

## Review Article

# Plant asparaginyl endopeptidases and their structural determinants of function

Samuel G. Nonis<sup>1,2</sup>,  Joel Haywood<sup>1,2</sup> and  Joshua S. Mylne<sup>1,2</sup>

<sup>1</sup>School of Molecular Sciences, The University of Western Australia, 35 Stirling Highway, Crawley, Perth 6009, Australia; <sup>2</sup>The ARC Centre of Excellence in Plant Energy Biology, The University of Western Australia, 35 Stirling Highway, Crawley, Perth 6009, Australia

**Correspondence:** Joshua S. Mylne ([joshua.mylne@uwa.edu.au](mailto:joshua.mylne@uwa.edu.au))



Asparaginyl endopeptidases (AEPs) are versatile enzymes that in biological systems are involved in producing three different catalytic outcomes for proteins, namely (i) routine cleavage by bond hydrolysis, (ii) peptide maturation, including macrocyclisation by a cleavage-coupled intramolecular transpeptidation and (iii) circular permutation involving separate cleavage and transpeptidation reactions resulting in a major reshuffling of protein sequence. AEPs differ in their preference for cleavage or transpeptidation reactions, catalytic efficiency, and preference for asparagine or aspartate target residues. We look at structural analyses of various AEPs that have laid the groundwork for identifying important determinants of AEP function in recent years, with much of the research impetus arising from the potential biotechnological and pharmaceutical applications.

## Introduction

Asparaginyl endopeptidases (AEPs) are cysteine proteases of the C13 family (clan CD) that cleave at the carboxyl-terminal side of Asx (Asp or Asn). First discovered in legume plants [1], these enzymes were thus categorised as legumains (EC 3.4.22.34) [2] and are referred to as such, even for mammalian orthologues [3,4]. AEPs are widely distributed in land plants, where they are also sometimes called vacuolar processing enzymes due to their localisation in vacuoles and their role in processing vacuolar proteins [1,5].

The discovery of butelase-1, an extremely efficient AEP (catalytic efficiency of up to  $1\,340\,000\text{ M}^{-1}\text{ s}^{-1}$  on a non-native substrate) [6,7] that forms peptide bonds via a cleavage-coupled transpeptidation reaction, has led to a surge of interest in AEPs. Transpeptidation is an ATP-independent reaction. The putative mechanism in AEP involves an initial cleavage reaction, where an acyl-enzyme intermediate is first formed between the catalytic cysteine and the carbonyl carbon of the substrate P1 residue (nomenclature according to Schechter and Berger [8]). This intermediate is then resolved by nucleophilic attack by either a water molecule or a nearby N-terminal  $\alpha$ -amine, resulting in hydrolysis or aminolysis (transpeptidation), respectively [9,10]. Transpeptidation by plant AEPs is often intramolecular, in that the free N-terminal nucleophile originates from the same molecule in which the initial cleavage reaction occurs, resulting in peptide macrocyclisation. Many plant AEPs are therefore often referred to as cyclases as well. The term AEP is used to emphasise inherent peptidase activity, but the discovery of AEPs with efficient transpeptidase activity has led to certain AEPs being referred to as ligases or Peptide Asparaginyl Ligases instead; this is useful in emphasising the most notable enzyme activity but introduces a grey area for AEPs that have different enzyme efficiency and dominant activity depending on the pH and substrate [10–15].

Perhaps fitting for an enzyme with so many names, AEPs are also involved in processing a wide range of proteins in plants. They have been proposed to activate other proteases involved in seed storage protein processing [16]. Their role in vacuolar rupture to initiate a proteolytic cascade in programmed cell death has also been hypothesised to be through the activation of other vacuolar enzymes [17]. AEPs were for some time believed to also play a role in protein degradation during

Received: 18 December 2020  
 Revised: 9 February 2021  
 Accepted: 10 February 2021

Version of Record published:  
 5 March 2021

plant senescence, a process closely related to programmed cell death, but that has been disproven [18]. Here, we discuss the more well-characterised types of modifications performed by AEPs, including the maturation of seed storage proteins by hydrolysis, peptide processing and macrocyclisation, and the unique circular permutation of concanavalin A. Additionally, in light of knowledge generated from the recent wave of AEP structural analyses, we summarise recent studies on the structural features of AEP that influence their ability to perform transpeptidation.

## Maturation of seed storage proteins by hydrolysis

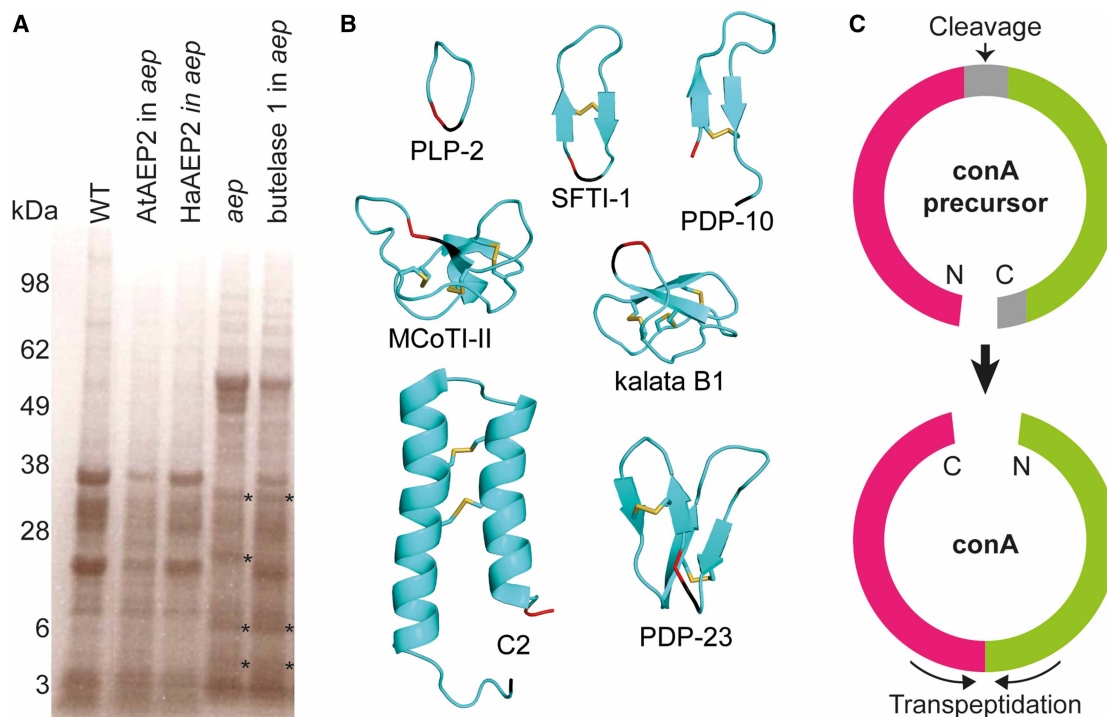
Seed storage proteins are a source of amino acids and elemental nutrients for the plant during germination [19]; they make up to half of the seed protein content and have no apparent function during seed development [20,21]. As the first systematic study on seed storage proteins was carried out by a chemist [22], these proteins were broadly classified based on a chemical property — their solubility in various solvents, namely the water-soluble albumins, salt-soluble globulins, alcohol-soluble prolamins, and acid- or alkali-soluble glutelins. Depending on the plant species, different types of storage proteins predominate [20,21].

It was the investigation into the maturation process of globulins that led to the discovery of asparagine-specific thiol proteases [1,23], which we now refer to as AEPs. During seed development, AEPs are thought to degrade unassembled or misfolded proteins [24], and have been shown to cleave seed storage proteins to facilitate their oligomeric assembly for final deposition in seeds [25–27]. AEP involvement in seed storage maturation was demonstrated *in vivo* with an *Arabidopsis thaliana* AEP quadruple knockout line, which produced alternatively processed storage proteins [28], that could be fully or partially rescued by transgenic expression of native and even non-native AEPs [29] (Figure 1A). The altered storage protein profile in the absence of AEPs is a combination of failure to cleave at the usual cut sites and abnormal cleavage at other sites by other proteases [30]. This is illustrated by the inability for 11S globulin to adopt the mature oligomeric complex conformation in the absence of AEP cleavage and is thus exposed to digestion by other proteases [31]. Unexpectedly, the altered seed storage protein profile in the AEP knockout *A. thaliana* line appears not to affect plant viability or development in any detectable way [28,30]. It remains to be seen if phenotypic effects can be discerned in AEP-deficient mutant plants when exposed to various stresses.

## Peptide processing and macrocyclisation

Macrocytic peptides are small backbone-cyclic peptides. These plant peptides are structurally diverse and typically less than 40 amino acids in size (Figure 1B). They are categorised into four broad classes: cyclotides or kalata-type cyclic peptides (e.g. kalata B1), which contain an embedded cystine knot of three disulfide bonds that can interact with lipid membranes [32]; cyclic knottins (e.g. MCoTI-II), often grouped with cyclotides as they also contain a cystine knot, but are more closely related to acyclic knottins and are protease inhibitors [33,34]; Preproalbumin with SFTI (PawS)-Derived Peptides (PDPs) (e.g. SFTI-1 and PDP-23), which are genetically embedded in the seed-storage albumin gene and contain one disulfide bond [35–37] while just PDP-23 contains two [38]; and orbitides (e.g. PLP-2), which are 5–16 amino acids in size and lack disulfide bonds [39,40]. The albumin-buried PLP orbitides appear to be evolutionary precursors of PDPs, and some PDPs have gained the ability to inhibit proteases as well [41,42]. Due to their observed *in vitro* bioactivity, many macrocyclic peptides are thought to confer resistance to pests and pathogens [32,35,43–46]. Macrocyclic peptides have also garnered considerable interest as drug scaffolds [47,48] as their small size, constrained rigid structure and absence of termini make them highly resistant to thermal, chemical, or enzymatic breakdown [33,36,49,50]. One of the most interesting scaffolds is that of the recently discovered and unusual PDP-23 (Figure 1B), which forms a symmetrical homodimer that dissociates upon interacting with membranes [38].

In nature, macrocyclic peptides are produced either by thioesterase-mediated non-ribosomal peptide synthesis [51] or by post-translational modification of ribosomally synthesised linear peptide precursors [9,10]. The latter process, which produces all four classes of macrocyclic peptides mentioned above, involves excising a short peptide from a protein chain via two cleavage reactions. AEP involvement in peptide macrocyclisation was initially inferred from a highly conserved Asx residue at the proto-C-terminus of peptide precursors [52–54]. This was confirmed *in vivo* by macrocyclisation of transgenic macrocyclic peptide precursors in plants devoid of macrocyclic peptides and demonstrating that inhibiting/silencing endogenous AEPs in the transformed plant results in failure to macrocyclise these precursors [34,37,55]. Several plant AEPs have since been characterised and shown to carry out peptide macrocyclisation *in vitro* [6,9,10,13,14,56]. Furthermore, a plant-based



**Figure 1. Types of modifications by plant AEPs.**

(A) AEPs process seed storage proteins. Total seed protein profile examined by SDS-PAGE demonstrate that expression of AtAEP2 (lane 2) and HaAEP2 (lane 3), which are inefficient transpeptidases, fully rescue the seed protein profile of *A. thaliana* *aep* quadruple null mutant (lane 4), resulting in a wild type seed protein profile (WT, lane 1). Expression of butelase 1, which is an efficient transpeptidase, only partially rescues the seed protein profile (lane 5). Asterisks indicate alternatively processed proteins. Refer to previous study for details [29]. (B) AEPs are implicated in the processing of a variety of cyclic and acyclic peptides. Examples of macrocyclic peptides include the orbitide PLP-2 (PDB: 6AXI), the PawS-Derived Peptides SFTI-1 (PDB: 6U7S) and PDP-23 [38], the cyclic knottin MCoTI-II (PDB: 1IB9), and the cyclotide kalata B1 (PDB: 1NB1). Examples of acyclic peptides include PDP-10 (PDB: 2NB6) and the vicilin-buried peptide C2 (6WQL). PDP-23 is unique among PDPs as it has two disulfide bonds. (Proto)-N-terminus shown in red. (Proto)-C-terminus shown in black. (C) Circular permutation of concanavalin A (conA) by jack bean AEP. AEP excises fifteen amino acids in the middle of the conA precursor and ligates the two proto-termini via a separate transpeptidation reaction, which involves cleaving off nine amino acids at the C terminus. Amino acids that are cleaved off are shown in grey in the precursor.

expression system has been recently developed, whereby transgenic AEPs and peptide precursors are co-expressed to produce macrocyclic peptides [57]. Orbitides are a diverse class, and cyclisation of some have been shown to be mediated by the serine protease peptide cyclase 1 in some plant species [58,59], but analogy to the PDPs and sequence conservation in the recently discovered large PLP family of orbitides suggest AEP involvement for these [46]. These findings suggest that macrocyclic peptides evolved by exploiting an already-existing biochemical machinery in plants [55]. It was hypothesised that this evolutionary convergence on AEPs was due to the reactive thioester acyl-intermediate that AEPs form after cleavage of the scissile peptide bond, and how the active site is structured to encourage interactions with an N-terminal nucleophile instead of water [34]. Cleavage at the proto-N-terminus of the peptide must therefore occur first so that it can participate in cleavage-coupled transpeptidation reaction at the proto-C-terminus [9,10,34,37,55]. Besides the conserved proto-C-terminal Asx residue, other hallmarks of AEP-mediated macrocyclisation are the highly conserved proto-N-terminal glycine, the small P1' residue and the hydrophobic P2' leucine [34].

The conserved Asx processing sites are found in acyclic peptides as well. PDP-10 (Figure 1B) is an example of a peptide similar to SFTI-1 but appears to undergo AEP-mediated hydrolysis rather than transpeptidation [36]. A recently discovered large group of non-cyclic peptides known as vicilin-buried peptides (e.g. C2) also contain

the conserved Asx processing sites [60]. These peptides are characterised by disulfide bonds forming between CXXXC motifs. Over 20 years ago, the prototypic member of what would be later called the vicilin-buried peptide family was shown *in vitro* to be matured by AEP cleavages [61], but the trypsin-inhibitory activity it displayed in 1999 was not reproducible with a synthetic version [62].

## Circular permutation of concanavalin A

Concanavalin A (conA), found in the seeds of jack bean plants (*Canavalia ensiformis*), is a carbohydrate-binding protein (lectin) that binds specifically and reversibly to mannose and glucose [63]. Even though conA has been functionalised in many ways with widespread use especially in glycoprotein purification [63–69], conA function *in vivo* is far from certain. ConA has been hypothesised to be involved in seed storage and plant defence, and their likely roles are mostly inferred from studies on other legume lectins [70–73]. The biosynthesis of conA and the highly-similar conA-like lectins found in plants that, like *C. ensiformis*, belong to the Diocleinae subtribe, involve a maturation process that employs the only known example of post-translational circular permutation in nature [74].

ConA circular permutation represents not only one of the earliest examples of protein splicing observed in nature [75,76], but also offered the first hint of the existence of an enzyme capable of transpeptidation in higher organisms [77]. Consequently, the jack bean AEP was one of the first AEPs to be studied extensively [78,79]. Circular permutation of conA involves cleaving the precursor of conA in the middle and ligating the two original termini via a transpeptidation reaction, resulting in a shuffling of the primary sequence such that the two halves of the protein are swapped [75] (Figure 1C). Circular permutation does not affect the general tertiary structure of conA or its functional sites, which is reflected in the apparently similar carbohydrate-binding abilities between conA and its precursor [73]. However, conA was recently shown to be more thermal and pH stable than its precursor *in vitro*, with a difference in complex formation in the crystal structures providing a possible explanation, revealing a functional consequence for what has been thought to be an inconsequential modification since its discovery 35 years ago [73,75].

## AEP structure

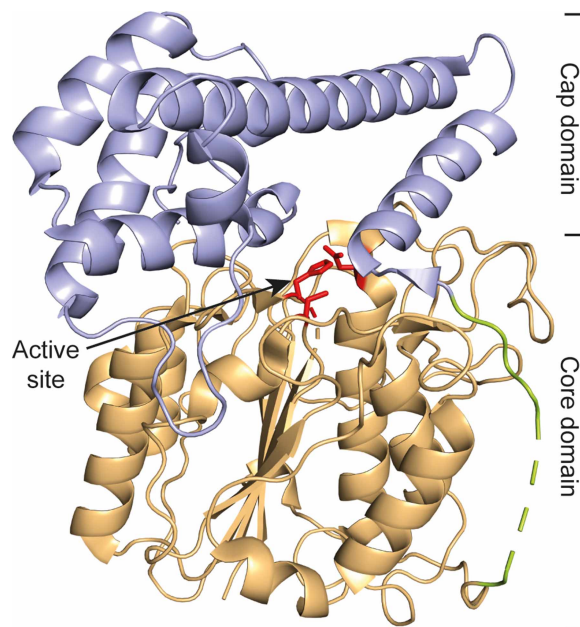
AEPs are synthesised in the inactive form consisting of the N-terminal ‘core’ domain, which contains the active site, and a smaller C-terminal auto-inhibitory ‘cap’ domain, which is also termed the Legumain Stabilisation and Activity Modulation domain as it blocks access to the active site, conferring enzymatic latency and protein stability at neutral pH conditions (Figure 2) [80]. Conversion of AEP into the active form has been studied in exquisite detail in two isoforms of *A. thaliana* AEP (AtAEP2 and AtAEP3) [80–82]. AEP activation occurs via autocatalytic cleavage around a flexible linker region between the cap and core domains, and requires disruption of electrostatic interactions between the two domains under acidic conditions [80–82].

The catalytic triad (or dyad) of all AEPs are structurally conserved (Figure 3). The catalytic cysteine was confirmed by covalent linking with a peptidic chloromethylketone-based inhibitor in human AEP [3]. The imidazole group of the catalytic histidine has been hypothesised to act as a general base. It has been proposed to activate the catalytic cysteine by deprotonating the cysteine thiol sidechain [83]. However, all plant AEP structures solved so far show that the catalytic cysteine and histidine sidechains are more than 5.4 Å apart. The catalytic histidine more likely acts upon the carbonyl carbon of the substrate directly or via a catalytic water in a pH-dependent manner [3,84]. In either scenario, the result is an acyl-enzyme intermediate that is resolved by hydrolysis or transpeptidation. A nearby asparagine makes up the last residue of the catalytic triad. It was hypothesised to facilitate substrate deprotonation by influencing the orientation of the catalytic histidine [85], but was shown to be a non-crucial catalytic residue in other cysteine proteases [86]. Studies on sunflower AEP1 (HaAEP1), a predominant hydrolase, showed that a catalytic cysteine to serine mutation abolished enzyme activity, a catalytic histidine to alanine mutation drastically reduced activity, and a catalytic asparagine to alanine mutation resulted in a higher transpeptidation to hydrolysis ratio [11].

## Structural features that influence AEP function

Unlike the specialised enzyme PatG, which evolved an additional dedicated protein domain to facilitate transpeptidation in cyanobacteria by shielding its reaction intermediates from water [87], AEPs that are efficient at transpeptidation do not possess an obvious additional domain. The structural features that convert an AEP from a predominant protease to a predominant transpeptidase must be discerned from subtle differences in the





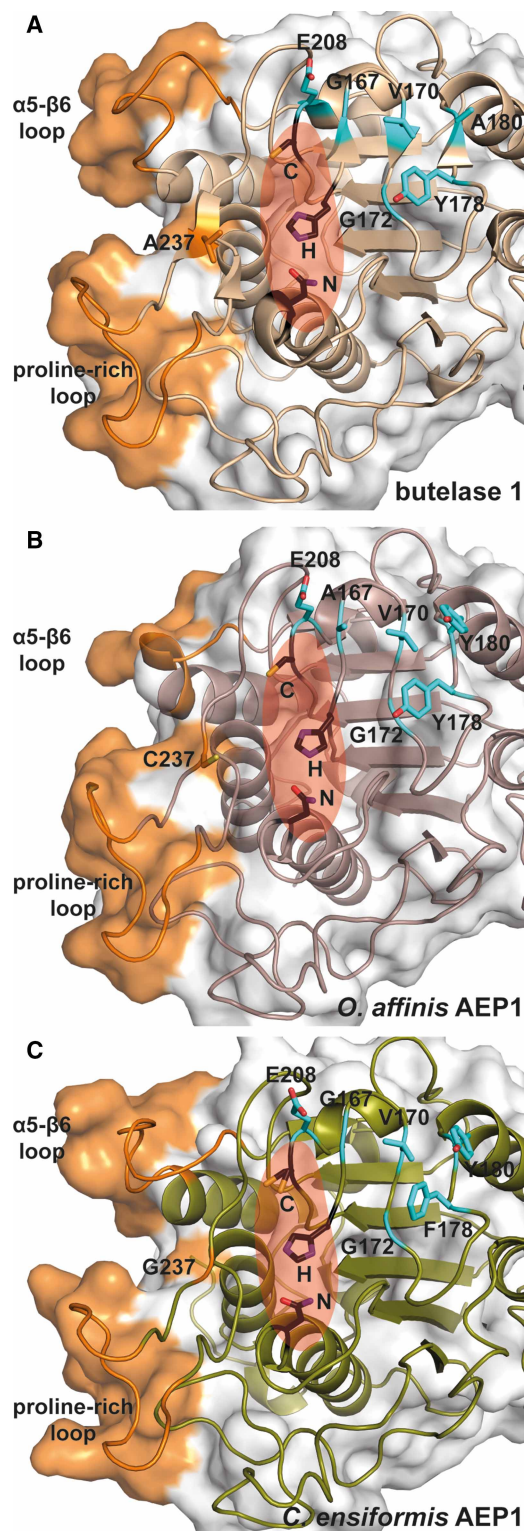
**Figure 2. AEP structure.**

Cartoon representation of inactive butelase 1 (PDB: 6DHI [29]) illustrating the core domain (wheat), cap domain (purple), flexible linker region (green) and active site (red). Association of the cap and core domains occludes the active site, conferring inactivity and stability. Dashed line in linker region represents region of low electron density.

substrate-binding pocket. Delineation of the substrate-binding sites was made possible by the structural elucidation of activated plant AEPs in covalent complex with various substrates [11,12].

To facilitate comparison of equivalent residues amongst different AEPs, all residue numberings used in this review are based on the butelase 1 sequence. Residues around the active site have been proposed to facilitate interaction between the acyl-enzyme intermediate and the N-terminal nucleophile. Asn59, which is part of the catalytic triad, and Glu208 (Figure 3) were shown to influence transpeptidation efficiency in HaAEP1 [11]. Residue 59 was hypothesised to influence the conformational flexibility of the localised catalytic histidine, a smaller sidechain here may therefore reduce steric hindrance for nucleophilic attack on the acyl-enzyme intermediate by an incoming N-terminal residue [11]. Glu208 was shown in docking studies with an AtAEP to form an ionic interaction with the N-terminal nucleophile [12] and was suggested to deprotonate the incoming N-terminus in the transpeptidation reaction [11].

The primed region of the AEP active site, i.e. AEP residues that interact with substrate residues proceeding the scissile peptide bond, has been hypothesised and corroborated with molecular dynamics analyses to promote transpeptidation by maintaining affinity with the cleaved substrate until they are displaced by an incoming N-terminal nucleophile [12–14]. Mutating residue 167 in the S1' pocket (Figure 3) to an amino acid with a small sidechain resulted in a higher transpeptidation/hydrolysis ratio and increased catalytic efficiency in two AEPs from different plant species [13]. It was suggested that a bulky sidechain at this position interferes with proper replacement of the leaving group by the N-terminal nucleophile, thus allowing water to enter the active site [13]. The highly conserved hydrophobic P2' residue in cyclic peptide precursors [10,52,88] and the conserved deep hydrophobic S2' pocket (residues 170, 172, 178, 180) (Figure 3) indicate that the S2'-P2' interaction is crucial for transpeptidation. The depth of the S2' pocket is attributable to the lack of sidechain at residue 172 and a large aromatic sidechain at residue 178. Molecular dynamic simulations showed that non-hydrophobic P2' residues resulted in poor retention of residues C-terminal of the scissile bond [12]. Reduction in hydrophobicity at the S2' pocket also negatively impacts cyclisation —HaAEP1 and AtAEP2, which have low to intermediate cyclisation efficiencies, contain a histidine rather than an aromatic amino acid at residue 178 [9,11,12]. The S2' pocket was also hypothesised to shield the active site from water [56].



**Figure 3. Structural features that influence AEP activity.**

Part 1 of 2

Cartoon and surface representation of the substrate binding site of butelase 1 (PDB: 6DHI [29]) from *Clitoria ternatea* (A), *Oldenlandia affinis* AEP1 (PDB: 5H0I [89]) (B), and *C. ensiformis* AEP1 (PDB: 6XT5 [73]) (C). Catalytic residues cysteine (C), histidine (H) and asparagine (N) are represented as black sticks and shaded in red. Residue 237<sup>butelase1</sup>,  $\alpha$ 5- $\beta$ 6 loop and proline-rich loop (orange surface) are on the left of the active site and interact with substrate residues preceding the scissile

**Figure 3. Structural features that influence AEP activity.**

Part 2 of 2

peptide bond (non-primed region). Residues 167, 170, 172, 178, 180 and 208 (butelase 1 numbering) (cyan sticks) are on the right of the active site and interact with substrate residues preceding the scissile peptide bond (primed region). Residue numbering for *O. affinis* AEP1 and *C. ensiformis* AEP1 are based on butelase 1 numbering to facilitate comparison.

Several residues in the non-primed region, i.e. AEP residues that interact with substrate residues preceding the scissile peptide bond, have been shown to influence transpeptidation efficiency. Substrate affinity, hypothesised by some to positively influence transpeptidase activity, was attributed to the non-primed region [81]. However, the complex interplay amongst the residues in this region makes it difficult to rationalise mechanistic explanations for their influence on AEP activity. Residue 237, the  $\alpha 5$ – $\beta 6$  loop, and the proline-rich loop (also previously referred to as the ‘Gate-keeper residue’, ‘Marker of Ligase Activity’, and ‘poly-proline loop’, respectively) of the non-primed region have been the most widely discussed (Figure 3). One of the most prolific single point mutations was of that performed at residue 237 in *Oldenlandia affinis* AEP1 (OaAEP1), which improved cyclisation kinetics ( $k_{\text{cat}}$ ) by 160-fold [89]. However, mutating this residue to the same residues that are found in other efficient transpeptidases resulted in poorer transpeptidation efficiency in OaAEP1 [6,13,89], but improved cyclisation ratio of a dominant hydrolase in a *Clitoria ternatea* AEP [15]. In another study, mutating residue 237 did not improve transpeptidation in a petunia AEP unless coupled with mutations in the  $\alpha 5$ – $\beta 6$  loop [90]. This flexible  $\alpha 5$ – $\beta 6$  loop was shown to influence the yield of macrocyclic peptides, but not the activity preference as the deletion of this loop in the dominant ligase OaAEP1 did not result in enhanced hydrolysis, indicating that it likely influences substrate affinity, but is not a determinant of AEP functional preference [90]. The large difference in  $\alpha 5$ – $\beta 6$  loop size in two extremely efficient AEP transpeptidases, butelase 1 (Figure 3A) and OaAEP1, (Figure 3B) suggests that transpeptidation efficiency is not necessarily correlated to the size of the  $\alpha 5$ – $\beta 6$  loop. Furthermore, AEPs that are inefficient at transpeptidation, such as *C. ensiformis* AEP1 (Figure 3C), have a similar  $\alpha 5$ – $\beta 6$  loop size to butelase 1. The more hydrophobic characteristic of the  $\alpha 5$ – $\beta 6$  loop in butelase 1, as compared to other inefficient transpeptidases, was postulated to be a distinguishing feature of an efficient transpeptidase [90]. The conserved proline-rich loop has been shown to interact with the substrate non-primed region [12]. However, it is not known if substrate-recognition at the proline-rich loop influences transpeptidation efficiency.

AEP activity is also substrate- and condition-dependent [10–14]. AEPs such as HaAEP1 that have been structurally and biochemically characterised as predominant peptidases have turned out to also have high cyclic to linear product ratio when assayed with different substrates [14]. Asparagine/aspartate preference is pH-dependent — aspartate was hypothesised to be accepted as a P1 residue below pH 5.5 because protonating the aspartate carboxylic sidechain makes it sterically and electrostatically similar to asparagine, which is the preferred P1 residue at higher pH conditions [3,81,91]. Studies on AtAEP3 showed that transpeptidation activity is typically confined to near-neutral pH conditions [80] where N-terminal amino groups are more likely deprotonated, either on their own or by a deprotonated catalytic histidine, and therefore able to perform nucleophilic attack on the acyl-enzyme intermediate [11]. Zauner and colleagues [80] have hypothesised that the association of the cap and core domain in AtAEP3 results in the exclusion of water from the active site, and that a deprotonated N-terminal nucleophile can interfere with this association to resolve the acyl-enzyme intermediate.

## Conclusion

Recent studies highlight the fact that the preferred AEP activity and catalytic efficiency are dependent not only on the structural features of a given AEP, but also on the nature of the substrate and the pH condition. This is illustrated in the structural diversity of both the peptide substrates (Figure 1B) and AEP substrate binding sites (Figure 3). Mutagenesis studies point to a complex interplay amongst residues of the non-primed region in the AEP substrate binding site, as single substitution of a residue or a loop in this region do not consistently result in the expected enhancement of transpeptidase activity, whereas multiple substitutions are more likely to do so [14,56]. Nevertheless, developments in understanding AEP function, especially that of butelase 1 from *Clitoria ternatea* and a mutant variant of OaAEP1 from *Oldenlandia affinis*, have led to their use as protein engineering tools especially for biotechnological and pharmaceutical applications [7,92–98]. These AEPs are efficient transpeptidases that leave only a P1 Asx residue and typically an aliphatic P2'' residue, such as isoleucine, leucine or valine, as processing artefacts [6,10,90,92]. The broad specificity for any nucleophilic P1'' amino acid except

proline [6,99] in the transpeptidation reaction makes AEPs ideal enzymes for a variety of protein or peptide substrates. However, this promiscuity comes at a cost as ingenious workarounds are required to minimise nucleophilic attack by the cleaved by-product [92,100,101]. The study and genetic engineering of AEP provides an alternative to the chemical ligation of peptides and is therefore crucial for developing novel macrocyclic peptides for use in agriculture and medicine. Even though applications of synthetic macrocyclic peptides have to date been modest, much research continues to evaluate their potential as drug leads for human diseases [102–105].

## Perspectives

- Asparaginyl endopeptidases perform a variety of functions in plants. There is enormous potential to harness their versatility for protein engineering, especially for the development of peptidic compounds and drugs.
- Recent discoveries and developments have improved our understanding of the structural features that can make an asparaginyl endopeptidase an efficient transpeptidase.
- This knowledge will aid not only our search for better transpeptidases but also the development of the current known ones into effective biotechnological tools.

## Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

## Funding

This work was supported in part by Australian Research Council (ARC) grant DP160100107 to J.S.M.

J.H. was supported by an ARC Discovery Early Career Researcher Award DE180101445. S.G.N. was supported by the Australian Research Training Program.

## Open Access Statement

Open access for this article was enabled by the participation of University of Western Australia in an all-inclusive *Read & Publish* pilot with Portland Press and the Biochemical Society under a transformative agreement with CAUL.

## Author Contributions

S.G.N. and J.S.M. conceptualised the manuscript. S.G.N. drafted the manuscript. J.H. assisted with generating figures. S.G.N., J.S.M., and J.H. revised the manuscript.

## Acknowledgements

The authors thank Amy James for providing an SDS gel image of the seed protein profiles which is similar to the SDS gel image used in her published work in *Plant Journal* [29].

## Abbreviations

AEPs, asparaginyl endopeptidases; conA, concanavalin A; PDPs, PawS-Derived Peptides; AtAEP, *Arabidopsis thaliana* AEP; OaAEP1, *Oldenlandia affinis* AEP1; HaAEP, *Helianthus annuus* AEP.

## References

- 1 Hara-Nishimura, I., Inoue, K. and Nishimura, M. (1991) A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. *FEBS Lett.* **294**, 89–93 [https://doi.org/10.1016/0014-5793\(91\)81349-D](https://doi.org/10.1016/0014-5793(91)81349-D)
- 2 Kembhavi, A.A., Buttle, D.J., Knight, C.G. and Barrett, A.J. (1993) The two cysteine endopeptidases of legume seeds - purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* **303**, 208–213 <https://doi.org/10.1006/abbi.1993.1274>



- 3 Dall, E. and Brandstetter, H. (2013) Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proc. Natl Acad. Sci. U.S.A.* **110**, 10940–10945 <https://doi.org/10.1073/pnas.1300686110>
- 4 Zhao, L., Hua, T., Crowley, C., Ru, H., Ni, X., Shaw, N. et al. (2014) Structural analysis of asparaginyl endopeptidase reveals the activation mechanism and a reversible intermediate maturation stage. *Cell Res.* **24**, 344–358 <https://doi.org/10.1038/cr.2014.4>
- 5 Yamada, K., Basak, A.K., Goto-Yamada, S., Tarnawska-Glatt, K. and Hara-Nishimura, I. (2020) Vacuolar processing enzymes in the plant life cycle. *New Phytol.* **226**, 21–31 <https://doi.org/10.1111/nph.16306>
- 6 Nguyen, G.K.T., Wang, S., Qiu, Y., Hemu, X., Lian, Y. and Tam, J.P. (2014) Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol.* **10**, 732–738 <https://doi.org/10.1038/nchembio.1586>
- 7 Nguyen, G.K.T., Kam, A., Loo, S., Jansson, A.E., Pan, L.X. and Tam, J.P. (2015) Butelase 1: a versatile ligase for peptide and protein macrocyclization. *J. Am. Chem. Soc.* **137**, 15398–15401 <https://doi.org/10.1021/jacs.5b11014>
- 8 Schechter, I. and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162 [https://doi.org/10.1016/S0006-291X\(67\)80055-X](https://doi.org/10.1016/S0006-291X(67)80055-X)
- 9 Bernath-Levin, K., Nelson, C., Elliott, A.G., Jayasena, A.S., Millar, A.H., Craik, D.J. et al. (2015) Peptide macrocyclization by a bifunctional endoprotease. *Chem. Biol.* **22**, 571–582 <https://doi.org/10.1016/j.chembiol.2015.04.010>
- 10 Harris, K.S., Durek, T., Kaas, Q., Poth, A.G., Gilding, E.K., Conlan, B.F. et al. (2015) Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat. Commun.* **6**, 10199 <https://doi.org/10.1038/ncomms10199>
- 11 Haywood, J., Schmidberger, J.W., James, A.M., Nonis, S.G., Sukhovkov, K.V., Elias, M. et al. (2018) Structural basis of ribosomal peptide macrocyclization in plants. *eLife* **7**, e32955 <https://doi.org/10.7554/eLife.32955>
- 12 Zauner, F.B., Elsässer, B., Dall, E., Cabrele, C. and Brandstetter, H. (2018) Structural analyses of *Arabidopsis thaliana* legumain gamma reveal the differential recognition and processing of proteolysis and ligation substrates. *J. Biol. Chem.* **293**, 8934–8946 <https://doi.org/10.1074/jbc.M117.817031>
- 13 Hemu, X., El Sahili, A., Hu, S., Wong, K., Chen, Y., Wong, Y.H. et al. (2019) Structural determinants for peptide-bond formation by asparaginyl ligases. *Proc. Natl Acad. Sci. U.S.A.* **116**, 11737–11746 <https://doi.org/10.1073/pnas.1818568116>
- 14 Du, J., Yap, K., Chan, L.Y., Rehm, F.B.H., Looi, F.Y., Poth, A.G. et al. (2020) A bifunctional asparaginyl endopeptidase efficiently catalyzes both cleavage and cyclization of cyclic trypsin inhibitors. *Nat. Commun.* **11**, 1575 <https://doi.org/10.1038/s41467-020-15418-2>
- 15 Hemu, X., El Sahili, A., Hu, S.D., Zhang, X.H., Serra, A., Goh, B.C. et al. (2020) Turning an asparaginyl endopeptidase into a peptide ligase. *ACS Catal.* **10**, 8825–8834 <https://doi.org/10.1021/acscatal.0c02078>
- 16 Okamoto, T. and Minamikawa, T. (1999) Molecular cloning and characterization of *Vigna mungo* processing enzyme 1 (VmPE-1), an asparaginyl endopeptidase possibly involved in post-translational processing of a vacuolar cysteine endopeptidase (SH-EP). *Plant Mol. Biol.* **39**, 63–73 <https://doi.org/10.1023/A:1006170518002>
- 17 Hatsugai, N., Yamada, K., Goto-Yamada, S. and Hara-Nishimura, I. (2015) Vacuolar processing enzyme in plant programmed cell death. *Front. Plant Sci.* **6**, 234 <https://doi.org/10.3389/fpls.2015.00234>
- 18 Pruzinska, A., Shindo, T., Niessen, S., Kaschani, F., Toth, R., Millar, A.H. et al. (2017) Major Cys protease activities are not essential for senescence in individually darkened *Arabidopsis* leaves. *BMC Plant Biol.* **17**, 4 <https://doi.org/10.1186/s12870-016-0955-5>
- 19 Barber, J.T. and Boulter, D. (1963) Argininosuccinic acid in germinating seeds of *Vicia faba* L. *Nature* **197**, 1112 <https://doi.org/10.1038/1971112a0>
- 20 Higgins, T.J., Chandler, P.M., Randall, P.J., Spencer, D., Beach, L.R., Blagrove, R.J. et al. (1986) Gene structure, protein structure, and regulation of the synthesis of a sulfur-rich protein in pea seeds. *J. Biol. Chem.* **261**, 11124–11130 [https://doi.org/10.1016/S0021-9258\(18\)67357-0](https://doi.org/10.1016/S0021-9258(18)67357-0)
- 21 Shewry, P.R., Napier, J.A. and Tatham, A.S. (1995) Seed storage proteins: structures and biosynthesis. *Plant Cell.* **7**, 945–956 <https://doi.org/10.1105/tpc.7.7.945>
- 22 Osborne, T.B. (1908) Our present knowledge of plant proteins. *Science* **28**, 417–427 <https://doi.org/10.1126/science.28.718.417>
- 23 Hara-Nishimura, I. and Nishimura, M. (1987) Proglubulin processing enzyme in vacuoles isolated from developing pumpkin cotyledons. *Plant Physiol.* **85**, 440–445 <https://doi.org/10.1104/pp.85.2.440>
- 24 Jung, R., Saalbach, G., Nielsen, N.C. and Müntz, K. (1993) Site-specific limited proteolysis of legumin chloramphenicol acetyl transferase fusions *in vitro* and in transgenic tobacco seeds. *J. Exp. Bot.* **44**, 343–349 <https://www.jstor.org/stable/23694172>
- 25 Jung, R., Scott, M.P., Nam, Y.-W., Beaman, T.W., Bassüner, R., Saalbach, I. et al. (1998) The role of proteolysis in the processing and assembly of 11S seed globulins. *Plant Cell* **10**, 343–357 <https://doi.org/10.1105/tpc.10.3.343>
- 26 Adachi, M., Takenaka, Y., Gidamis, A.B., Mikami, B. and Utsumi, S. (2001) Crystal structure of soybean proglucinin A1aB1b homotrimer. *J. Mol. Biol.* **305**, 291–305 <https://doi.org/10.1006/jmbi.2000.4310>
- 27 Adachi, M., Kanamori, J., Masuda, T., Yagasaki, K., Kitamura, K., Mikami, B. et al. (2003) Crystal structure of soybean 11S globulin: glycinin A3B4 homohexamer. *Proc. Natl Acad. Sci. U.S.A.* **100**, 7395–7400 <https://doi.org/10.1073/pnas.0832158100>
- 28 Gruis, D., Schulze, J. and Jung, R. (2004) Storage protein accumulation in plants. *Trends Plant Sci.* **7**, 340–344 [https://doi.org/10.1016/S1360-1385\(02\)02298-7](https://doi.org/10.1016/S1360-1385(02)02298-7)
- 29 James, A.M., Haywood, J., Leroux, J., Ignasiak, K., Elliott, A.G., Schmidberger, J.W. et al. (2019) The macrocyclizing protease butelase 1 remains autocatalytic and reveals the structural basis for ligase activity. *Plant J.* **98**, 988–999 <https://doi.org/10.1111/tpj.14293>
- 30 Shimada, T., Yamada, K., Kataoka, M., Nakaune, S., Koumoto, Y., Kuroyanagi, M. et al. (2003) Vacuolar processing enzymes are essential for proper processing of seed storage proteins in *Arabidopsis thaliana*. *J. Biol. Chem.* **278**, 32292–32299 <https://doi.org/10.1074/jbc.M305740200>
- 31 Müntz, K. and Shutov, A.D. (2002) Legumains and their functions in plants. *Trends Plant Sci.* **7**, 340–344 [https://doi.org/10.1016/S1360-1385\(02\)02298-7](https://doi.org/10.1016/S1360-1385(02)02298-7)
- 32 de Veer, S.J., Kan, M.W. and Craik, D.J. (2019) Cyclotides: from structure to function. *Chem. Rev.* **119**, 12375–12421 <https://doi.org/10.1021/acs.chemrev.9b00402>
- 33 Kolmar, H. (2008) Alternative binding proteins: biological activity and therapeutic potential of cystine-knot miniproteins. *FEBS J.* **275**, 2684–2690 <https://doi.org/10.1111/j.1742-4658.2008.06440.x>
- 34 Mylne, J.S., Chan, L.Y., Chanson, A.H., Daly, N.L., Schaefer, H., Bailey, T.L. et al. (2012) Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis. *Plant Cell* **24**, 2765–2778 <https://doi.org/10.1105/tpc.112.099085>
- 35 Luckett, S., Garcia, R.S., Barker, J.J., Konarev, A.V., Shewry, P.R., Clarke, A.R. et al. (1999) High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *J. Mol. Biol.* **290**, 525–533 <https://doi.org/10.1006/jmbi.1999.2891>

- 36 Franke, B., Mylne, J.S. and Rosengren, K.J. (2018) Buried treasure: biosynthesis, structures and applications of cyclic peptides hidden in seed storage albumins. *Nat. Prod. Rep.* **35**, 137–146 <https://doi.org/10.1039/C7NP00066A>
- 37 Mylne, J.S., Colgrave, M.L., Daly, N.L., Chanson, A.H., Elliott, A.G., McCallum, E.J. et al. (2011) Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nat. Chem. Biol.* **7**, 257–259 <https://doi.org/10.1038/nchembio.542>
- 38 Payne, C.D., Franke, B., Fisher, M.F., Hajiaghaalipour, F., McAleese, C.E., Song, A. et al. (2020) A chameleonic macrocyclic peptide with drug delivery applications. *bioRxiv* 2020:2020.12.16.422786 <https://doi.org/10.1101/2020.12.16.422786>
- 39 Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G. et al. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 <https://doi.org/10.1039/C2NP20085F>
- 40 Fisher, M.F., Payne, C.D., Rosengren, K.J. and Mylne, J.S. (2019) An orbitide from *Ratibida columnifera* seed containing 16 amino acid residues. *J. Nat. Prod.* **82**, 2152–2158 <https://doi.org/10.1021/acs.jnatprod.9b00111>
- 41 Elliott, A.G., Delay, C., Liu, H., Phua, Z., Rosengren, K.J., Benfield, A.H. et al. (2014) Evolutionary origins of a bioactive peptide buried within preproalbumin. *Plant Cell* **26**, 981–995 <https://doi.org/10.1105/tpc.114.123620>
- 42 Jayasena, A.S., Fisher, M.F., Panero, J.L., Secco, D., Bernath-Levin, K., Berkowitz, O. et al. (2017) Stepwise evolution of a buried inhibitor peptide over 45 My. *Mol. Biol. Evol.* **34**, 1505–1516 <https://doi.org/10.1093/molbev/msx104>
- 43 Tan, N.-H. and Zhou, J. (2006) Plant cyclopeptides. *Chem. Rev.* **106**, 840–895 <https://doi.org/10.1021/cr040699h>
- 44 Craik, D.J. (2012) Host-defense activities of cyclotides. *Toxins (Basel)* **4**, 139–156 <https://doi.org/10.3390/toxins4020139>
- 45 Tian, J., Shen, Y., Yang, X., Liang, S., Shan, L., Li, H. et al. (2010) Antifungal cyclic peptides from *Psammosilene tunicoides*. *J. Nat. Prod.* **73**, 1987–1992 <https://doi.org/10.1021/np100363a>
- 46 Fisher, M.F., Zhang, J., Taylor, N.L., Howard, M.J., Berkowitz, O., Debowski, A.W. et al. (2018) A family of small, cyclic peptides buried in preproalbumin since the Eocene epoch. *Plant Direct* **2**, e00042 <https://doi.org/10.1002/pld3.42>
- 47 Lesner, A., Łęgoska, A., Wysocka, M. and Rolka, K. (2011) Sunflower trypsin inhibitor 1 as a molecular scaffold for drug discovery. *Curr. Pharm. Des.* **17**, 4308–4317 <https://doi.org/10.2174/138161211798999393>
- 48 Gould, A., Ji, Y., Aboye, T.L. and Camarero, J.A. (2011) Cyclotides, a novel ultrastable polypeptide scaffold for drug discovery. *Curr. Pharm. Des.* **17**, 4294–4307 <https://doi.org/10.2174/138161211798999438>
- 49 Colgrave, M.L. and Craik, D.J. (2004) Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. *Biochemistry* **43**, 5965–5975 <https://doi.org/10.1021/bi049711q>
- 50 Heitz, A., Avrutina, O., Le-Nguyen, D., Diederichsen, U., Hernandez, J.F., Gracy, J. et al. (2008) Knottin cyclization: impact on structure and dynamics. *BMC Struct. Biol.* **8** <https://doi.org/10.1186/1472-6807-8-54>
- 51 Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A. and Walsh, C.T. (2000) Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **407**, 215–218 <https://doi.org/10.1038/35025116>
- 52 Gillon, A.D., Saska, I., Jennings, C.V., Guarino, R.F., Craik, D.J. and Anderson, M.A. (2008) Biosynthesis of circular proteins in plants. *Plant J.* **53**, 505–515 <https://doi.org/10.1111/j.1365-313X.2007.03357.x>
- 53 Jennings, C., West, J., Waine, C., Craik, D. and Anderson, M. (2001) Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis*. *Proc. Natl Acad. Sci. U.S.A.* **98**, 10614–10619 <https://doi.org/10.1073/pnas.191366898>
- 54 Dutton, J.L., Renda, R.F., Waine, C., Clark, R.J., Daly, N.L., Jennings, C.V. et al. (2004) Conserved structural and sequence elements implicated in the processing of gene-encoded circular proteins. *J. Biol. Chem.* **279**, 46858–46867 <https://doi.org/10.1074/jbc.M407421200>
- 55 Saska, I., Gillon, A.D., Hatsugai, N., Dietzgen, R.G., Hara-Nishimura, I., Anderson, M.A. et al. (2007) An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclisation. *J. Biol. Chem.* **282**, 29721–29728 <https://doi.org/10.1074/jbc.M705185200>
- 56 Harris, K.S., Guarino, R.F., Dissanayake, R.S., Quimbar, P., McCorkelle, O.C., Poon, S. et al. (2019) A suite of kinetically superior AEP ligases can cyclise an intrinsically disordered protein. *Sci. Rep.* **9**, 10820 <https://doi.org/10.1038/s41598-019-47273-7>
- 57 Poon, S., Harris, K.S., Jackson, M.A., McCorkelle, O.C., Gilding, E.K., Durek, T. et al. (2018) Co-expression of a cyclizing asparaginyl endopeptidase enables efficient production of cyclic peptides in plants. *J. Exp. Bot.* **69**, 633–641 <https://doi.org/10.1093/jxb/erx422>
- 58 Barber, C.J.S., Pujara, P.T., Reed, D.W., Chivocha, S., Zhang, H. and Covello, P.S. (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. *J. Biol. Chem.* **288**, 12500–12510 <https://doi.org/10.1074/jbc.M112.437947>
- 59 Chekan, J.R., Estrada, P., Covello, P.S. and Nair, S.K. (2017) Characterization of the macrocyclase involved in the biosynthesis of RiPP cyclic peptides in plants. *Proc. Natl Acad. Sci. U.S.A.* **114**, 6551–6556 <https://doi.org/10.1073/pnas.1620499114>
- 60 Zhang, J., Payne, C.D., Pouvreau, B., Schaefer, H., Fisher, M.F., Taylor, N.L. et al. (2019) An ancient peptide family buried within vicilin precursors. *ACS Chem. Biol.* **14**, 979–993 <https://doi.org/10.1021/acschembio.9b00167>
- 61 Yamada, K., Shimada, T., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (1999) Multiple functional proteins are produced by cleaving Asn-Gln bonds of a single precursor by vacuolar processing enzyme. *J. Biol. Chem.* **274**, 2563–2570 <https://doi.org/10.1074/jbc.274.4.2563>
- 62 Payne, C.D., Vadlamani, G., Fisher, M.F., Zhang, J., Clark, R.J., Mylne, J.S. et al. (2020) Defining the familial fold of the vicilin-buried peptide family. *J. Nat. Prod.* **83**, 3030–3040 <https://doi.org/10.1021/acs.jnatprod.0c00594>
- 63 Lis, H. and Sharon, N. (1998) Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637–674 <https://doi.org/10.1021/cr940413g>
- 64 Bernhard, W. and Avrameas, S. (1971) Ultrastructural visualization of cellular carbohydrate components by means of concanavalin A. *Exp. Cell Res.* **64**, 232–236 [https://doi.org/10.1016/0014-4827\(71\)90217-5](https://doi.org/10.1016/0014-4827(71)90217-5)
- 65 Ogata, S., Muramatsu, T. and Kobata, A. (1975) Fractionation of glycopeptides by affinity column chromatography on concanavalin A-sepharose. *J. Biochem.* **78**, 687–696 <https://doi.org/10.1093/oxfordjournals.jbchem.a130956>
- 66 Dwyer, J.M. and Johnson, C. (1981) The use of concanavalin A to study the immunoregulation of human T cells. *Clin. Exp. Immunol.* **46**, 237–249 PMID: 6461456
- 67 Goldstein, J.I., Winter, C.H. and Poretz, D.R. (1997) Plant lectins: tools for the study of complex carbohydrates. In *Glycoproteins II*, 1 edn (Montreuil, J., Vliegthart, J.F.G., Schachter, H., eds), vol. 29, pp. 403–474, Elsevier B.V, Amsterdam, Netherlands

- 68 Locke, A.K., Cummins, B.M., Abraham, A.A. and Coté, G.L. (2014) PEGylation of concanavalin A to improve its stability for an *in vivo* glucose sensing assay. *Anal. Chem.* **86**, 9091–9097 <https://doi.org/10.1021/ac501791u>
- 69 Saleemuddin, M. and Husain, Q. (1991) Concanavalin A: a useful ligand for glycoenzyme immobilization—a review. *Enzyme Microb. Technol.* **13**, 290–295 [https://doi.org/10.1016/0141-0229\(91\)90146-2](https://doi.org/10.1016/0141-0229(91)90146-2)
- 70 Dalkin, K. and Bowles, D.J. (1983) Analysis of inter-relationship of jackbean seed components by two-dimensional mapping of iodinated tryptic peptides. *Planta* **157**, 536–539 <https://doi.org/10.1007/BF00396885>
- 71 Sharon, N. and Lis, H. (1990) Legume lectins—a large family of homologous proteins. *FASEB J.* **4**, 3198–3208 <https://doi.org/10.1096/fasebj.4.14.2227211>
- 72 Peumans, W.J. and Van Damme, E.J. (1995) Lectins as plant defense proteins. *Plant. Physiol.* **109**, 347–352 <https://doi.org/10.1104/pp.109.2.347>
- 73 Nonis, S.G., Haywood, J., Schmidberger, J.W., Bond, C.S. and Mylne, J.S. (2020) Structural basis for a natural circular permutation in proteins. *bioRxiv* 2020.10.28.360099 <https://doi.org/10.1101/2020.10.28.360099>
- 74 Cavada, B.S., Pinto-Junior, V.R., Osterne, V.J.S. and Nascimento, K.S. (2018) ConA-like lectins: high similarity proteins as models to study structure/biological activities relationships. *Int. J. Mol. Sci.* **20**, 30 <https://doi.org/10.3390/ijms20010030>
- 75 Carrington, D.M., Auffret, A. and Hanke, D.E. (1985) Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature* **313**, 64–67 <https://doi.org/10.1038/313064a0>
- 76 Hendrix, R.W. (1991) Protein carpentry. *Curr. Biol.* **1**, 71–73 [https://doi.org/10.1016/0960-9822\(91\)90280-A](https://doi.org/10.1016/0960-9822(91)90280-A)
- 77 Bowles, D.J., Marcus, S.E., Pappin, D.J., Findlay, J.B., Eliopoulos, E., Maycox, P.R. et al. (1986) Posttranslational processing of concanavalin A precursors in jackbean cotyledons. *J. Cell Biol.* **102**, 1284–1297 <https://doi.org/10.1083/jcb.102.4.1284>
- 78 Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I. et al. (1993) Asparaginyl endopeptidase of jack bean seeds. Purification, characterization, and high utility in protein sequence analysis. *J. Biol. Chem.* **268**, 3525–3529 [https://doi.org/10.1016/S0021-9258\(18\)53726-1](https://doi.org/10.1016/S0021-9258(18)53726-1)
- 79 Min, W. and Jones, D.H. (1994) *In vitro* splicing of concanavalin A is catalyzed by asparaginyl endopeptidase. *Nat. Struct. Mol. Biol.* **1**, 502–504 <https://doi.org/10.1038/nsb0894-502>
- 80 Zauner, F.B., Dall, E., Regl, C., Grassi, L., Huber, C.G., Cabrele, C. et al. (2018) Crystal structure of plant legumain reveals a unique two-chain state with pH-dependent activity regulation. *Plant Cell* **30**, 686–699 <https://doi.org/10.1105/tpc.17.00963>
- 81 Dall, E., Zauner, F.B., Soh, W.T., Demir, F., Dahms, S.O., Cabrele, C. et al. (2020) Structural and functional studies of *Arabidopsis thaliana* legumain beta reveal isoform specific mechanisms of activation and substrate recognition. *J. Biol. Chem.* **295**, 13047–13064 <https://doi.org/10.1074/jbc.RA120.014478>
- 82 Kuroyanagi, M., Nishimura, M. and Hara-Nishimura, I. (2002) Activation of *Arabidopsis* vacuolar processing enzyme by self-catalytic removal of an auto-inhibitory domain of the C-terminal propeptide. *Plant Cell Physiol.* **43**, 143–151 <https://doi.org/10.1093/pcp/pcf035>
- 83 James, A.M., Haywood, J. and Mylne, J.S. (2018) Macrocyclization by asparaginyl endopeptidases. *New Phytol.* **2118**, 923–928 <https://doi.org/10.1111/nph.14511>
- 84 Elsässer, B., Zauner, F.B., Messner, J., Soh, W.T., Dall, E. and Brandstetter, H. (2017) Distinct roles of catalytic cysteine and histidine in the protease and ligase mechanisms of human legumain As revealed by DFT-based QM/MM simulations. *ACS Catal.* **7**, 5585–5593 <https://doi.org/10.1021/acscatal.7b01505>
- 85 Dall, E. and Brandstetter, H. (2016) Structure and function of legumain in health and disease. *Biochimie* **122**, 126–150 <https://doi.org/10.1016/j.biochi.2015.09.022>
- 86 Vernet, T., Tessier, D.C., Chatellier, J., Plouffe, C., Lee, T.S., Thomas, D.Y. et al. (1995) Structural and functional roles of asparagine 175 in the cysteine protease papain. *J. Biol. Chem.* **270**, 16645–16652 <https://doi.org/10.1074/jbc.270.28.16645>
- 87 Koehnke, J., Bent, A., Houssen, W.E., Zollman, D., Morawitz, F., Shirran, S. et al. (2012) The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nat. Struct. Mol. Biol.* **19**, 767–772 <https://doi.org/10.1038/nsmb.2340>
- 88 Conlan, B.F., Colgrave, M.L., Gillon, A.D., Guarino, R., Craik, D.J. and Anderson, M.A. (2012) Insights into processing and cyclization events associated with biosynthesis of the cyclic peptide kalata B1. *J. Biol. Chem.* **287**, 28037–28046 <https://doi.org/10.1074/jbc.M112.347823>
- 89 Yang, R., Wong, Y.H., Nguyen, G.K.T., Tam, J.P., Lescar, J. and Wu, B. (2017) Engineering a catalytically efficient recombinant protein ligase. *J. Am. Chem. Soc.* **139**, 5351–5358 <https://doi.org/10.1021/jacs.6b12637>
- 90 Jackson, M.A., Gilding, E.K., Shafee, T., Harris, K.S., Kaas, Q., Poon, S. et al. (2018) Molecular basis for the production of cyclic peptides by plant asparaginyl endopeptidases. *Nat. Commun.* **9**, 2411 <https://doi.org/10.1038/s41467-018-04669-9>
- 91 Dall, E. and Brandstetter, H. (2012) Activation of legumain involves proteolytic and conformational events, resulting in a context- and substrate-dependent activity profile. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **68**(Pt 1), 24–31 <https://doi.org/10.1107/S1744309111048020>
- 92 Tang, T.M.S., Cardella, D., Lander, A.J., Li, X.F., Escudero, J.S., Tsai, Y.H. et al. (2020) Use of an asparaginyl endopeptidase for chemoenzymatic peptide and protein labeling. *Chem. Sci.* **11**, 5881–5888 <https://doi.org/10.1039/D0SC02023K>
- 93 Cao, Y., Nguyen, G.K., Tam, J.P. and Liu, C.F. (2015) Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation. *Chem. Commun. (Camb)* **51**, 17289–17292 <https://doi.org/10.1039/C5CC07227A>
- 94 Bi, X., Yin, J., Nguyen Giang, K.T., Rao, C., Halim Nurashikin Bte, A., Hemu, X. et al. (2017) Enzymatic engineering of live bacterial cell surfaces using butelase 1. *Angew. Chem. Int. Ed. Engl.* **56**, 7822–7825 <https://doi.org/10.1002/anie.201703317>
- 95 Hemu, X., Qiu, Y., Nguyen, G.K.T. and Tam, J.P. (2016) Total synthesis of circular bacteriocins by butelase 1. *J. Am. Chem. Soc.* **138**, 6968–6971 <https://doi.org/10.1021/jacs.6b04310>
- 96 Hemu, X., Zhang, X. and Tam, J.P. (2019) Ligase-controlled cyclo-oligomerization of peptides. *Org. Lett.* **21**, 2029–2032 <https://doi.org/10.1021/acs.orglett.9b00151>
- 97 Rehm, F.B.H., Harmand, T.J., Yap, K., Durek, T., Craik, D.J. and Ploegh, H.L. (2019) Site-specific sequential protein labeling catalyzed by a single recombinant ligase. *J. Am. Chem. Soc.* **141**, 17388–17393 <https://doi.org/10.1021/jacs.9b09166>
- 98 Jackson, M.A., Nguyen, L.T.T., Gilding, E.K., Durek, T. and Craik, D.J. (2020) Make it or break it: plant AEPs on stage in biotechnology. *Biotechnol. Adv.* **45**, 107651 <https://doi.org/10.1016/j.biotechadv.2020.107651>
- 99 Nguyen, G.K., Qiu, Y., Cao, Y., Hemu, X., Liu, C.F. and Tam, J.P. (2016) Butelase-mediated cyclization and ligation of peptides and proteins. *Nat. Protoc.* **11**, 1977–1988 <https://doi.org/10.1038/nprot.2016.118>

- 100 Nguyen, G.K., Cao, Y., Wang, W., Liu, C.F. and Tam, J.P. (2015) Site-specific N-terminal labeling of peptides and proteins using butelase 1 and thiodipeptide. *Angew. Chem.Int. Ed. Engl.* **54**, 15694–15698 <https://doi.org/10.1002/anie.201506810>
- 101 Rehm, F.B.H., Tyler, T.J., Yap, K., Durek, T. and Craik, D.J. (2021) Improved asparaginyl-ligase-catalyzed transpeptidation via selective nucleophile quenching. *Angew. Chem. Int. Ed. Engl.* **60**, 4004 <https://doi.org/10.1002/anie.202013584>
- 102 Giordanetto, F. and Kihlberg, J. (2014) Macrocyclic drugs and clinical candidates: what can medicinal chemists learn from their properties? *J. Med. Chem.* **57**, 278–295 <https://doi.org/10.1021/jm400887j>
- 103 White, A.M. and Craik, D.J. (2016) Discovery and optimization of peptide macrocycles. *Expert Opin Drug Discov.* **11**, 1151–1163 <https://doi.org/10.1080/17460441.2016.1245720>
- 104 Zorzi, A., Deyle, K. and Heinis, C. (2017) Cyclic peptide therapeutics: past, present and future. *Curr. Opin. Chem. Biol.* **38**, 24–29 <https://doi.org/10.1016/j.cbpa.2017.02.006>
- 105 Gründemann, C., Stenberg, K.G. and Gruber, C.W. (2019) T20k: an immunomodulatory cyclotide on its way to the clinic. *Int. J. Pept. Res. Ther.* **25**, 9–13 <https://doi.org/10.1007/s10989-018-9701-1>