatin via gene duplications. Although it is known that phytocystatins inhibit papain-like cysteine proteases (family C1) (Tajima *et al.*, 2011), it is yet unknown which proteases are regulated by multicystatins (Benchabane *et al.*, 2010).

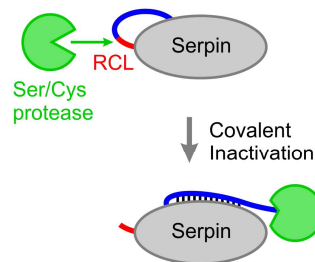
3. Promiscuous inhibitory folds: the mouse trap type

In this section, we describe two protease inhibitor families that utilize a mechanical trapping mechanism to sequester their target proteases. The nature of these mechanisms is destructive: inhibition is irreversible and dooms both the protease and the inhibitor for degradation. The active site that is recognized by the enzyme can vary in mouse trap type inhibitors without affecting functionality of the trap. Accordingly, mouse trap type inhibitors have developed multifunctionality in the sense that the same inhibitory fold can be used to target different proteases.

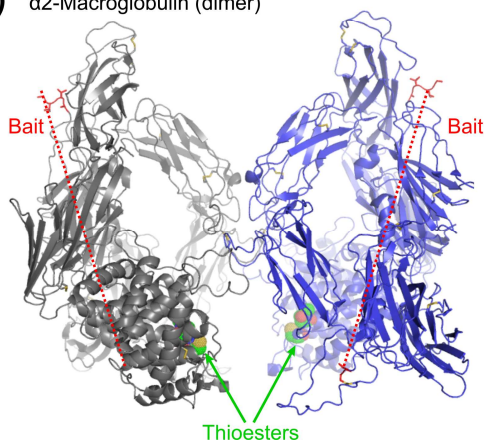
(a) AtSerpin1 (I4)



(b)



(c) α 2-Macroglobulin (dimer)



(d)

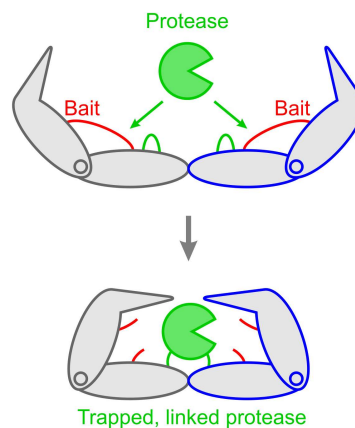


Fig. 4 Serpin and α 2-macroglobulin, the promiscuous mouse traps. (a) Top pane, crystal structure of the serpin AtSerp1 (family I4; PDB ID 3le2) showing the reactive centre loop (RCL; red) that can be cleaved by C1 papain-like Cys protease RD21 (family C1A) or metcaspase-9 (MC9, family C14). Cleavage causes a conformational change that inserts the loop (blue) as a new anti-parallel β -sheet between the two β -sheets (blue arrows). Bottom panel, schematic illustration of the mouse trap mechanism. Cleavage of the RCL results in a conformational change that inserts the blue strand of the loop into the serpin structure, irreversibly deforming the serpin and the covalently trapped protease. (b) Upper panel, crystal structure of human α 2-macroglobulin (family I39; PDB ID 4acq). Only two of the four subunits are shown, in grey and blue, respectively. The bait peptide structure is unresolved and indicated with a red dashed line. The Cys-Glu thioester (green) traps the protease covalently. Lower panel, simplified illustration of the inhibitory mechanism of α 2-macroglobulin. Any small-size protease that cleaves the promiscuous bait peptide triggers a conformational change that traps the protease in a cage. The thioester (green loop) reacts with any Lys residue on the surface of the protease, thereby covalently immobilizing the protease in the cage.

3.1 Serpins (I4): springing the mouse trap on Ser and Cys proteases

Ser and Cys protease inhibitors of the MEROPS family I4, called serpins, occur in the grains of cereals of the Triticeae tribe, including barley, rye, and wheat, and in oat (Poeae tribe), (Roberts *et al.*, 2003). As highly abundant grain proteins, serpins are found even in fully processed beer (Hejgaard & Kaersgaard, 1983; Hejgaard *et al.*, 1985). The biological role(s) of serpins remain somewhat unclear, although there are hints in two directions. First, artificial diets containing serpins impair growth and fertility of insect pests, indicating a potential role in defence and highlighting the potential serpins hold for crop improvement (Thomas *et al.*, 1994, 1995; Yoo *et al.*, 2000; Alvarez-Alfageme *et al.*, 2011). Second, two putative plant target proteases of the *Arabidopsis thaliana* serpin 1 (AtSerp1) are involved in programmed cell death (PCD), suggesting that AtSerp1 could have a pro-survival function. AtSerp1 inhibits metacaspase AtMC9 (family C14) *in vitro* and co-localizes with AtMC9 *in vivo* in the extracellular space (Vercammen *et al.*, 2006). Electron microscopic images of *atmc9* mutant *A. thaliana* leaf cells suggest a role for AtMC9 in clearance of the cell contents after tonoplast rupture (Bollhöner *et al.*, 2013). The second putative AtSerp1 target protease is the C1 protease RD21 (Responsive to Desiccation 21), which was identified as an *in vivo* interaction partner during pull-down of AtSerp1 from plant extracts (Lampl *et al.*, 2010). The RD21 precursor protein accumulates in ER-derived protease storage bodies, which fuse with each other and the vacuole under stress conditions. Fusion is believed to lead to activation of the proteases which assist in recycling of cellular contents during stress-induced PCD (Hayashi *et al.*,

2001). In line with this role of RD21, PCD during plant infection with the necrotrophic fungi *Botrytis cinerea* and *Sclerotinia sclerotiorum* is accelerated in the *atserpin1* knock-out, but prevented in plants lacking RD21. Interestingly, Lampl *et al.* detected AtSerp1-GFP fusions expressed from the endogenous promoter mainly in the cytoplasm (Lampl *et al.*, 2013), in contrast to the previously reported extracellular localization (Vercammen *et al.*, 2006). In addition, increased resistance of *rd21* mutant leaves to infection by *Botrytis cinerea* is in contrast to earlier findings that *rd21* mutant plants are more susceptible than the wild type to *Botrytis cinerea* (Shindo *et al.*, 2012). Some serpins act as traps with a versatile bait, like the barley serpin BSZx. Utilizing overlapping sites within its reactive center loop, BSZx can inhibit trypsin, chymotrypsin and, to some extent, cathepsin G (all family S1) *in vitro* (Dahl *et al.*, 1996). Evolutionarily, inhibitory serpins are conserved throughout the kingdoms of life, although they vary widely between uni- and multicellular organisms (Roberts *et al.*, 2004). *Chlamydomonas* serpins, for instance have a distinct intron-exon structure from higher plant serpins. The reactive centre loops differ remarkably between monocots and dicots, suggesting that serpins might have diverged with regard to target proteases and biological functions (Roberts & Hejgaard, 2007).

With regard to structure and mechanism, serpins are unique among plant protease inhibitors. The structure of the ca 43 kDa inhibitor AtSerp1 has been resolved by crystallography and consists of three conserved β -sheets and nine conserved α -helices, as common among animal serpins (Fig. 4a) (Lampl *et al.*, 2010). Serpins inhibit their target proteases in a unique, suicidal manner. Ser and Cys proteases cleave the serpin reactive loop, forming an acyl-enzyme intermediate with the serpin. Cleavage triggers a profound conformational change in the metastable fold of the serpin, which deforms the active site of the protease and irreversibly binds it to the inhibitor (Fig. 4b). This remarkable trapping mechanism (Huntington *et al.*, 2000) allows for a structural separation of inhibitory activity from protease specificity. Thus, members of the serpin family share the same fold, but target a range of Ser and Cys proteases (Fluhr *et al.*, 2012; Rawlings *et al.*, 2014).

3.2 α -macroglobulin (I39): the gilded cage

Functional macroglobulin genes are annotated in only a few plant species, according to the MEROPS database (Rawlings *et al.*, 2014) and a recent comparative genomics study (Santamaría *et al.*, 2014). This includes cucumber (*Cucumis sativus*), alpine strawberry (*Fragaria vesca*), the alga *Micromonas sp. RCC299* and black cottonwood (*Populus trichocarpa*). However, the structure, specificity and mechanism of action of this family are intriguing. Macroglobulins are large (ca 200 kDa) glycoproteins and their structure resembles a round cage (Fig. 4c) (Sottrup-Jensen, 1989). α -macroglobulins possess an exposed bait region

with recognition sites for various types of endopeptidases, cleavage of which triggers a conformational change. Thus, the protease gets trapped inside the large macroglobulin protein, much like in a cage (Fig. 4d) (Feldman *et al.*, 1985). The caged peptidase cannot bind large targets or inhibitors any more, but remains accessible for small molecules (Sottrup-Jensen, 1989). Variation of the bait region does not affect the inhibitory mechanism, allowing for multifunctionality of the macroglobulin fold (Sottrup-Jensen, 1989). The physiological role of macroglobulins in plants remains obscure. The versatile functions of this protease inhibitor family in animals and bacteria are reviewed elsewhere (Budd *et al.*, 2004; Rehman *et al.*, 2013).

Significance of multifunctional protease inhibitors in the plant research arena

Multifunctional protease inhibitors represent hubs that regulate distinct branches of the plant physiological network, for instance defence and tuber sprouting in the case of the potato multicystatin (family I25). For one inhibitor to do several biological jobs, it must often target multiple proteases of different families or even different catalytic classes. This can be achieved through several inhibitory interfaces on the same protein, as seen with the Janus-type and the multidomain I20 inhibitors, or through one promiscuous interface, as seen in the mouse trap type inhibitors. A different way to assign multiple biological functions to one inhibitor is to produce it on different occasions in space and time. For instance, multicystatins (family I25) accumulate in tomato and potato leaves in response to wounding as well as in potato tubers when building up storage protein reserves. Through whichever route inhibitors acquire multifunctionality, the result is a protein that provides a link between the biological processes it regulates. Increased knowledge on multifunctional protease inhibitors will therefore promote a network-level understanding of plant physiology. The evolutionary history of inhibitors in their role as regulatory hubs could then reveal how the network was restructured over time.

Multifunctional protease inhibitors share stabilizing structural features, most notably a compact fold linked covalently by disulphide bridges. This structural similarity may resemble a common evolutionary road to multifunctionality. It appears that small, stable proteins, such as protease inhibitors, are well suited to acquire a (second) inhibitory function, since they are already fit to persist in harsh environments with high proteolytic activity. Additional protease inhibitory activities may await discovery in many small, stable proteins, including the known multifunctional inhibitors. A second, even more obvious road towards multifunctionality is duplication of inhibitory domains, which can further lead to

neofunctionalization of the duplicate, rearrangements in the overall protein structure or emergence of new regulatory mechanisms (sections 1.1 and 2).

On the applied side, recombinant expression of multifunctional inhibitors from the families I12, I13, I20 and I25 has been successfully used to generate pest-resistant crop plants (Orr *et al.*, 1994; Xu *et al.*, 1996; Siqueira-Júnior *et al.*, 2002; Dunse *et al.*, 2010). Furthermore, I3 and I20 inhibitors were utilized to limit proteolysis of recombinant human proteins produced in plants, tackling a major issue in molecular farming (Kim *et al.*, 2008; Goulet *et al.*, 2012). Increasing our understanding how multifunctional inhibitors link physiological networks could facilitate new applications. For instance, recombinant protein degradation could be prevented very effectively by controlling regulators of proteolytic cascades in the plant. Inhibitors that interfere with endogenous as well as exogenous proteases could protect crops from pests and premature senescence. In medicine, multifunctional plant protease inhibitors of the family I12 have been used to limit undesired protease activity as potential anticancer drugs (Souza *et al.*, 2014).

In one case, the concept of multidomain inhibitors has been taken further using artificial multidomain inhibitors consisting of five naturally occurring domains (families I25 and I31) to create a more stable and more potent inhibitor than its natural, single-domain counterparts. Overexpression of this custom-made multidomain inhibitor in potato increases resistance to the insect pest *Frankliniella occidentalis* (Outchkourov *et al.*, 2004). Some impressive examples of custom-made inhibitors based on a known multifunctional fold are found among the cyclotides (section 2.2). However, the relatively low affinity of the engineered cyclotides for their targets highlights that a deeper understanding is needed to develop synthetic approaches. Knowing all essential features of an inhibitor structure, one could design completely novel inhibitors, customized for desired applications.

From the examples discussed throughout this review, it is clear that several roads to multifunctionality exist, starting from adaptation of target recognition sites in stable proteins or gene duplication, with the latter branching out in multiple directions. Fine-mapping these roads will facilitate the construction of custom multifunctional inhibitors while, at the same time, enhancing our knowledge about regulatory hubs in the plant physiological network.

Box 1: The MEROPS database of proteases and their inhibitors (<http://merops.sanger.ac.uk>, (Rawlings *et al.*, 2014))

MEROPS groups proteins into families based on sequence homology and families into clans based on structural homology. Proteases or inhibitors in a clan are assumed to have evolved from a common ancestor, but are less closely related than the proteins within a family. Inhibitor families are named I1 to

193. Protease families are named with a letter indicating the catalytic type (i.e. A for Aspartic, S for Serine and C for Cysteine proteases), followed by a consecutive number. We refer to MEROPS release 9.12 throughout this article.

Box 2: The Laskowski Mechanism of Protease Inhibition

The Laskowski mechanism of inhibition is probably the most common scenario of protease inhibition by proteinaceous inhibitors (Rawlings *et al.*, 2014). Michael Laskowski described the “standard mechanism” of protease inhibition, where the inhibitor acts as a “limited proteolysis substrate” (reviewed in Laskowski and Kato, 1980). A reactive peptide bond on this limited substrate is bound by the target protease and an acyl intermediate is formed with a high association constant. However, the rate of completion of proteolytic cleavage and dissociation is very low, resulting in an apparent equilibrium between the free enzyme and inhibitor on the one hand and the complex on the other. Both the intact and the cleaved inhibitor can bind and inhibit the protease, and cleavage is reversible.

References

- 531 **Abdul-Hussain S, Paulsen GM. 1989.** Role of proteinaceous α -amylase enzyme inhibitors in preharvest
532 sprouting of wheat grain. *Journal of Agricultural and Food Chemistry* **37**: 295–299.
- 533 **Akers CP, Hoff JE. 1980.** Simultaneous formation of chymopapain inhibitor activity and cubical crystals in
534 tomato leaves. *Canadian Journal of Botany* **58**: 1000–1003.
- 535 **Alvarez-Alfageme F, Maharramov J, Carrillo L, Vandenabeele S, Vercammen D, Van Breusegem F,**
536 **Smagghe G. 2011.** Potential use of a serpin from *Arabidopsis* for pest control. *PLoS ONE* **6**: e20278.
- 537 **Alvarez-Fernandez M, Barrett AJ, Gerhartz B, Dando PM, Ni J, Abrahamson M. 1999.** Inhibition of
538 mammalian legumain by some cystatins is due to a novel second reactive site. *Journal of Biological*
539 *Chemistry* **274**: 19195–19203.
- 540 **Atkinson AH, Heath RL, Simpson RJ, Clarke AE, Anderson MA. 1993.** Proteinase inhibitors in *Nicotiana*
541 *glauca* stigmas are derived from a precursor protein which is processed into five homologous inhibitors.
542 *The Plant Cell* **5**: 203–213.
- 543 **Azarkan M, Martinez-Rodriguez S, Buts L, Baeyens-Volant D, Garcia-Pino A. 2011.** The plasticity of the
544 β -trefoil fold constitutes an evolutionary platform for protease inhibition. *Journal of Biological Chemistry*
545 **286**: 43726–43734.
- 546 **Barbeta BL, Marshall AT, Gillon AD, Craik DJ, Anderson MA. 2008.** Plant cyclotides disrupt epithelial
547 cells in the midgut of lepidopteran larvae. *Proceedings of the National Academy of Sciences of the United*
548 *States of America* **105**: 1221–1225.
- 549 **Barrette-Ng IH, Ng KK-S, Cherney MM, Pearce G, Ryan CA, James MNG. 2003.** Structural basis of
550 inhibition revealed by a 1:2 complex of the two-headed tomato inhibitor-II and subtilisin Carlsberg.
551 *Journal of Biological Chemistry* **278**: 24062–24071.
- 552 **Beadle GW, Tatum EL. 1941.** Genetic control of biochemical reactions in *Neurospora*. *Proceedings of the*
553 *National Academy of Sciences of the United States of America* **27**: 499–506.
- 554 **Bellincampi D, Camardella L, Delcour JA, Desseaux V, D'Ovidio R, Durand A, Elliot G, Gebruers K,**
555 **Giovane A, Juge N, et al. 2004.** Potential physiological role of plant glycosidase inhibitors. *Biochimica et*
556 *Biophysica Acta (BBA) - Proteins and Proteomics* **1696**: 265–274.
- 557 **Benchabane M, Schlüter U, Vorster J, Goulet M-C, Michaud D. 2010.** Plant cystatins. *Biochimie* **92**:
558 1657–1666.
- 559 **Birk Y. 1961.** Purification and some properties of a highly active inhibitor of trypsin and α -chymotrypsin
560 from soybeans. *Biochimica et Biophysica Acta* **54**: 378–381.
- 561 **Bollhöner B, Zhang B, Stael S, Denancé N, Overmyer K, Goffner D, Van Breusegem F, Tuominen H.**
562 **2013.** Post mortem function of AtMC9 in xylem vessel elements. *New Phytologist* **200**: 498–510.
- 563 **Bowman DE. 1946.** Differentiation of soybean antitryptic factors. *Proceedings of the Society for*
564 *Experimental Biology and Medicine* **63**: 547–550.

- 565 **Budd A, Blandin S, Levashina EA, Gibson TJ. 2004.** Bacterial α 2-macroglobulins: colonization factors
566 acquired by horizontal gene transfer from the metazoan genome? *Genome biology* **5**: R38.
- 567 **Burman R, Gunasekera S, Strömstedt AA, Göransson U. 2014.** Chemistry and biology of cyclotides:
568 circular plant peptides outside the box. *Journal of Natural Products* **77**: 724–736.
- 569 **Chen P, Rose J, Love R, Wei CH, Wang B-C. 1992.** Reactive sites of an anticarcinogenic Bowman-Birk
570 proteinase inhibitor are similar to other trypsin inhibitors. *Journal of Biological Chemistry* **267**: 1990–
571 1994.
- 572 **Cohn, F. 1859.** Über Proteinkristalle in der Kartoffel. *Jahres-Bericht der Schlesischen Gesellschaft für*
573 *Vaterländische Cultur* **37**: 72–82.
- 574 **Craik DJ, Daly NL, Bond T, Waite C. 1999.** Plant cyclotides: A unique family of cyclic and knotted
575 proteins that defines the cyclic cystine knot structural motif. *Journal of Molecular Biology* **294**: 1327–
576 1336.
- 577 **Dahl SW, Rasmussen SK, Hejgaard J. 1996.** Heterologous Expression of Three Plant Serpins with Distinct
578 Inhibitory Specificities. *Journal of Biological Chemistry* **271**: 25083–25088.
- 579 **Da Silva DS, de Oliveira CFR, Parra JRP, Marangoni S, Macedo MLR. 2014.** Short and long-term
580 antinutritional effect of the trypsin inhibitor ApTI for biological control of sugarcane borer. *Journal of*
581 *Insect Physiology* **61**: 1–7.
- 582 **Dattagupta JK, Podder A, Chakrabarti C, Sen U, Mukhopadhyay D, Dutta SK, Singh M. 1999.** Refined
583 crystal structure (2.3 Å) of a double-headed winged bean α -chymotrypsin inhibitor and location of its
584 second reactive site. *Proteins* **35**: 321–331.
- 585 **Dunse KM, Stevens JA, Lay FT, Gaspar YM, Heath RL, Anderson MA. 2010.** Coexpression of potato type
586 I and II proteinase inhibitors gives cotton plants protection against insect damage in the field.
587 *Proceedings of the National Academy of Sciences of the United States of America* **107**: 15011–15015.
- 588 **Feldman SR, Gonias SL, Pizzo SV. 1985.** Model of α 2-macroglobulin structure and function. *Proceedings*
589 *of the National Academy of Sciences of the United States of America* **82**: 5700–5704.
- 590 **Fincher GB. 1989.** Molecular and cellular biology associated with endosperm mobilization in germinating
591 cereal grains. *Annual Review of Plant Physiology and Plant Molecular Biology* **40**: 305–346.
- 592 **Fluhr R, Lampl N, Roberts TH. 2012.** Serpin protease inhibitors in plant biology. *Physiologia Plantarum*
593 **145**: 95–102.
- 594 **Franco OL, Grossi de Sá MF, Sales MP, Mello LV, Oliveira AS, Rigden DJ. 2002.** Overlapping binding sites
595 for trypsin and papain on a Kunitz-type proteinase inhibitor from *Prosopis juliflora*. *Proteins: Structure,*
596 *Function, and Bioinformatics* **49**: 335–341.
- 597 **Gillon AD, Saska I, Jennings CV, Guarino RF, Craik DJ, Anderson MA. 2008.** Biosynthesis of circular
598 proteins in plants. *The Plant Journal* **53**: 505–515.

- 599 **Girard C, Rivard D, Kiggundu A, Kunert K, Gleddie SC, Cloutier C, Michaud D. 2007a.** A multicomponent,
600 elicitor-inducible cystatin complex in tomato, *Solanum lycopersicum*. *The New phytologist* **173**: 841–851.
- 601 **Girard C, Rivard D, Kiggundu A, Kunert K, Gleddie SC, Cloutier C, Michaud D. 2007b.** A multicomponent,
602 elicitor-inducible cystatin complex in tomato, *Solanum lycopersicum*. *New Phytologist* **173**: 841–851.
- 603 **Goulet C, Khalf M, Sainsbury F, D'Aoust M-A, Michaud D. 2012.** A protease activity-depleted
604 environment for heterologous proteins migrating towards the leaf cell apoplast. *Plant biotechnology*
605 *journal* **10**: 83–94.
- 606 **Graham JS, Pearce G, Merryweather J, Titani K, Ericsson LH, Ryan CA. 1985.** Wound-induced proteinase
607 inhibitors from tomato leaves. II. The cDNA-deduced primary structure of pre-inhibitor II. *Journal of*
608 *Biological Chemistry* **260**: 6561–6564.
- 609 **Gran L. 1973.** Oxytocic principles of *Oldenlandia affinis*. *Lloydia* **36**: 174–178.
- 610 **Greenblatt HM, Ryan CA, James MNG. 1989.** Structure of the complex of *Streptomyces griseus*
611 proteinase B and polypeptide chymotrypsin inhibitor-1 from Russet Burbank potato tubers at 2.1 Å
612 resolution. *Journal of Molecular Biology* **205**: 201–228.
- 613 **Green AR, Nissen MS, Kumar GNM, Knowles NR, Kang C. 2013.** Characterization of *Solanum tuberosum*
614 multicystatin and the significance of core domains. *The Plant Cell* **25**: 5043–5052.
- 615 **Guerra FP, Reyes L, Vergara-Jaque A, Campos-Hernández C, Gutiérrez A, Pérez-Díaz J, Pérez-Díaz R,**
616 **Blaudez D, Ruíz-Lara S. 2015.** *Populus deltoides* Kunitz trypsin inhibitor 3 confers metal tolerance and
617 binds copper, revealing a new defensive role against heavy metal stress. *Environmental and*
618 *Experimental Botany* **115**: 28–37.
- 619 **Hara-Nishimura I, Inoue K, Nishimura M. 1991.** A unique vacuolar processing enzyme responsible for
620 conversion of several proprotein precursors into the mature forms. *FEBS Letters* **294**: 89–93.
- 621 **Hayashi Y, Yamada K, Shimada T, Matsushima R, Nishizawa N, Nishimura M, Hara-Nishimura I. 2001.** A
622 proteinase-storing body that prepares for cell death or stresses in the epidermal cells of Arabidopsis.
623 *Plant and Cell Physiology* **42**: 894–899.
- 624 **Heath RL, Barton PA, Simpson RJ, Reid GE, Lim G, Anderson MA. 1995.** Characterization of the protease
625 processing sites in a multidomain proteinase inhibitor precursor from *Nicotiana glauca*. *European Journal*
626 *of Biochemistry* **230**: 250–257.
- 627 **Hedstrom L, Szilagyi L, Rutter WJ. 1992.** Converting trypsin to chymotrypsin: the role of surface loops.
628 *Science* **255**: 1249–1253.
- 629 **Hejgaard J, Kaersgaard P. 1983.** Purification and properties of the major antigenic beer protein of barley
630 origin. *Journal of the Institute of Brewing* **89**: 402–410.
- 631 **Hejgaard J, Rasmussen SK, Brandt A, Svendsen I. 1985.** Sequence homology between barley endosperm
632 protein Z and protease inhibitors of the α 1-antitrypsin family. *FEBS Letters* **180**: 89–94.

- 633 **Hernandez J-F, Gagnon J, Chiche L, Nguyen TM, Andrieu J-P, Heitz A, Trinh Hong T, Pham TTC, Le**
 634 **Nguyen D. 2000.** Squash trypsin inhibitors from *Momordica cochinchinensis* exhibit an atypical
 635 macrocyclic structure. *Biochemistry* **39**: 5722–5730.
- 636 **Hiraiwa N, Nishimura M, Hara-Nishimura I. 1999.** Vacuolar processing enzyme is self-catalytically
 637 activated by sequential removal of the C-terminal and N-terminal propeptides. *FEBS Letters* **447**: 213–
 638 216.
- 639 **Huntington JA, Read RJ, Carrell RW. 2000.** Structure of a serpin–protease complex shows inhibition by
 640 deformation. *Nature* **407**: 923–926.
- 641 **Hwang JE, Hong JK, Je JH, Lee KO, Kim DY, Lee SY, Lim CO. 2009.** Regulation of seed germination and
 642 seedling growth by an Arabidopsis phytocystatin isoform, AtCYS6. *Plant Cell Reports* **28**: 1623–1632.
- 643 **Iwasaki T, Kiyohara T, Yoshikawa M. 1976.** Amino acid sequence of an active fragment of potato
 644 proteinase inhibitor IIa. *Journal of Biochemistry* **79**: 381–391.
- 645 **Jennings C, West J, Waite C, Craik D, Anderson M. 2001.** Biosynthesis and insecticidal properties of
 646 plant cyclotides: The cyclic knotted proteins from *Oldenlandia affinis*. *Proceedings of the National*
 647 *Academy of Sciences of the United States of America* **98**: 10614–10619.
- 648 **Johnson R, Narvaez J, An G, Ryan C. 1989.** Expression of proteinase inhibitors I and II in transgenic
 649 tobacco plants: effects on natural defense against *Manduca sexta* larvae. *Proceedings of the National*
 650 *Academy of Sciences of the United States of America* **86**: 9871–9875.
- 651 **Joshi RS, Gupta VS, Giri AP. 2014.** Differential antibiosis against *Helicoverpa armigera* exerted by distinct
 652 inhibitory repeat domains of *Capsicum annuum* proteinase inhibitors. *Phytochemistry* **101**: 16–22.
- 653 **Kim T-G, Lee H-J, Jang Y-S, Shin Y-J, Kwon T-H, Yang M-S. 2008.** Co-expression of proteinase inhibitor
 654 enhances recombinant human granulocyte–macrophage colony stimulating factor production in
 655 transgenic rice cell suspension culture. *Protein Expression and Purification* **61**: 117–121.
- 656 **Kong L, Ranganathan S. 2008.** Tandem duplication, circular permutation, molecular adaptation: how
 657 *Solanaceae* resist pests via inhibitors. *BMC Bioinformatics* **9**: S22.
- 658 **Kumar GNM, Houtz RL, Knowles NR. 1999.** Age-induced protein modifications and increased proteolysis
 659 in potato seed-tubers. *Plant Physiology* **119**: 89–100.
- 660 **Kunitz M. 1945.** Crystallization of a trypsin inhibitor from soybean. *Science* **101**: 668–669.
- 661 **Lampl N, Alkan N, Davydov O, Fluhr R. 2013.** Set-point control of RD21 protease activity by AtSerp1
 662 controls cell death in Arabidopsis. *The Plant Journal* **74**: 498–510.
- 663 **Lampl N, Budai-Hadrian O, Davydov O, Joss TV, Harrop SJ, Curmi PMG, Roberts TH, Fluhr R. 2010.**
 664 Arabidopsis AtSerp1, crystal structure and in vivo interaction with its target protease RESPONSIVE TO
 665 DESICCATION-21 (RD21). *Journal of Biological Chemistry* **285**: 13550–13560.
- 666 **Laskowski M, Kato I. 1980.** Protein inhibitors of proteinases. *Annual Review of Biochemistry* **49**: 593–
 667 626.

- 668 **Leah R, Mundy J. 1989.** The bifunctional α -amylase/subtilisin inhibitor of barley: nucleotide sequence
669 and patterns of seed-specific expression. *Plant Molecular Biology* **12**: 673–682.
- 670 **Lee MCS, Scanlon MJ, Craik DJ, Anderson MA. 1999.** A novel two-chain proteinase inhibitor generated
671 by circularization of a multidomain precursor protein. *Nature Structural & Molecular Biology* **6**: 526–530.
- 672 **Madureira HC, Da Cunha M, Jacinto T. 2006.** Immunolocalization of a defense-related 87 kDa cystatin in
673 leaf blade of tomato plants. *Environmental and Experimental Botany* **55**: 201–208.
- 674 **Mahatmanto T. 2015.** Seed biopharmaceutical cyclic peptides: From discovery to applications. *Peptide*
675 *Science (in press)*.
- 676 **Mahatmanto T, Mylne JS, Poth AG, Swedberg JE, Kaas Q, Schaefer H, Craik DJ. 2015.** The evolution of
677 Momordica cyclic peptides. *Molecular Biology and Evolution* **32**: 392–405.
- 678 **Major IT, Constabel CP. 2008.** Functional analysis of the Kunitz Trypsin Inhibitor family in poplar reveals
679 biochemical diversity and multiplicity in defence against herbivores. *Plant Physiology* **146**: 888–903.
- 680 **Margis R, Reis EM, Villeret V. 1998.** Structural and phylogenetic relationships among plant and animal
681 cystatins. *Archives of Biochemistry and Biophysics* **359**: 24–30.
- 682 **Martinez M, Diaz I. 2008.** The origin and evolution of plant cystatins and their target cysteine
683 proteinases indicate a complex functional relationship. *BMC Evolutionary Biology* **8**: 198.
- 684 **Martinez M, Diaz-Mendoza M, Carrillo L, Diaz I. 2007.** Carboxy terminal extended phytocystatins are
685 bifunctional inhibitors of papain and legumain cysteine proteinases. *FEBS Letters* **581**: 2914–2918.
- 686 **Maskos K, Huber-Wunderlich M, Glockshuber R. 1996.** RBI, a one-domain α -amylase/trypsin inhibitor
687 with completely independent binding sites. *FEBS Letters* **397**: 11–16.
- 688 **Meulenbroek EM, Thomassen EAJ, Pouvreau L, Abrahams JP, Gruppen H, Pannu NS. 2012.** Structure of
689 a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor
690 from potato. *Acta Crystallographica Section D* **68**: 794–799.
- 691 **Micheelsen PO, Vévodová J, De Maria L, Østergaard PR, Friis EP, Wilson K, Skjøl M. 2008.** Structural
692 and mutational analyses of the interaction between the barley α -amylase/subtilisin inhibitor and the
693 subtilisin savinase reveal a novel mode of inhibition. *Journal of Molecular Biology* **380**: 681–690.
- 694 **Migliolo L, de Oliveira AS, Santos EA, Franco OL, de Sales MP. 2010.** Structural and mechanistic insights
695 into a novel non-competitive Kunitz trypsin inhibitor from *Adenanthera pavonina* L. seeds with double
696 activity toward serine- and cysteine-proteinases. *Journal of Molecular Graphics and Modelling* **29**: 148–
697 156.
- 698 **Mignery GA, Pikaard CS, Park WD. 1988.** Molecular characterization of the patatin multigene family of
699 potato. *Gene* **62**: 27–44.
- 700 **Mundy J, Svendsen IB, Hejgaard J. 1983.** Barley α -amylase/subtilisin inhibitor. I. Isolation and
701 characterization. *Carlsberg Research Communications* **48**: 81–90.

- 702 **Mylne JS, Chan LY, Chanson AH, Daly NL, Schaefer H, Bailey TL, Nguyencong P, Cascales L, Craik DJ.**
 703 **2012.** Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated
 704 biosynthesis. *The Plant Cell* **24**: 2765–2778.
- 705 **Mylne JS, Colgrave ML, Daly NL, Chanson AH, Elliott AG, McCallum EJ, Jones A, Craik DJ. 2011.**
 706 Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nature Chemical*
 707 *Biology* **7**: 257–259.
- 708 **Nielsen PK, Bønsager BC, Berland CR, Sigurskjold BW, Svensson B. 2003.** Kinetics and energetics of the
 709 binding between barley α -amylase/subtilisin inhibitor and barley α -amylase 2 analyzed by surface
 710 plasmon resonance and isothermal titration calorimetry. *Biochemistry* **42**: 1478–1487.
- 711 **Nielsen KJ, Heath RL, Anderson MA, Craik DJ. 1995.** Structures of a series of 6-kDa trypsin inhibitors
 712 isolated from the stigma of *Nicotiana glauca*. *Biochemistry* **34**: 14304–14311.
- 713 **Nissen MS, Kumar GNM, Youn B, Knowles DB, Lam KS, Ballinger WJ, Knowles NR, Kang C. 2009.**
 714 Characterization of *Solanum tuberosum* multicystatin and its structural comparison with other cystatins.
 715 *The Plant Cell Online* **21**: 861–875.
- 716 **Odani S, Ikenaka T. 1973.** Scission of soybean Bowman-Birk proteinase inhibitor into two small
 717 fragments having either trypsin or chymotrypsin inhibitory activity. *Journal of Biochemistry* **74**: 857–860.
- 718 **Orr GL, Strickland JA, Walsh TA. 1994.** Inhibition of *Diabrotica* larval growth by a multicystatin from
 719 potato tubers. *Journal of Insect Physiology* **40**: 893–900.
- 720 **Otlewski J, Zbyryt T. 1994.** Single peptide bond hydrolysis/resynthesis in squash inhibitors of serine
 721 proteinases. 1. Kinetics and thermodynamics of the interaction between squash inhibitors and bovine
 722 beta-trypsin. *Biochemistry* **33**: 200–207.
- 723 **Outchkourov NS, De Kogel WJ, Wiegers GL, Abrahamson M, Jongsma MA. 2004.** Engineered
 724 multidomain cysteine protease inhibitors yield resistance against western flower thrips (*Frankliniella*
 725 *occidentalis*) in greenhouse trials. *Plant Biotechnology Journal* **2**: 449–458.
- 726 **Park EY, Kim J-A, Kim H-W, Kim YS, Song HK. 2004.** Crystal Structure of the Bowman–Birk inhibitor from
 727 barley seeds in ternary complex with porcine trypsin. *Journal of Molecular Biology* **343**: 173–186.
- 728 **Pekkarinen AI, Jones BL. 2003.** Purification and identification of barley proteins that inhibit the alkaline
 729 Serine proteinases of *Fusarium culmorum*. *Journal of Agricultural and Food Chemistry* **51**: 1710–1717.
- 730 **Pekkarinen AI, Longstaff C, Jones BL. 2007.** Kinetics of the inhibition of *Fusarium* Serine proteinases by
 731 barley (*Hordeum vulgare* L.) inhibitors. *Journal of Agricultural and Food Chemistry* **55**: 2736–2742.
- 732 **Pouvreau L, Gruppen H, Piersma SR, van den Broek LAM, van Koningsveld GA, Voragen AGJ. 2001.**
 733 Relative abundance and inhibitory distribution of protease inhibitors in potato juice from cv. Elkana.
 734 *Journal of Agricultural and Food Chemistry* **49**: 2864–2874.

- 735 **Qu L-J, Chen J, Liu M, Pan N, Okamoto H, Lin Z, Li C, Li D, Wang J, Zhu G, et al. 2003.** Molecular cloning
736 and functional analysis of a novel type of Bowman-Birk inhibitor gene family in rice. *Plant Physiology*
737 **133**: 560–570.
- 738 **Rakwal R, Kumar Agrawal G, Jwa N-S. 2001.** Characterization of a rice (*Oryza sativa* L.) Bowman–Birk
739 proteinase inhibitor: tightly light regulated induction in response to cut, jasmonic acid, ethylene and
740 protein phosphatase 2A inhibitors. *Gene* **263**: 189–198.
- 741 **Rawlings ND, Waller M, Barrett AJ, Bateman A. 2014.** MEROPS: the database of proteolytic enzymes,
742 their substrates and inhibitors. *Nucleic Acids Research* **42**: D503–D509.
- 743 **Rehman AA, Ahsan H, Khan FH. 2013.** α 2-macroglobulin: A physiological guardian. *Journal of Cellular*
744 *Physiology* **228**: 1665–1675.
- 745 **Renko M, Sabotič J, Turk D. 2012.** β -Trefoil inhibitors – from the work of Kunitz onward. *Biological*
746 *Chemistry* **393**: 1043–1054.
- 747 **Roberts TH, Hejgaard J. 2007.** Serpins in plants and green algae. *Functional & Integrative Genomics* **8**: 1–
748 27.
- 749 **Roberts TH, Hejgaard J, Saunders NFW, Cavicchioli R, Curmi PMG. 2004.** Serpins in Unicellular Eukarya,
750 Archaea, and Bacteria: Sequence Analysis and Evolution. *Journal of Molecular Evolution* **59**: 437–447.
- 751 **Roberts TH, Marttila S, Rasmussen SK, Hejgaard J. 2003.** Differential gene expression for suicide-
752 substrate serine proteinase inhibitors (serpins) in vegetative and grain tissues of barley. *Journal of*
753 *experimental botany* **54**: 2251–2263.
- 754 **Rodis P, Hoff JE. 1984.** Naturally occurring protein crystals in the potato. *Plant Physiology* **74**: 907–911.
- 755 **Saether O, Craik DJ, Campbell ID, Sletten K, Juul J, Norman DG. 1995.** Elucidation of the primary and
756 three-dimensional structure of the uterotonic polypeptide kalata B1. *Biochemistry* **34**: 4147–4158.
- 757 **Santamaría ME, Diaz-Mendoza M, Diaz I, Martinez M. 2014.** Plant protein peptidase inhibitors: an
758 evolutionary overview based on comparative genomics. *BMC Genomics* **15**: 812–826.
- 759 **Saska I, Gillon AD, Hatsugai N, Dietzgen RG, Hara-Nishimura I, Anderson MA, Craik DJ. 2007.** An
760 asparaginyl endopeptidase mediates in vivo protein backbone cyclization. *Journal of Biological Chemistry*
761 **282**: 29721–29728.
- 762 **Scanlon MJ, Lee MC, Anderson MA, Craik DJ. 1999.** Structure of a putative ancestral protein encoded by
763 a single sequence repeat from a multidomain proteinase inhibitor gene from *Nicotiana glauca*. *Structure*
764 **7**: 793–802.
- 765 **Shindo T, Misas-Villamil JC, Hörger AC, Song J, Van der Hoorn, Renier A L. 2012.** A role in immunity for
766 Arabidopsis cysteine protease RD21, the ortholog of the tomato immune protease C14. *PloS one* **7**:
767 e29317.
- 768 **Shivaraj N, Pattabiraman T. 1981.** Natural plant enzyme inhibitors. Characterization of an unusual
769 α -amylase/trypsin inhibitor from ragi (*Eleusine coracana* Gaertn.). *Biochemical Journal* **193**: 29–36

- 770 **Siqueira-Júnior CL, Fernandes KVS, Machado OLT, da Cunha M, Gomes VM, Moura D, Jacinto T. 2002.**
771 87 kDa tomato cystatin exhibits properties of a defense protein and forms protein crystals in
772 prosystemin overexpressing transgenic plants. *Plant Physiology and Biochemistry* **40**: 247–254.
- 773 **Song HK, Kim YS, Yang JK, Moon J, Lee JY, Suh SW. 1999.** Crystal structure of a 16 kDa double-headed
774 Bowman-Birk trypsin inhibitor from barley seeds at 1.9 Å resolution. *Journal of Molecular Biology* **293**:
775 1133–1144.
- 776 **Sottrup-Jensen L. 1989.** α-macroglobulins: structure, shape, and mechanism of proteinase complex
777 formation. *Journal of Biological Chemistry* **264**: 11539–11542.
- 778 **Souza L da C, Camargo R, Demasi M, Santana JM, Sá CM de, de Freitas SM. 2014.** Effects of an
779 anticarcinogenic Bowman-Birk protease inhibitor on purified 20S proteasome and MCF-7 breast cancer
780 cells. *PLoS ONE* **9**: e86600.
- 781 **Strobl S, Maskos K, Wiegand G, Huber R, Gomis-Rüth FX, Glockshuber R. 1998.** A novel strategy for
782 inhibition of α-amylases: yellow meal worm α-amylase in complex with the Ragi bifunctional inhibitor at
783 2.5 Å resolution. *Structure* **6**: 911–921.
- 784 **Strobl S, Muehlhahn P, Bernstein R, Wiltschek R, Maskos K, Wunderlich M, Huber R, Glockshuber R,**
785 **Holak TA. 1995.** Determination of the three-dimensional structure of the bifunctional α-amylase/trypsin
786 inhibitor from Ragi seeds by NMR spectroscopy. *Biochemistry* **34**: 8281–8293.
- 787 **Stubbs MT, Laber B, Bode W, Huber R, Jerala R, Lenarcic B, Turk V. 1990.** The refined 2.4 Å x-ray crystal
788 structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of
789 proteinase inhibitor interaction. *The EMBO Journal* **9**: 1939–1947.
- 790 **Sweet RM, Wright HT, Janin J, Chothia CH, Blow DM. 1974.** Crystal structure of the complex of porcine
791 trypsin with soybean trypsin inhibitor (Kunitz) at 2.6-Å resolution. *Biochemistry* **13**: 4212–4228.
- 792 **Tajima T, Yamaguchi A, Matsushima S, Satoh M, Hayasaka S, Yoshimatsu K, Shioi Y. 2011.** Biochemical
793 and molecular characterization of senescence-related cysteine protease–cystatin complex from spinach
794 leaf. *Physiologia Plantarum* **141**: 97–116.
- 795 **Tamhane VA, Chougule NP, Giri AP, Dixit AR, Sainani MN, Gupta VS. 2005.** In vivo and in vitro effect of
796 Capsicum annum proteinase inhibitors on *Helicoverpa armigera* gut proteinases. *Biochimica Et*
797 *Biophysica Acta* **1722**: 156–167.
- 798 **Tamhane VA, Giri AP, Kumar P, Gupta VS. 2009.** Spatial and temporal expression patterns of diverse
799 Pin-II proteinase inhibitor genes in *Capsicum annum* Linn. *Gene* **442**: 88–98.
- 800 **Thomas JC, Adams DG, Keppenne VD, Wasmann CC, Brown JK, Kanost MR, Bohnert HJ. 1995.** Protease
801 inhibitors of *Manduca sexta* expressed in transgenic cotton. *Plant Cell Reports* **14**: 758–762.
- 802 **Thomas JC, Wasmann CC, Echt C, Dunn RL, Bohnert HJ, McCoy TJ. 1994.** Introduction and expression of
803 an insect proteinase inhibitor in alfalfa *Medicago sativa* L. *Plant Cell Reports* **14**: 31–36.

- 804 **Thongyoo P, Bonomelli C, Leatherbarrow RJ, Tate EW. 2009.** Potent inhibitors of β -tryptase and human
805 leukocyte elastase based on the MCoTI-II scaffold. *Journal of Medicinal Chemistry* **52**: 6197–6200.
- 806 **Thongyoo P, Roqué-Rosell N, Leatherbarrow RJ, Tate EW. 2008.** Chemical and biomimetic total
807 syntheses of natural and engineered MCoTI cyclotides. *Organic & Biomolecular Chemistry* **6**: 1462–1470.
- 808 **Uppalapati SR, Ayoubi P, Weng H, Palmer DA, Mitchell RE, Jones W, Bender CL. 2005.** The phytotoxin
809 coronatine and methyl jasmonate impact multiple phytohormone pathways in tomato. *The Plant Journal*
810 **42**: 201–217.
- 811 **Vallée F, Kadziola A, Bourne Y, Juy M, Rodenburg KW, Svensson B, Haser R. 1998.** Barley α -amylase
812 bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 Å resolution.
813 *Structure* **6**: 649–659.
- 814 **Valueva TA, Revina TA, Mosolov VV, Mentele R. 2000.** Primary structure of potato kunitz-type serine
815 proteinase inhibitor. *Biological Chemistry* **381**: 1215–1221.
- 816 **Vercammen D, Belenghi B, van de Cotte B, Beunens T, Gavigan J-A, De Rycke R, Brackenier A, Inzé D,
817 Harris JL, Van Breusegem F. 2006.** Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for Metacaspase
818 9. *Journal of Molecular Biology* **364**: 625–636.
- 819 **Voss RH, Ermler U, Essen LO, Wenzl G, Kim YM, Flecker P. 1996.** Crystal structure of the bifunctional
820 soybean Bowman-Birk inhibitor at 0.28-nm resolution. Structural peculiarities in a folded protein
821 conformation. *European Journal of Biochemistry* **242**: 122–131.
- 822 **Walsh TA, Strickland JA. 1993.** Proteolysis of the 85-kilodalton crystalline cysteine proteinase inhibitor
823 from potato releases functional cystatin domains. *Plant Physiology* **103**: 1227–1234.
- 824 **Weeda SM, Mohan Kumar GN, Richard Knowles N. 2009.** Developmentally linked changes in proteases
825 and protease inhibitors suggest a role for potato multicystatin in regulating protein content of potato
826 tubers. *Planta* **230**: 73–84.
- 827 **Xu D, Xue Q, McElroy D, Mawal Y, Hilder VA, Wu R. 1996.** Constitutive expression of a cowpea trypsin
828 inhibitor gene, CpTi, in transgenic rice plants confers resistance to two major rice insect pests. *Molecular*
829 *Breeding* **2**: 167–173.
- 830 **Yamagata H, Kunitatsu K, Kamasaka H, Kuramoto T, Iwasaki T. 1998.** Rice bifunctional alpha-
831 amylase/subtilisin inhibitor: characterization, localization, and changes in developing and germinating
832 seeds. *Bioscience, Biotechnology, and Biochemistry* **62**: 978–985.
- 833 **Yoo B-C, Aoki K, Xiang Y, Campbell LR, Hull RJ, Xoconostle-Cazares B, Monzer J, Lee J-Y, Ullman DE,
834 Lucas WJ. 2000.** Characterization of *Cucurbita maxima* Phloem Serpin-1 (CmPS-1): A developmentally
835 regulated elastase inhibitor. *Journal of Biological Chemistry* **275**: 35122–35128.
- 836 **Zakharov A, Carchilan M, Stepurina T, Rotari V, Wilson K, Vaintraub I. 2004.** A comparative study of the
837 role of the major proteinases of germinated common bean (*Phaseolus vulgaris* L.) and soybean (*Glycine*
838 *max* (L.) Merrill) seeds in the degradation of their storage proteins. *Journal of Experimental Botany* **55**:
839 2241–2249.

841 Figures & Figure legends

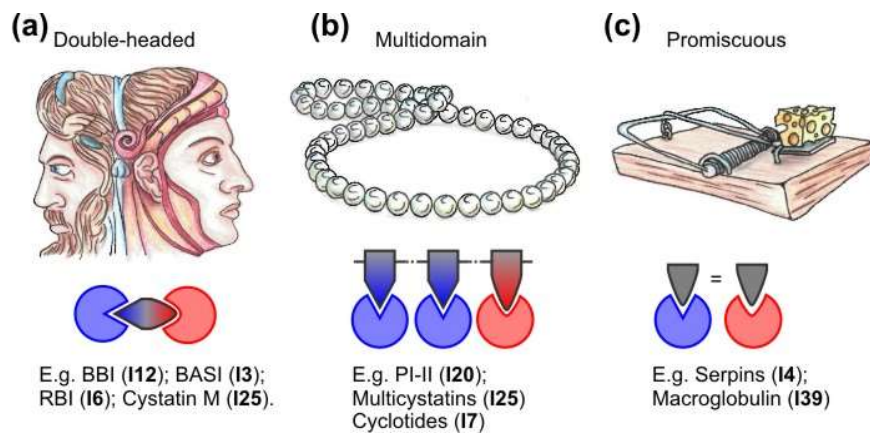


Figure 1: Three classes of multifunctional protease inhibitors.

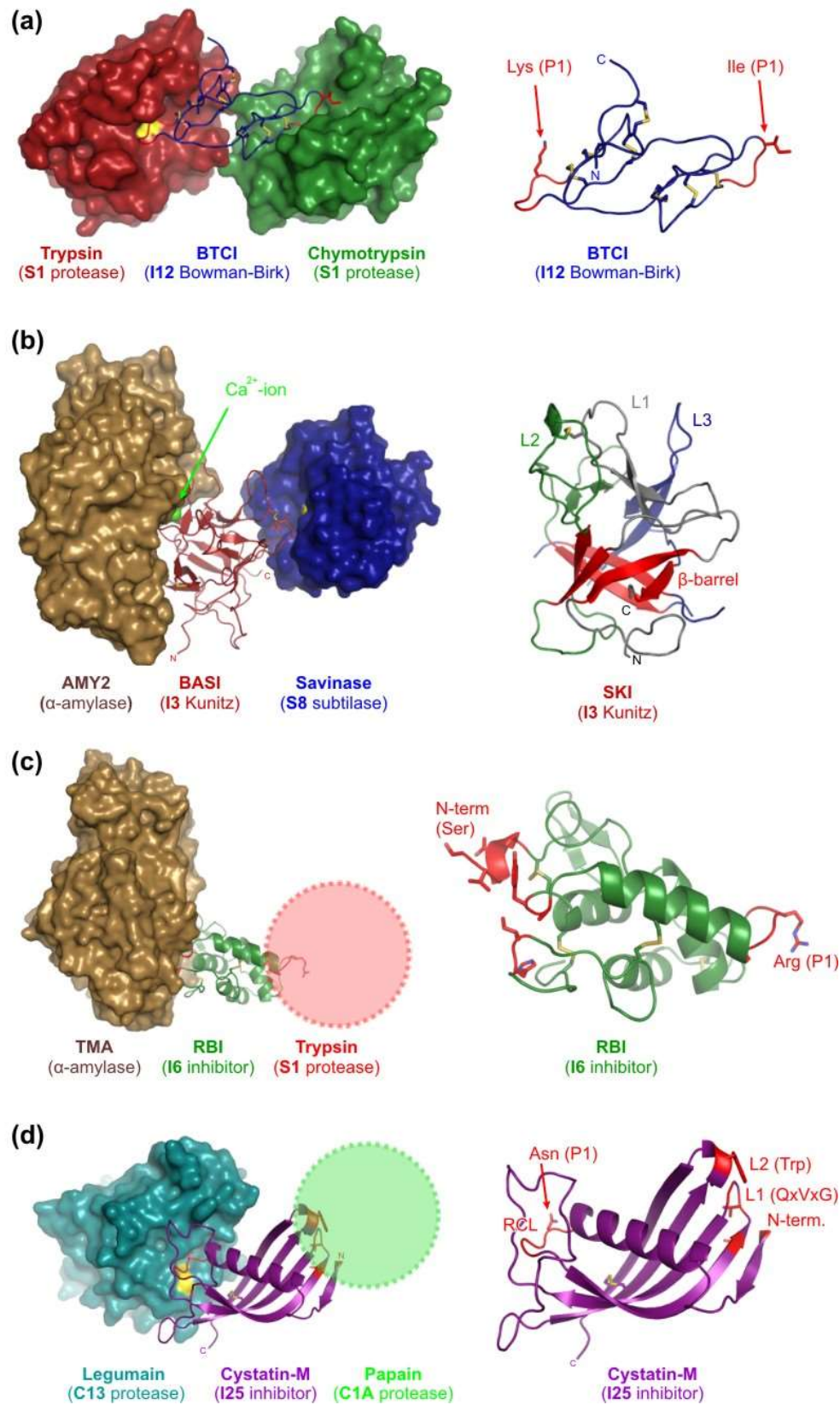


Figure 2: Four examples of double-headed inhibitors.

(a) Left: Crystal structure (PDB ID 3ru4) of *Vigna unguiculata* **Bowman-Birk** trypsin and chymotrypsin Inhibitor (purple, BTCL, **family I12**) in complex with bovine trypsin (red, family S1) and bovine chymotrypsin (green, family S1). Right: BTCL inhibits trypsin and chymotrypsin using loops (red) that contain Lys and Ile residues mimicking the P1 substrate recognition sites of the respective target enzymes.

(b) Left: Crystal structures of *Hordeum vulgare* (barley) **α -amylase/subtilisin inhibitor** (BASI, **family I3**, PDB ID 3bx1) in complex with barley α -amylase (AMY2, brown, PDB ID 1ava) and *Bacillus lentus* Savinase (blue, family S8, PDB ID 3bx1). A calcium ion (green) acts as molecular glue in the AMY2-BASI interaction. Right: The *Glycine max* (soybean) Kunitz inhibitor (SKI, family I3, PDB ID 1avu) illustrates the overall structure of Kunitz inhibitors: a β -barrel (red) with three extended loops (L1-L3), representing a tree trunk and branches, respectively.

(c) Left: Crystal structure of **Ragi Bifunctional Inhibitor** (RBI, **family I6**, PDB ID 1tmq, green) from *Eleusine coracana* (Ragi or Indian finger millet) with *Tenebrio molitor* (yellow meal worm) α -amylase (TMA, brown, PDB ID 1tmq). The location where a trypsin-like protease could bind is indicated by a red circle. Right: Inhibition is facilitated by the N-terminal Ser residue of RBI (red) that interacts with the active site of TMA. The loop that can interact with trypsin-like proteases (family S1) contains an Arg residue (red) that mimics the P1 substrate recognition motif for trypsin.

(d) Left: Crystal structure (PDB ID 4n6o) of human **Cystatin M** (purple, **family I25**) with human legumain (cyan, family C13). The location where a papain-like protease could bind is indicated by a green circle. Right: Legumains specifically cleave after Asn (P1 = Asn), therefore selectivity of the interaction is facilitated by an Asn residue in the Reactive Center Loop (RCL, red). The three regions of Cystatin M that would bind the substrate-binding groove of papain-like Cys proteases are located on the opposing side (red).

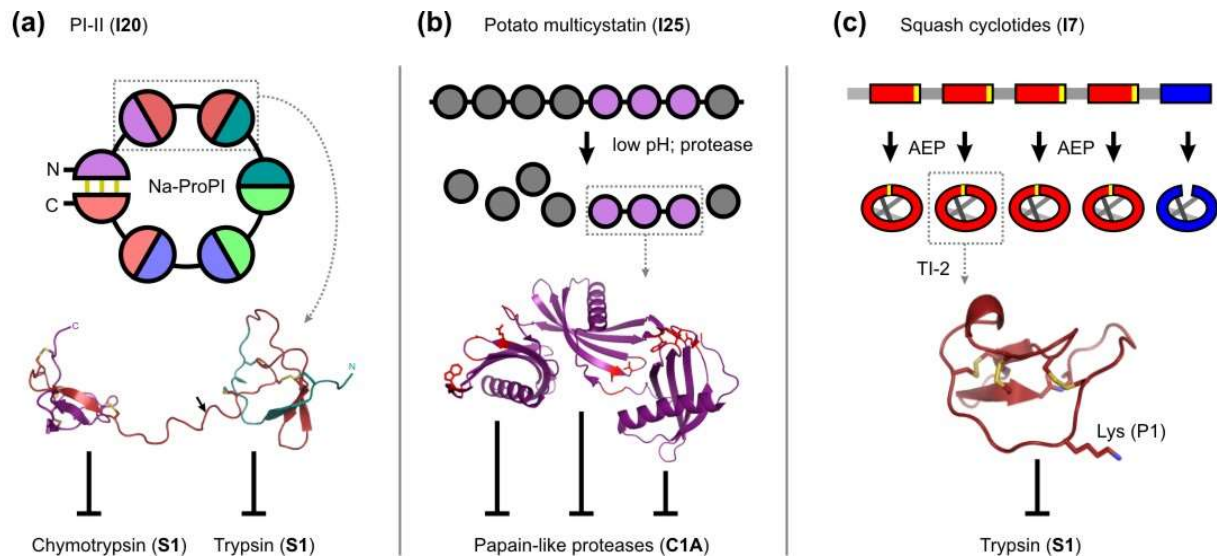


Figure 3: Folding and processing of multidomain inhibitors.

(a) Folding and processing of the **PI-II inhibitor** of *Nicotiana alata* (Na-ProPI, **family I20**). Na-ProPI contains six sequence repeats (highlighted in different colours) that fold into six hybrid inhibitory domains (depicted as pearls). One domain consists of the N- and C-termini linked by three disulphide bridges. The inhibitory domains are released through proteolytic cleavage by Cys proteases (presumably family C13). The NMR structure of the 2nd and 3rd domain with the cleavable linker in between them has been resolved (bottom, PDB ID 1fyb). The cleavage site for release of the two domains, which target chymotrypsin and trypsin (both family S1), respectively, is indicated by a black arrow.

(b) Folding and processing of potato **multicystatin** (PMC, **family I25**). The precursor consists of eight cystatin domains (depicted as pearls) that are released through proteolytic cleavage by Ser proteases at low pH. Domains 5-7 remain together and have been crystallized (PDB ID 4lzi). All released cystatin domains target papain-like Cys proteases (family C1) using a tripartite wedge inhibitory site (highlighted in red).

(c) Folding and processing of squash **cyclotides** (PMC, **family I7**). The precursor consists of five sequence repeats, each but the last ending in the AEP/VPE cleavage site Asp or Asn, indicated in yellow. Four of the released peptides (red) undergo cyclization by AEP/VPE, while the last repeat (blue) remains acyclic. All five released cyclotides share the cystine knot motif with three disulphide bridges (grey). The crystal structure of the inhibitory cyclotide MCoTI-II (PDB ID 4gux) shows the disulphide bridges linking the central loop to the periphery (yellow) and the Lys residue recognized by trypsin (family S1).

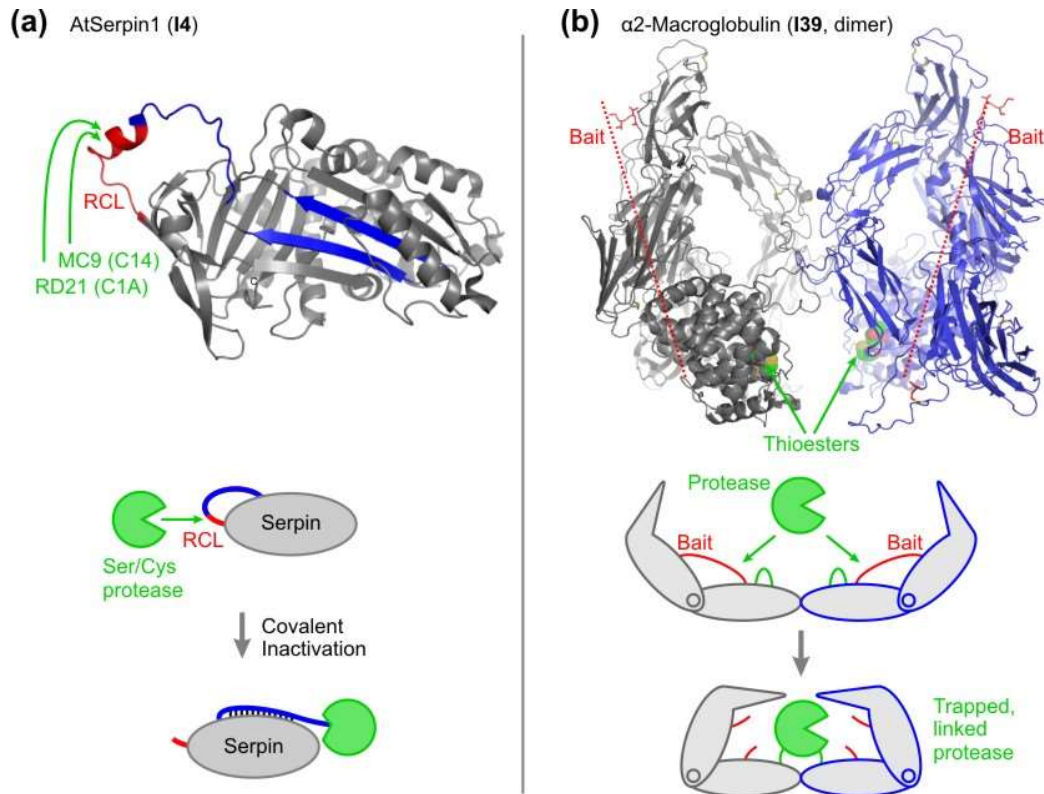


Figure 4: Serpin and α 2-Macroglobulin, the promiscuous mouse traps

(a) Top: Crystal structure of the **serpin** AtSerpin1 (**family I4**, PDB ID 3le2) showing the reactive centre loop (RCL, red) that can be cleaved by C1 papain-like Cys protease RD21 (family C1A) or Metcaspase-9 (MC9, family C14). Cleavage causes a conformational change that inserts the loop (blue) as a new anti-parallel β -sheet in between the two β -sheets (blue arrows). Bottom: Schematic illustration of the mouse trap mechanism. Cleavage of the RCL results in a conformational change that insets the blue strand of the loop into the serpin structure, irreversibly deforming the serpin and the covalently trapped protease.

(b) Top: Crystal structure of human **α 2-macroglobulin** (**family I39**, PDB ID 4acq). Only two of the four subunits are shown in grey and blue, respectively. The bait peptide structure is unresolved and indicated with a red dashed line. The Cys-Glu thioester (green) traps the protease covalently. Bottom: Simplified illustration of the inhibitory mechanism of α 2-macroglobulin. Any small size protease that cleaves the promiscuous bait peptide triggers a conformational change that traps the protease in a cage. The thioester (green loop) reacts with any Lys residue on the surface of the protease, thereby covalently immobilizing the protease in the cage.

Table 1: Plant protease inhibitor families discussed in this review

name	family	type	inhibitory mechanism	target enzymes	implicated in defence	implicated in development	examples discussed in this review
Bowman-Birk inhibitors (BBIs)	I12	DH	Laskowski	trypsin (S1), chymotrypsin (S1)	- expression induced during plant immune responses ¹ - overexpression increases resistance to a fungal pathogen ¹	ND	BBI (<i>Glycine max</i> , <i>Medicago truncatula</i> , <i>Oryza sativa</i> , <i>Zea mays</i>)
Ragi bifunctional inhibitor (RBI)	I6	DH	Laskowski	α -amylase, trypsin (S1)	- inhibits α -amylases from organisms that feed on seed, such as insects ³ and mammals ⁴	ND	RBI (<i>Eleusine coracana</i>)
Kunitz inhibitors targeting different proteases	I3	DH	Laskowski (BASI: modified Laskowski)	α -amylase, subtilisin-like proteases (S8), PLCPs (C1), trypsin (S1), chymotrypsin (S1), cathepsin D (A1)	- inhibit insect gut proteases ⁷ and α -amylases ² - reduce viability and fertility of insects ⁸	- inhibit α -amylases involved in germination ⁵	BASI (<i>Hordeum vulgare</i>) ApKTI (<i>Adenantha pavonina</i>), PSPI (<i>Solanum tuberosum</i>), KTI3 (<i>Populus trichocarpa</i> x <i>Populus deltoides</i>)
legumain-inhibiting cystatins	I25	DH	cystatin mechanism (C1), substrate mimicry (C13)	PLCPs (C1), legumains (C13)	ND	- may prevent early germination ⁶	AtCYS6 (<i>Arabidopsis thaliana</i>)
multicystatins		MD		PLCPs (C1)	- expression induced during plant immune responses ⁹ - inhibit insect gut proteases ¹⁰ - impair the growth of insects and fungal pathogens ¹¹	- regulate storage protein accumulation in tubers ¹²	PMC (<i>Solanum tuberosum</i>)
multidomain Potato peptidase inhibitor II	I20	MD	Laskowski	trypsin (S1), chymotrypsin (S1), subtilisin (S8)	- expression increases upon wounding ¹³ - inhibit insect gut proteases ¹⁴ - overexpression increases resistance to insect pests ¹⁵	ND	PI II (<i>Solanum tuberosum</i>), NaProPI (<i>Nicotiana glauca</i>)
cyclotides	I7	MD	Laskowski	trypsin (S1)	- non-inhibitory cyclotides disrupt gut cell membranes in insect predators ¹⁶	ND	MCoTI-II (<i>Momordica cochinchinensis</i>), kalata B1 (<i>Oldenlandia affinis</i> , non-inhibitory cyclotide)
serpins	I4	P	serpin mechanism (deformation)	PLCPs (C1), metacaspases (C14), chymotrypsin (S1)	- impair insect growth and fertility ¹⁷	ND	Atserpin1 (<i>Arabidopsis thaliana</i>), BSZx (<i>Hordeum vulgare</i>)
α -macroglobulins	I39	P	macroglobulin mechanism (caging)	endopeptidases	ND	ND	A2M (<i>Cucumis sativus</i> , <i>Fragaria vesca</i> , <i>Micromonas</i> sp. RCC299, <i>Populus trichocarpa</i>)

DH = double headed (Janus type); MD = multidomain (pearls on a string type); P = promiscuous inhibitors (mouse trap type); PLCPs = papain-like Cys proteases, ND = not determined.

¹(Rakwal *et al.*, 2001; Qu *et al.*, 2003), ²(Pekkarinen & Jones, 2003), ³(Strobl *et al.*, 1998), ⁴(Maskos *et al.*, 1996), ⁵(Abdul-Hussain & Paulsen, 1989; Vallée *et al.*, 1998; Nielsen *et al.*, 2003), ⁶(Hwang *et al.*, 2009), ⁷(Da Silva *et al.*, 2014), ⁹(Siqueira-Júnior *et al.*, 2002; Uppalapati *et al.*, 2005; Girard *et al.*, 2007b), ¹⁰(Orr *et al.*, 1994; Siqueira-Júnior *et al.*, 2002), ¹¹(Orr *et al.*, 1994; Siqueira-Júnior *et al.*, 2002), ¹²(Mignery *et al.*, 1988; Pouvreau *et al.*, 2001; Weeda *et al.*, 2009), ¹³(Graham *et al.*, 1985; Kong & Ranganathan, 2008), ¹⁴(Tamhane *et al.*, 2009; Joshi *et al.*, 2014), ¹⁵(Johnson *et al.*, 1989; Tamhane *et al.*, 2009; Dunse *et al.*, 2010; Joshi *et al.*, 2014), ¹⁶(Barbeta *et al.*, 2008), ¹⁷(Fluhr *et al.*, 2012)