Macrocyclization by asparaginyl endopeptidases

Amy M. James[1], Joel Haywood[1], and Joshua S. Mylne[1*]

[1] The University of Western Australia, School of Chemistry and Biochemistry & The ARC Centre of Excellence in Plant Energy Biology, 35 Stirling Highway, Crawley, Perth 6009, Australia

[*] Corresponding author. E-mail: joshua.mylne@uwa.edu.au, telephone: +61 8 6488 4415

Twitter handle: @j_mylne

Word Counts: Summary 132 words; Body of text 2,563 words

Figures: Figure 1 (colour); Figure 2 (colour)

Contents list

Summary

Plant asparaginyl endopeptidases (AEPs) are important for the post-translational processing of seed storage proteins by cleaving precursor proteins. Some AEPs also function as peptide bond-makers during the biosynthesis of several unrelated classes of cyclic peptides, namely the kalata-type cyclic peptides, PawS-Derived Peptides and cyclic knottins. These three families of gene-encoded peptides have different evolutionary origins, but all have recruited AEP for their maturation. In the last few years, the field has advanced rapidly with the biochemical characterization of three plant AEPs capable of peptide macrocyclization and insights have arisen from the first AEP crystal structures, albeit mammalian ones. Although the biochemical studies have improved our understanding of the mechanism of action, the focus now is to understand what changes in AEP sequence and structure enable some plant AEPs to perform macrocyclization reactions.

Key words

asparaginyl endopeptidase, cyclic peptides, macrocyclization , protease, transpeptidation, macrocycle

I. Introduction

Asparaginyl endopeptidases (AEPs) are cysteine endopeptidases that target Asn (N) and Asp (D) residues for cleavage. They have been identified in animals, plants and the protozoan, *Schistosoma mansoni* (Hara-Nishimura & Nishimura, 1987; Chen *et al.*, 1997; Caffrey *et al.*, 2000). AEPs were first discovered in plants, functioning in Asn-specific processing of seed storage proteins (Hara-Nishimura & Nishimura, 1987; Hara-Nishimura *et al.*, 1991; Gruis *et*

al., 2004). Since their discovery, plant AEPs have been shown to be upregulated in vegetative tissues upon wounding and in response to pathogen stress and are therefore predicted to have a function in defence (Kinoshita *et al.*, 1999; Rojo *et al.*, 2004; Kuroyanagi *et al.*, 2005).

AEPs traffic to the endoplasmic reticulum where their signal peptide is removed, forming the folded inactive precursor. Not until it reaches its final destination of the vacuole is AEP processed into its mature form via pH-dependent removal of an N-terminal and a C-terminal pro-domain (Hiraiwa *et al.*, 1999). The C-terminal pro-domain is thought to cap and restrict access to the active site based on work with mammalian AEPs (Dall & Brandstetter, 2013). By analogy to mammalian AEP, in a low pH environment, such as the plant vacuole, the electrostatic interactions between the C-terminal cap/pro-domain and the central core are potentially lost enabling AEP auto-activation through exposure of the active site.

Most plants contain multiple copies of *AEP* genes, whose encoded proteins mainly function in the cleavage of seed storage protein precursors to form mature seed storage proteins. Some *AEP* genes however encode AEPs that can perform a different enzymatic activity in the synthesis of cyclic peptides by joining residues within a linear precursor protein to make a macrocyclic product. There are four major classes of cyclic peptides in plants: the orbitides, the kalata-type cyclic peptides, cyclic knottins and the PawS-Derived Peptides. An orbitide from *Saponaria vaccaria,* segetalin A, is synthesized by cleavage of the linear precursor by an oligopeptidase followed by cyclization by a serine peptidase (Condie *et al.*, 2011; Barber *et al.*, 2013). With the exception of the orbitides, the other three plant cyclic peptide classes are processed by AEPs.

AEPs were first implicated in the processing of cyclic peptides in plants due to the highly conserved proto-C-terminal Asp/Asn within precursor sequences for kalata B1 and related peptides (Jennings *et al.*, 2001). Plants lacking cyclic peptides (*Nicotiana benthamiana* and *Arabidopsis thaliana*) could produce mass spectral evidence for the accumulation of kalata B1 when transformed with a construct for its precursor protein. The mass spectrometric signal for kalata B1 decreased relative to other masses when AEP expression was reduced transiently, providing the first direct evidence that AEP was involved in kalata B1 biosynthesis(Saska *et al.*, 2007; Gillon *et al.*, 2008). In recent years, several AEPs have been characterized and their roles in the synthesis of cyclic peptides confirmed.

II. **Plant AEPs with macrocyclizing ability**

The unusual reactions AEP could catalyse in addition to its endopeptidase activity were first shown in the processing of seed proteins in the legume species *Canavalia ensiformis*, commonly known as jack bean. An early investigation of the biosynthesis of the lectin, concanavalin A, predicted a simultaneous cleavage and ligation event occurred (Carrington *et al.*, 1985; Bowles *et al.*, 1986) and this exciting new kind of post-translational modification was for a time termed 'protein carpentry' (Hendrix, 1991). Subsequent *in vitro* studies demonstrated AEP processed concanavalin A via a transpeptidation reaction (Min & Jones, 1994; Sheldon *et al.*, 1996).

The role of AEP in the biosynthesis of cyclic peptides was demonstrated by transgene approaches with an *A. thaliana aep* knockout line and precursors for two very different cyclic trypsin inhibitor peptides (Mylne *et al.*, 2011; Mylne *et al.*, 2012). A gene from the common sunflower called *PawS1* encodes a precursor protein for Sunflower Trypsin Inhibitor 1 (SFTI-1), a small macrocyclic peptide inhibitor of trypsin (Luckett *et al.*, 1999; Mylne *et al.*, 2011). Heterologous expression of *PawS1* in the *A. thaliana aep* null showed that AEP was critical for producing SFTI-1. A later study similarly examined the precursor for cyclic knottins from the squash species, *Momordica cochinchinensis* (Mylne *et al.*, 2012). Heterologous expression of a construct encoding the precursor protein TIPTOP2 produced cyclic and acyclic knottins in a wild type background, but only the acyclic knottin in the *aep* null background (Mylne et al., 2012). Critically, these two studies and that of Saska *et al.* (2007) demonstrated by modulating AEP activity that three different cyclic peptide biosyntheses involved AEP. The repeated, independent evolutionary recruitment of AEP in macrocycle biosyntheses was termed biosynthetic parallelism and inferred AEP performed the macrocyclization reaction (Mylne *et al.*, 2012). The genetic demonstration of the role of AEP in the synthesis of different types of cyclic peptide was followed by the identification and biochemical characterization of three plant AEPs capable of cyclization that are expressed in the seeds of *Clitoria ternatea* (Nguyen *et al.*, 2014), *Canavalia ensiformis* seeds (Bernath-Levin *et al.*, 2015), and the leaves of *Oldenlandia affinis* (Harris *et al.*, 2015). These studies proved the role plant AEPs have in macrocyclization and rapidly advanced our understanding of the mechanism of macrocyclization for plant cyclic peptides.

The first isolation and characterization of a macrocyclizing AEP was from the legume *Clitoria ternatea* (Nguyen *et al.*, 2014). An AEP capable of macrocyclization was isolated

from native protein fractions from *Clitoria ternatea* pods using a substrate that assayed for cyclizing ability (Nguyen *et al.*, 2014). Named after the local name of the plant, bunga telang, Butelase 1 is an AEP purported to have fully specialized as a ligase and lost any cleavage activity. A plant extract of this *C. ternatea* AEP was capable of cyclizing several non-native substrates with the strict requirement of a C-terminal His-Val dipeptide following the Asn residue recognised for cleavage. Mutating the Asn residue to Asp or removing one or both of the C-terminal residues dramatically reduced its cyclizing ability (Nguyen *et al.*, 2014). Homology based modelling of Butelase 1 suggested that a putative extended C-terminal domain could contribute to its ligase activity by binding the cleaved substrate at the active site and protecting the acyl intermediate from hydrolysis in a similar manner to PatG, a cyanobacterial serine protease capable of cyclization (Koehnke *et al.*, 2012; Nguyen *et al.*, 2014). The *C. ternatea*-extracted AEP was highly efficient and has been subsequently used to cyclize or ligate and impressive host of non-native targets (Nguyen *et al.*, 2015; Hemu *et al.*, 2016; Nguyen *et al.*, 2016a; Nguyen *et al.*, 2016b), however attempts by Nguyen *et al.* (2014) to generate recombinant protein failed, limiting its potential for commercial use. A second AEP from *C. ternatea* was recently produced recombinantly and characterised, but could only perform cleavage reactions (Serra *et al.*, 2016).

Jack bean AEP had been shown to have transpeptidation ability in the synthesis of concanavalin A (Min & Jones, 1994; Sheldon *et al.*, 1996) and was later found to be capable of macrocyclization reactions with precursor substrates for the macrocycle from sunflower, SFTI-1 (Bernath-Levin *et al.*, 2015). SFTI-1 has an unusual biosynthetic origin as its sequence is buried in a preproprotein that codes for two unrelated proteins; SFTI-1 and a seed storage protein albumin (Mylne *et al.*, 2011). The processing of SFTI-1 from its precursor was predicted to be performed by an AEP as the sequence is flanked by Asn residues and albumins are known to be processed by AEPs. A study by Bernath-Levin *et al.* (2015) illustrated the role AEP could play in the synthesis of SFTI-1 using recombinant AEPs. The ability to produce recombinant AEP in *E. coli* allowed for the macrocyclization reaction to be reconstituted *in vitro* with pure substrate and homogenous enzyme. Although most recombinant AEPs investigated only cleaved the peptide substrate, jack bean AEP could cleave as well as perform a cleavage-coupled macrocyclization to form SFTI-1 suggesting a similar AEP from sunflower performs this reaction. The predicted mechanism of SFTI-1 biosynthesis requires three cleavages and the recognition of both Asp and Asn by AEP. The first cleavage is Asn-specific and separates an 18-residue pro-peptide from the proalbumin

precursor by cleaving at Asn residues that flank the SFTI-1 sequence N- and C- terminally. A second AEP cleavage recognizes Asp14 in a macrocyclization reaction to produce the 14-residue SFTI-1, releasing a C-terminal Gly-Leu-Asp-Asn tail (Bernath-Levin *et al.*, 2015). Any acyclic product formed via a cleavage reaction was readily degraded in the *in situ* assays (Bernath-Levin *et al.*, 2015). This highlighted the importance of macrocyclization of SFTI-1 by AEP. Furthermore, upon cyclization, the Gly_1-Asp_{14} ligation point became unrecognizable by AEP for cleavage so the macrocyclization reaction cannot be reversed by the AEP that made it (Bernath-Levin *et al.*, 2015). Thus, the cyclic structure, which results in protection against exoproteases, is further stabilized by resistance to endoproteases. This resistance to cleavage by AEP is predicted to be due to the position of the Asp residue no longer being able to adopt a beta-strand conformation, which proteases require to recognize and bind a target site (Tyndall *et al.*, 2005).

The first example of an AEP that macrocyclized its native substrate was from *O. affinis* which processed the cyclic peptide kalata B1 (Harris *et al.*, 2015). Kalata B1 is the prototypic member of a large and widespread family of plant cyclic peptides. The processing of kalata B1 and its relatives was predicted to be performed by AEPs due to the presence of a highly conserved Asn residue at the proto-C-terminus of each peptide domain of the precursor (Jennings *et al.*, 2001). Four *AEP* transcripts were identified and the protein encoded by one was made and shown to mature kalata B1 from a precursor peptide (Harris *et al.*, 2015). The sequence of this active, recombinant *O. affinis* AEP was experimentally defined and lacked the putative extended C-terminal domain of Butelase 1, suggesting extra C-terminal residues in the mature enzyme are probably not what makes it an efficient at macrocyclization. The AEP studied could only perform the C-terminal processing of kalata B1; a second enzyme is required for N-terminal processing that must occur before C-terminal processing by AEP. The authors demonstrated the macrocyclization was dependent on a C-terminal Asn-Gly-Leu AEP recognition motif and removal of the Leu prevented cyclization. This supports previous findings investigating kalata B1 cyclization showing the importance of the C-terminal tail residues for cyclization through mutagenesis (Gillon *et al.*, 2008; Conlan *et al.*, 2012).

III. Mechanism of macrocyclization by AEPs

The three studies investigating cyclizing AEPs have provided valuable insight into the mechanism of cyclic peptide biosynthesis. The proposed mechanism of macrocyclization by AEPs is an intramolecular transpeptidation reaction that energetically couples cleavage of a

C-terminal tail to the transpeptidation reaction (Fig. 1). This was demonstrated through the use of substrates lacking a C-terminal region trailing the Asn/Asp residue targeted by AEP for macrocyclization showing that cleavage of the C-terminal tail is linked to the macrocyclization reaction (Nguyen *et al.*, 2014; Bernath-Levin *et al.*, 2015; Harris *et al.*, 2015). The possible alternative mechanism of hydrolysis followed by ligation was addressed using ^{18}O-water. If hydrolysis took place, there would be an opportunity for ^{18}O to be incorporated into the cyclic product resulting in a measurable difference in mass, but no mass difference was observed for macrocyclized products (Bernath-Levin *et al.*, 2015; Harris *et al.*, 2015). In the proposed hydrolysis-independent transpeptidation reaction leading to macrocyclization, rather than using water as a nucleophile AEP uses an amino terminus. The use of capped N-terminal amine substrates illustrated how, by blocking this nucleophilic attack, cyclization could be inhibited and hydrolysis of the thioester intermediate favoured (Nguyen *et al.*, 2014; Harris *et al.*, 2015). It seems likely that AEPs may favour nucleophilic attack of a substrates N-terminus over hydrolysis through the combination of a variety of factors including C-terminus interactions and active site hydrophobicity, especially given that substrates lacking disulfide bonds have been shown to be readily macrocyclized (Nguyen *et al.*, 2014; Harris *et al.*, 2015).

A ligation capability has also been reported *in vitro* for human AEP (Dall *et al.*, 2015), but it is not known if this is physiologically relevant. The cleavage-coupled transpeptidation reaction proposed for macrocyclizing plant AEPs differs from the mechanism of ligation for human AEP in which the ligation reaction is predicted to occur independently of the cleavage reaction and the active site Cys. The ligation ability of human AEP is said to result from a pH dependent conversion of an Asp residue to succinimide, which acts as coupling reagent for peptide bond synthesis (Dall *et al.*, 2015).

Despite a greater understanding of the mechanism, it is still unclear what changes in sequence and structure help AEPs perform macrocyclization efficiently, but based on the high degree of similarity and lack of indels seen in an alignment of macrocyclizing and non-macrocyclizing AEPs, it is bound to be a subtle change (Bernath-Levin *et al.*, 2015). Currently, there are no three dimensional structures available for plant AEPs to provide structural insights into the mechanism of macrocyclization. However, the crystal structures of human and mouse AEPs in their active and inactive forms were recently published (Dall & Brandstetter, 2013; Zhao *et al.*, 2014). These structures have 53.9 and 54.3% amino acid sequence identity, respectively, to *A. thaliana* over the catalytic-domain and have been used

to construct tentative models of plant AEPs (Nguyen *et al.*, 2014; Santana *et al.*, 2016). Given the moderate sequence similarity between plant and mammalian AEPs and their analogous pH dependent activation, it is expected that plant AEPs will adopt a similar architecture with the exception being predicted to arise through the insertion of an extended poly-proline loop in the substrate specificity loop (Shafee *et al.*, 2015). Moreover, it is likely that subtle structural differences within and around the active site are responsible for the different catalytic mechanisms each AEP is capable of. A sequence comparison of plant AEPs with those residues integral to the catalysis by mouse and human AEPs (Fig. 2) illustrates that the majority of non-conserved amino acid residues between plant and mammalian AEPs are found in the previously annotated substrate specificity region. Comparison of *O. affinis* to human AEP showed different substrate requirements for cleavage further suggesting that differences in the substrate specificity loop could be the main determinant of activity (Harris *et al.*, 2015). Future structural studies are needed to identify the residues responsible for the ability some plant AEPs have to macrocyclize.

IV. Conclusions

AEPs have been recruited for processing of cyclic peptides multiple times during plant evolution. The kalata-type cyclic peptides, cyclic knottins and the PawS-type all have independent evolutionary origins. The emergence of the biosynthesis of TIPTOP-derived cyclic peptides resulted from the new recruitment of AEP for synthesis as their acyclic predecessors did not require AEP for synthesis. The SFTI-1-producing lineage appears to have co-opted the pre-existing processing machinery of seed storage albumins. In this way, a new protein evolved with a completely different function to the host protein. It is clear from these examples that the evolution of macrocyclizing AEPs was essential for the synthesis of these cyclic peptides in plants. The ability to form a cyclic structure is likely selected for as it imparts stability though minimizing degradation by exoproteases. With the recent discovery and characterization of macrocyclizing AEPs we have gained a greater understanding of the biochemical mechanism. It remains to be shown which changes in structure enable macrocyclization in addition to (or instead of) the standard endopeptidase activity and how much the substrate contributes to a cyclic peptide outcome.

Acknowledgements

A.M.J. was supported by a University International Stipend and a Scholarship for International Research Fees from The University of Western Australia. A.M.J. and J.H. were supported by Australian Research Council grants DP130101191 and DP160100107 respectively. The authors thank Charlie Bond and Jason Schmidberger for helpful comments. J.S.M. was supported by an Australian Research Council Future Fellowship (FT120100013).

References

Barber CJS, Pujara PT, Reed DW, Chiwocha S, Zhang H, Covello PS. 2013. The two-step biosynthesis of cyclic peptides from linear precursors in a member of the plant family Caryophyllaceae involves cyclization by a serine protease-like enzyme. *Journal of Biological Chemistry* **288**(18): 12500-12510.

Bernath-Levin K, Nelson C, Elliott Alysha G, Jayasena Achala S, Millar AH, Craik David J, Mylne Joshua S. 2015. Peptide macrocyclization by a bifunctional endoprotease. *Chemistry & Biology* **22**(5): 571-582.

Bowles DJ, Marcus SE, Pappin DJ, Findlay JB, Eliopoulos E, Maycox PR, Burgess J. 1986. Posttranslational processing of concanavalin A precursors in jackbean cotyledons. *Journal of Cell Biology* **102**(4): 1284-1297.

Caffrey CR, Mathieu MA, Gaffney AM, Salter JP, Sajid M, Lucas KD, Franklin C, Bogyo M, McKerrow JH. 2000. Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*. *FEBS Letters* **466**(2–3): 244-248.

Carrington DM, Auffret A, Hanke DE. 1985. Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature* **313**(5997): 64-67.

Chen JM, Dando PM, Rawlings ND, Brown MA, Young NE, Stevens RA, Hewitt E, Watts C, Barrett AJ. 1997. Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase. *Journal of Biological Chemistry* **272**(12): 8090-8098.

Condie JA, Nowak G, Reed DW, John Balsevich J, Reaney MJT, Arnison PG, Covello PS. 2011. The biosynthesis of Caryophyllaceae-like cyclic peptides in *Saponaria vaccaria* L. from DNA-encoded precursors. *The Plant Journal* **67**: 682–690.

Conlan BF, Colgrave ML, Gillon AD, Guarino R, Craik DJ, Anderson MA. 2012. Insights into processing and cyclization events associated with biosynthesis of the cyclic peptide kalata B1. *Journal of Biological Chemistry* **287**(33): 28037-28046.

Dall E, Brandstetter H. 2013. Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proceedings of the National Academy of Sciences of the United States of America* **110**(27): 10940-10945.

Dall E, Fegg JC, Briza P, Brandstetter H. 2015. Structure and mechanism of an aspartimide-dependent peptide ligase in human legumain. *Angewandte Chemie International Edition* **54**(10): 2917-2921.

Gillon AD, Saska I, Jennings CV, Guarino RF, Craik DJ, Anderson MA. 2008. Biosynthesis of circular proteins in plants. *Plant Journal* **53**(3): 505-515.

Gruis D, Schulze J, Jung R. 2004. Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases. *Plant Cell* **16**(1): 270-290.

Hara-Nishimura I, Inoue K, Nishimura M. 1991. A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. *FEBS Letters* **294**(1-2): 89-93.

Hara-Nishimura I, Nishimura M. 1987. Proglobulin processing enzyme in vacuoles isolated from developing pumpkin cotyledons. *Plant Physiology* **85**(2): 440-445.

Harris KS, Durek T, Kaas Q, Poth AG, Gilding EK, Conlan BF, Saska I, Daly NL, van der Weerden NL, Craik DJ, et al. 2015. Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat Commun* **6**: 10199.

Hemu X, Qiu Y, Nguyen GKT, Tam JP. 2016. Total Synthesis of Circular Bacteriocins by Butelase 1. *Journal of the American Chemical Society* **138**(22): 6968-6971.

Hendrix RW. 1991. Protein carpentry. *Current Biology* **1**(2): 71-73.

Hiraiwa N, Nishimura M, Hara-Nishimura I. 1999. Vacuolar processing enzyme is self-catalytically activated by sequential removal of the C-terminal and N-terminal propeptides. *FEBS Letters* **447**(2-3): 213-216.

Jennings C, West J, Waine C, Craik D, Anderson M. 2001. Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis*. *Proceedings of the National Academy of Sciences of the United States of America* **98**(19): 10614-10619.

Kinoshita T, Yamada K, Hiraiwa N, Kondo M, Nishimura M, Hara-Nishimura I. 1999. Vacuolar processing enzyme is up-regulated in the lytic vacuoles of vegetative tissues during senescence and under various stressed conditions. *Plant Journal* **19**(1): 43-53.

Koehnke J, Bent A, Houssen WE, Zollman D, Morawitz F, Shirran S, Vendome J, Nneoyiegbe AF, Trembleau L, Botting CH, et al. 2012. The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nature Structural & Molecular Biology* **19**(8): 767-772.

Kuroyanagi M, Yamada K, Hatsugai N, Kondo M, Nishimura M, Hara-Nishimura I. 2005. Vacuolar processing enzyme is essential for mycotoxin-induced cell death in *Arabidopsis thaliana*. *Journal of Biological Chemistry* **280**(38): 32914-32920.

Luckett S, Garcia RS, Barker JJ, Konarev AV, Shewry PR, Clarke AR, Brady RL. 1999. High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *Journal of Molecular Biology* **290**(2): 525-533.

Min W, Jones DH. 1994. *In vitro* splicing of concanavalin A is catalyzed by asparaginyl endopeptidase. *Nature Structural & Molecular Biology* **1**(8): 502-504.

Mylne JS, Chan LY, Chanson AH, Daly NL, Schaefer H, Bailey TL, Nguyencong P, Cascales L, Craik DJ. 2012. Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis. *Plant Cell* **24**: 2765-2778.

Mylne JS, Colgrave ML, Daly NL, Chanson AH, Elliott AG, McCallum EJ, Jones A, Craik DJ. 2011. Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nature Chemical Biology* **7**(5): 257-259.

Nguyen GKT, Hemu X, Quek J-P, Tam JP. 2016a. Butelase-Mediated Macrocyclization of d-Amino-Acid-Containing Peptides. *Angewandte Chemie International Edition* **55**(41): 12802-12806.

Nguyen GKT, Kam A, Loo S, Jansson AE, Pan LX, Tam JP. 2015. Butelase 1: A Versatile Ligase for Peptide and Protein Macrocyclization. *Journal of the American Chemical Society* **137**(49): 15398-15401.

Nguyen GKT, Qiu Y, Cao Y, Hemu X, Liu C-F, Tam JP. 2016b. Butelase-mediated cyclization and ligation of peptides and proteins. *Nat. Protocols* **11**(10): 1977-1988.

Nguyen GKT, Wang S, Qiu Y, Hemu X, Lian Y, Tam JP. 2014. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nature Chemical Biology* **10**(9): 732-738.

Rojo E, Martin R, Carter C, Zouhar J, Pan S, Plotnikova J, Jin H, Paneque M, Sanchez-Serrano JJ, Baker B. 2004. VPEγ exhibits a caspase-like activity that contributes to defense against pathogens. *Current Biology* **14**(21): 1897-1906.

Santana JO, Freire L, de Sousa AO, Fontes Soares VL, Gramacho KP, Pirovani CP. 2016. Characterization of the legumains encoded by the genome of *Theobroma cacao* L. *Plant Physiology and Biochemistry* **98**: 162-170.

Saska I, Gillon AD, Hatsugai N, Dietzgen RG, Hara-Nishimura I, Anderson MA, Craik DJ. 2007. An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclisation. *Journal of Biological Chemistry* **282**(40): 29721-29728.

Serra A, Hemu X, Nguyen GKT, Nguyen NTK, Sze SK, Tam JP. 2016. A high-throughput peptidomic strategy to decipher the molecular diversity of cyclic cysteine-rich peptides. *Scientific reports* **6**: 23005.

Shafee T, Harris K, Anderson M 2015. Biosynthesis of cyclotides. *Advances in Botanical Research*: Academic Press, 227-269.

Sheldon PS, Keen JN, Bowles DJ. 1996. Post-translational peptide bond formation during concanavalin A processing *in vitro*. *Biochemical Journal* **320**: 865-870.

Tyndall JDA, Nall T, Fairlie DP. 2005. Proteases universally recognize beta strands in their active sites. *Chemical Reviews* **105**(3): 973-1000.

Zhao L, Hua T, Crowley C, Ru H, Ni X, Shaw N, Jiao L, Ding W, Qu L, Hung L-W, et al. 2014. Structural analysis of asparaginyl endopeptidase reveals the activation mechanism and a reversible intermediate maturation stage. *Cell Research* **24**: 344-358.

Figure captions

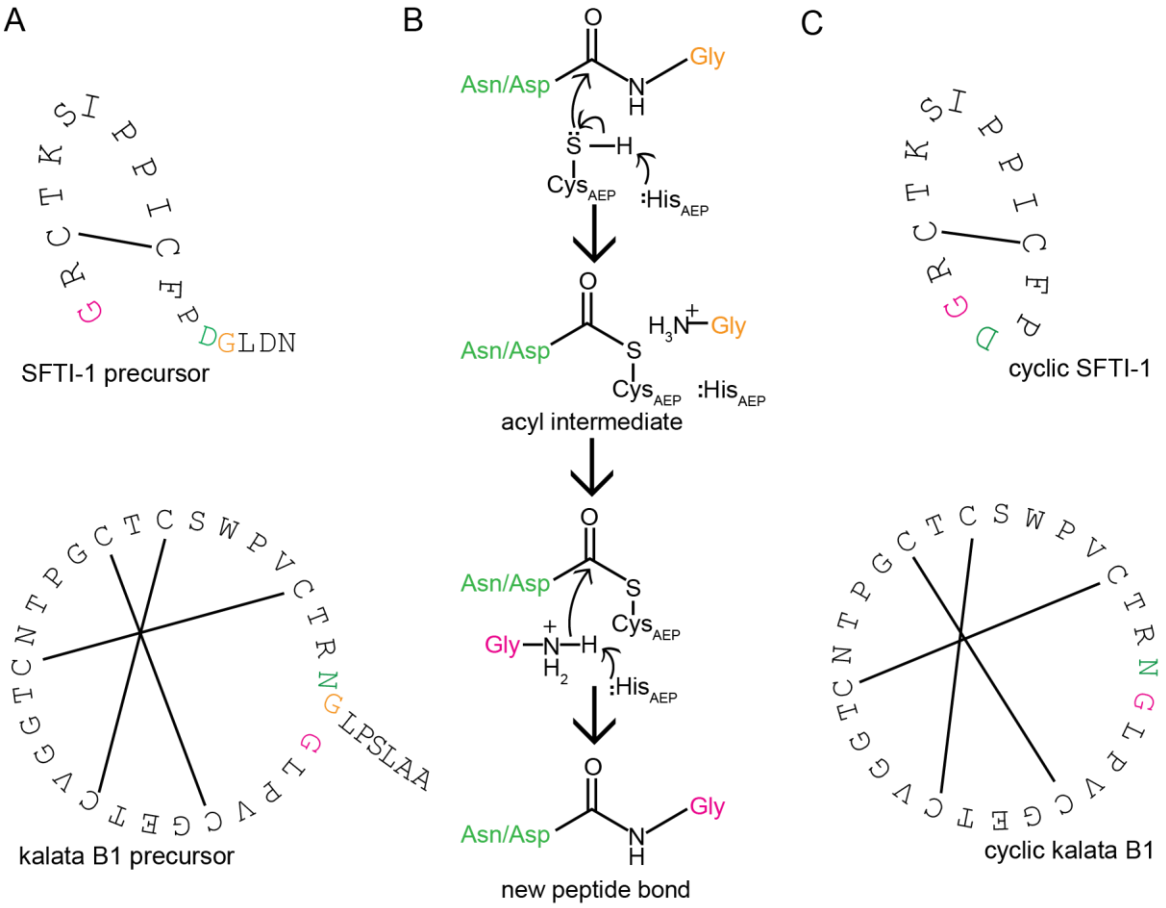

Figure 1. Predicted mechanism of the macrocyclization reaction by plant AEPs using SFTI-1 and kalata B1 as examples.

(***A***) Acyclic precursor peptides prior to cyclization are shown following processing from a preproprotein.

(***B***) The predicted mechanism of macrocyclization involves a cleavage-coupled transpeptidation reaction. The Cys sulfhydryl group is deprotonated by the His residue allowing for the Cys sulphur to serve as a nucleophile for attack on the carbonyl carbon of the Asp or Asn residue of the precursor peptide resulting in the cleavage of the tail sequence upon formation of an acyl intermediate. The His once again acts as a base to deprotonate the N-terminal Gly residue which serves as a nucleophile to resolve the acyl intermediate and form a new peptide bond and;

(***C***) The macrocyclic products.

A

Species (% identity)	Catalytic centre																Substrate specificity loop									
	57	58	59	60	159	160	161	**162**	163	200	201	202	203	**204**	205	206	227	230	231	232	233	234	235	236	251	254
A. thaliana	**Asn**	Y	R	H	Y	S	D	**His**	G	Y	L	E	A	**Cys**	E	S	N	E	S	S	W	G	T	Y	C	D
O. affinis (69.8%)	**Asn**	Y	R	H	Y	T	D	**His**	G	Y	L	E	A	**Cys**	E	S	N	E	S	S	W	C	Y	Y	C	D
C. ternatea (68.1%)	**Asn**	Y	R	H	Y	T	D	**His**	G	Y	V	E	S	**Cys**	E	S	D	E	S	S	W	V	T	Y	C	D
C. ensiformis (70.3%)	**Asn**	Y	R	H	Y	S	D	**His**	G	Y	I	E	A	**Cys**	E	S	N	E	N	S	F	G	T	Y	C	D
H. annuus (73.3%)	**Asn**	Y	R	H	Y	S	D	**His**	G	Y	L	E	A	**Cys**	E	S	G	E	N	S	Y	G	T	Y	C	D
M. musculus (54.3%)	**Asn**	Y	R	H	F	T	D	**His**	G	Y	I	E	A	**Cys**	E	S	N	E	S	S	Y	A	C	Y	Y	D
H. sapiens (53.9%)	**Asn**	Y	R	H	F	T	D	**His**	G	Y	I	E	A	**Cys**	E	S	N	E	S	S	Y	A	C	Y	Y	D

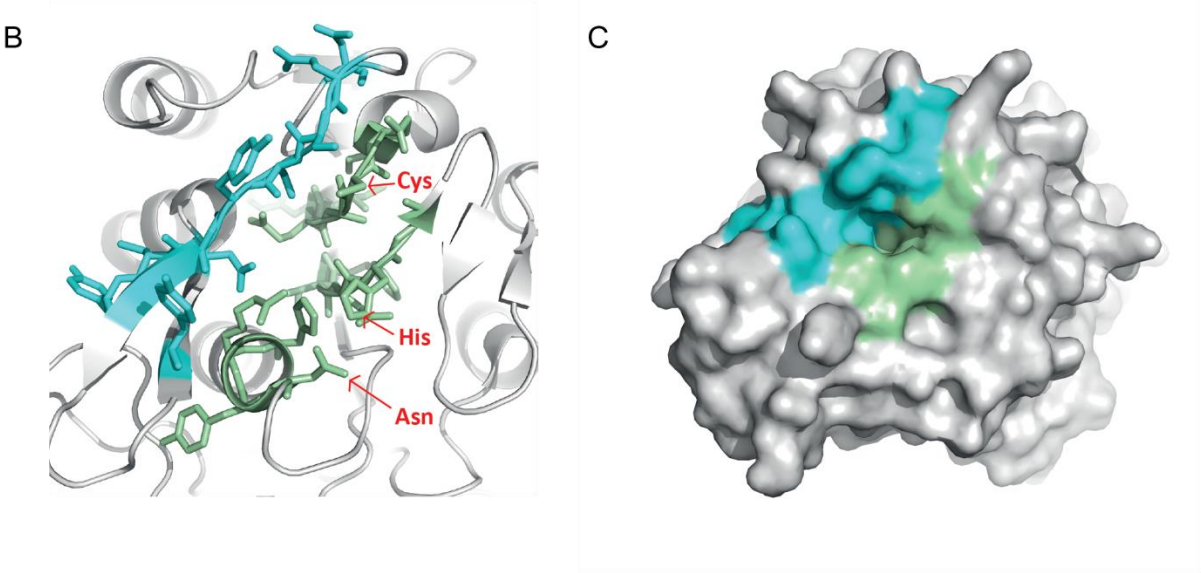

Figure 2. Sequence comparison of the catalytic centre and substrate specificity loop of AEPs as annotated by Dall and Brandsetter (2013).

(***A***) Summary of the residues involved in structural formation of the AEP active site in mammalian AEPs compared with plant AEPs. The catalytic centre (green) with the catalytic triad Asn, His and Cys (bold) shows high conservation across all residues (yellow background indicates conserved, dark grey indicates strongly similar properties, light grey indicates similar properties, white indicates non-conserved). Most differences between AEP active sites are within the substrate-specificity loop (blue).

(***B***) Cartoon representation of human AEP (PDB: 4AWB) active site with catalytic centre (green) and substrate specificity loop (blue) shown with sticks and catalytic triad labelled

(***C***) Surface representation of human AEP similarly coloured as in panel (B). Sequence analysis was carried out on predicted catalytic domain regions with ClustalO with *A. thaliana* used as the reference sequence. Accession numbers: P49047, ALG36103, AIB06797, P49046, AIZ09514, O89017, Q99538.