Biosynthesis of cyclotides

Thomas Shafee¹, Karen Harris¹, Marilyn Anderson¹

¹La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Vic 3086, Australia

1 ABSTRACT

Cyclotides are gene-encoded, ribosomally produced cyclic proteins. Their biosynthesis is a multistep process involving several enzymes for translation, disulphide isomerization, transport, cleavage and, finally, cyclization. Here, we describe this process, focusing on the cyclization step and the vacuolar processing enzyme (VPE) that performs it.

Cyclotide precursor proteins contain prodomains that direct organelle targeting and contain sequences necessary for cyclization in addition to the one or more domains that make up the final cyclic product. The VPEs are a family of cysteine proteases that perform cyclization by an adaptation of their more commonly described proteolysis mechanism. VPE-mediated cyclization is dependent on the presence of short recognition sequences (within and flanking the mature cyclotide domain) which have been characterized in vivo and in vitro for the prototypical cyclotide kalata B1. Finally, we compare cyclization by VPEs to other methods of protein cyclization found in nature or developed for biotechnology. Understanding cyclotide biosynthesis is important for both fundamental science, and for broader protein cyclization applications.

Key words: Cyclotide biosynthesis, vacuolar processing enzyme, transpeptidation, acyl-enzyme intermediate, cyclisation, ligation, cyclic proteins
INTRODUCTION

Cyclotides are small, disulphide-rich plant proteins with the unusual property of having a cyclic peptide backbone (Craik et al., 1999). They are highly thermally and proteolytically stable due to the lack of free peptide termini, the knotted topology of their disulphides and their compact size (around 30 amino acids) (Colgrave et al., 2004).

Cyclotides perform a variety of pathogen defence roles in the plants that produce them and have been described variously as insecticidal (Gruber, Cemazar, et al., 2007; Jennings et al., 2001; Mylne et al., 2010), nematocidal (Colgrave et al., 2008; Colgrave, Huang, et al., 2010; Simonsen et al., 2008), molluscicidal (Plan et al., 2008), bactericidal (Pränting et al., 2010), anti-viral (Wang et al., 2008), haemolytic (Henriques et al., 2011), and phytotoxic (Ovesen et al., 2011). Their activity marks them as interesting agricultural and pharmaceutical engineering targets (Craik et al., 2010, 2012; Dörnenburg, 2010), and their stability allows them to be used as scaffolds for displaying engineered peptides (Ji et al., 2013).

Investigations into the biosynthesis of cyclic peptides are important to both fundamental and applied science. Firstly, cyclotides are circularised by a modified protease mechanism that appears to have convergently evolved in multiple other cyclic protein biosynthesis pathways. Additionally, understanding cyclisation is necessary for recombinant cyclotide production and as a biotechnological tool for cyclising other proteins of interest. This may also provide the means to overcome the poor expression of correctly folded, mature cyclotides in transgenic plants, enhancing the agricultural value of this group of peptides.

In this chapter we review the current evidence on cyclotide biosynthesis, focussing on the processing events that occur, and the enzymes that facilitate them.
3 CYCLOTIDES ARE GENE PRODUCTS

Plant cyclotides are gene-encoded, unlike many small cyclic peptides from bacteria and fungi that are synthesised by multienzyme complexes (F. Conlan et al., 2011). The cyclotides are produced by cleavage and cyclisation of cyclotide domains embedded in precursor proteins which have been described for the Rubiaceae (Jennings et al., 2001), Violaceae (Dutton et al., 2004), Cucurbitaceae (Mylne et al., 2012), Solanaceae (Poth et al., 2012) and Fabaceae (Poth, Colgrave, Philip, et al., 2011).

The first cyclotide transcripts were obtained from *Oldenlandia affinis* (Rubiaceae) which produces the prototypical kalata B1 (kB1) and B2 (kB2) cyclotides of the Mobius subclass (Jennings et al., 2001). Transcripts encoding cycloviolacin cyclotides from both the Mobius and Bracelet subclasses were then obtained from *Viola odorata* (Violaceae) (Dutton et al., 2004). Genes from the Rubiaceae and the Violaceae encode similar precursors comprised of an endoplasmic reticulum (ER) signal sequence and a propeptide, followed by up to three cyclotide domains each flanked by conserved N-terminal and C-terminal prodomains (Figure 1). The cyclotide domains in the Mobius group shared greater than 79% sequence identity at the amino acid level irrespective of whether they were from the Rubiaceae or Violaceae. However when the Mobius and Bracelet mature cyclotide domains were compared they shared only 40% identity (Dutton et al., 2004) suggesting that these subclasses were inherited from an earlier ancestor in the Eudicot lineage and have evolved separately. Conservation of the 16 amino acid N-terminal repeat (Ntr) sequences in the *O. affinis* precursors led to the original hypothesis that that the Ntr had a role in cyclotide processing. However there is a lack of sequence conservation between the Ntrs from *O. affinis* and *V. odorata* precursors and thus a role in processing was considered unlikely. The observation that the Ntrs from both species form amphipathic helices led to the new hypothesis that they may function in protein folding or subcellular targeting (Dutton et al., 2004).

Relatively little is known about the regulation of cyclotide gene expression. Some cyclotides are expressed constitutively, whilst others vary during the year (Trabi, Svangård, et al., 2004), presumably induced by seasonal changes in biotic and abiotic stress (Basse, 2005; Seydel et al., 2007). Similarly cyclotides appear to be localised to different plant tissues (Mylne et al., 2010; Trabi & Craik, 2004).
Figure 1 | Schematic of cyclotide precursors

**Violaceae and Rubiaceae**: Cyclotide precursors are composed of an endoplasmic reticulum (ER) signal peptide, an N-terminal propeptide and a region composed of an N-terminal repeat, a mature cyclotide domain and a C-terminal repeat. This region can be repeated up to three times and is followed by an additional short sequence at the C-terminus.

**Solanaceae**: Cyclotide transcripts from *Petunia hybrida* encode a precursor with an ER sequence and one N-terminal propeptide of 17-19 amino acids followed by a mature cyclotide domain and a C-terminal propeptide of 6 amino acids.

**Cucurbitaceae**: Cyclotide transcripts from *Momordica cochinchinensis* in the Cucurbitaceae family encode precursors with an ER signal sequence followed by an N-terminal propeptide of 2-9 amino acids and a region composed of an N-terminal repeat sequence of 16 amino acids and a mature cyclotide domain. This region is repeated up to 7 times and is separated by 8 amino acids from a terminal acyclotide domain which is a typical non-cyclic knottin.

**Fabaceae**: Cyclotide transcripts from the butterfly pea, *Clitoria ternatea*, from the Fabaceae family encode a precursor with an ER signal peptide followed immediately by a mature cyclotide domain that is flanked at the C-terminus by a linking peptide and the albumin-1a chain. The cyclotide domain replaces the PA1b subunit that is present in typical albumin-1 genes.

The cyclotide precursors from the Solanaceae, the Cucubitaceae and the Fabaceae all have substantial differences to those from the Rubiacea and the Violaceae (Figure 1). Precursors from the Solanaceae have only one mature cyclotide domain and do not have an N-terminal propeptide. In contrast, precursors from *Momordica cochinchinensis* from the Cucubitaceae have up to seven cyclotide domains as well as one acyclotide domain at the C-terminus and are potent trypsin inhibitors (Mylne et al., 2012). Finally, the cyclotides from the Fabaceae are embedded within an albumin seed storage protein (Poth, Colgrave, Philip, et al., 2011). Thus, there has been more than one independent origin for the cyclotide genes in the Eudicots or they were present in a very early ancestor. If the latter is true they will be found in many more
plant families. This also has implications for the level of conservation of the maturation process, which will be discussed in section 4.4.
4 THE CYCLOTIDE PROCESSING PATHWAY

All the cyclotide precursors from the five Eudicot families illustrated in Figure 1 have typical ER targeting sequences and thus they enter the secretory pathway where folding and disulphide bond formation, aided by protein disulphide isomerase occurs (Gruber, Čemažar, et al., 2007) (Figure 2). After folding the precursor leaves the ER and transits the Golgi apparatus. In the absence of specific targeting signals, proteins are then secreted via the default pathway. However, the cyclotide precursors are targeted to the prevacuolar compartment and the vacuole for proteolytic processing and cyclisation (Figure 2).

Subcellular targeting to the vacuole was demonstrated using a series of green fluorescent protein (GFP) constructs and transient expression in Nicotiana benthamiana (Conlan et al., 2011). The full length cyclotide precursor from O. affinis, Oak1, directed GFP to the vacuole (Figure 3) demonstrating that excision and cyclisation of the kB1 cyclotide occurred in the vacuole and that cyclotides were stored in this compartment. Further GFP constructs were used to identify the vacuolar targeting signals within the precursor. The N-terminal propeptide and the N-terminal repeat sequence from Oak1 were both sufficient to target the precursor to the vacuole. The C-terminal propeptide, with the C-terminal repeat sequence which is essential for cyclisation, was not a targeting signal. Targeting of the precursor to the acidic vacuolar compartment supported the early hypothesis that a vacuolar processing enzyme (VPE) had a pivotal role in excision and cyclisation of the cyclotide domains from the cyclotide precursors (Gillon et al., 2008).
Figure 2 | Cyclotide processing pathway

The cyclotide precursor is comprised of several parts: Firstly a hydrophobic signal peptide (thought to be α-helical (Nakamura et al., 1993)) targets the precursor to the ER. The N-terminal propeptide and the helical N-terminal repeat contain vacuolar targeting sequences (Conlan et al., 2011) (PDB:1WN8). The cyclotide domain contains the cysteine knot motif (Rosengren et al., 2003) (PDB:1NB1). A final C-terminal propeptide of unknown structure is necessary for cyclisation. The cyclotide domain folds in the ER. Disulphide bond formation is mediated by the oxidising environment and protein disulphide isomerase. The first signal peptide is cleaved and the remaining protein is secreted to the vacuole, whereupon the N-terminal propeptide is also cleaved by an unknown protease. Finally, the protease, vacuolar processing enzyme performs backbone cyclisation by transpeptidation to yield the final cyclic product.
Figure 3 | Vacuolar targeting by the signal peptide and Oak1 precursor

Transient expression of GFP linked to the Oak1 cyclotide precursor from Oldenlandia affinis in leaf epidermal cells from Nicotiana benthamiana. Left panel: GFP fluorescence from GFP-Oak1 in the ER and Golgi on the periphery of the cell and GFP in the vacuole (v) after proteolytic processing and production of the kB1 cyclotide. Middle panel: Fluorescence from the membrane dye FM 5-95 highlighting the plasma membrane (pm), the tonoplast (t, vacuolar membrane) and other small membranous components. Right panel: Diagram illustrating the relevant subcellular compartments of one of the cells shown in the fluorescent micrographs. (Modified from (Conlan et al., 2011)).
SEQUENCE REQUIREMENTS FOR CYCLISATION

The maturation of cyclotides from their precursor proteins to the final, mature products requires several processing steps (Figure 2). The penultimate processing event removes the N-terminal region, producing a linear precursor containing the C-terminal propeptide. In the final processing step, the C-terminal propeptide is removed and the free N- and C-termini are ligated to yield the mature, cyclic peptide. This is thought to be achieved enzymatically by intramolecular transpeptidation; an event that requires the N- and C-termini to be brought into close proximity (Figure 4). The active site must therefore accommodate residues from both the N- and C-termini of the nascent cyclotide, suggesting that the propensity of a given substrate to be cyclised may be driven by the amino acid composition at both termini. In this section, the sequence requirements around these processing sites will be discussed, particularly in relation to the prototypic cyclotide kB1, drawing on evidence from precursor sequence conservation, transgenic cyclotide production, and substrate preference of putative cyclases.

**Figure 4 | Enzyme binding sites**
The cyclising enzyme binds firstly to the C-terminus of the cyclotide precursor and subsequently to the N-terminus using the same binding site. The binding sites are labelled S1 to S3 before the cleavage site, and S1’ to S3’ after the cleavage site (indicated by a black triangle). The peptide residues that bind to these sites are similarly labelled P1 to P3 before and P1’ to P3’ after the cleavage site. Since the cyclotide N-terminus binds to the enzyme as a second substrate after cleavage, its residues are labelled P1” to P3”.
5.1 SEQUENCE CONSERVATION

Sequence analysis of the N- and C-terminal processing sites of 36 cyclotide precursors from plants of the Rubiaceae and Violaceae families led to the identification of a conserved C-terminal tripeptide motif at each cleavage site (Gillon et al., 2008). Extending this to include a total of 113 cyclotide sequences supports this early conclusion (Figure 5). The first position of this motif prefers a residue with a small side chain (typically Gly or Ser); the second position is represented by a conserved hydrophobic residue (typically Leu or Ile); and the third site is most often occupied by Pro, although this is not well maintained at the C-terminal processing site. Gly₁ and Leu₃₁ are particularly highly favoured and are present in >90% of sequences analysed here indicating they are under strong selective pressure. However, whilst Gly₁ itself is highly conserved, some cyclotide sequences include an extra residue on its C-terminal side suggesting that sequence conservation of subsequent residues in the N-terminal tripeptide motif may not be as crucial for cyclotide biosynthesis. Strong sequence conservation of Asn/Asp at the C-terminal P1 site has also been observed and of the 282 cyclotide sequences currently available in CyBase, 78% contain Asn₂₉ and 15% contain Asp at this position (Mulvenna et al., 2006). This, in combination with other lines of evidence, led to the prediction that VPEs, an Asx-specific group of vacuolar proteases introduced in section 3, were responsible for cleavage of the C-terminal propeptide and head-tail ligation (discussed in detail in section 5).

![Figure 5 | Cyclotide terminal processing sequences](https://example.com/figure5.png)

The N-terminal sequence (P₁'' to P₃'') and C-terminal processing sequence (P₃ to P₁ and P₁' to P₃') based on an alignment of 113 cyclotide precursor sequences from the Rubiaceae and Violaceae. The consensus sequence is shown below, where ‘φ’ is Leu/Ile, ‘+’ is Arg/Lys and ‘X’ is any residue. The bar-chart shows the percentage conservation at each residue position. Note the similarity of GφP to SLX that have to bind to the same active site pockets of the cyclising enzyme.

5.2 MUTAGENESIS IN TRANSGENIC PLANTS

Analysis of the sequence conservation around the N- and C-terminal processing sites highlighted residues that were likely to be important for cyclotide maturation and guided in
planta mutagenesis of precursors of the prototypic cyclotide kB1 (encoded by the Oak1 gene) (summarized in Figure 6) (Gillon et al., 2008). In this cyclotide, the tripeptide motif is represented by Gly-Leu-Pro at both the C- and N-terminal processing sites. Both Nicotiana tabacum and Arabidopsis thaliana transformed with the wt Oak1 gene could stably express correctly processed cyclic kB1, although a number of linear processing variants were also evident. Following mutation of the conserved Asn29 to either Asp or Ala, no cyclic product was detected. To further investigate sequence requirements, transient expressions of additional variants were carried out in N. benthamiana. This series of experiments showed that truncation of the C-terminal propeptide from GLPSLAA to GLP did not affect cyclisation, however no cyclisation was evident when Oak1 was terminated after Asn29. Similarly, no cyclisation was observed when Gly1 or Leu2 within the N-terminal Gly-Leu-Pro motif were substituted with Ala.

![Figure 6](https://doi.org/10.1016/bs.abr.2015.08.005)

**Figure 6 | Processing efficiency of Oak1 gene mutants**

The Oak1 gene was mutated to encode amino acid substitutions and/or truncations in the N-terminal propeptide, mature cyclotide domain and the C-terminal propeptide and was then introduced into tobacco plants. The presence or absence of cyclic product was determined by mass spectrometry and is indicated by ✓ or ✗ respectively. * denotes C-terminal truncation after the indicated residue. Brackets denote combined truncation and mutation of the C-terminal propeptide. Adapted from references (Conlan et al., 2012; Gillon et al., 2008)

A further suite of Oak1 mutants were analysed in a more recent report (summarized in Figure 6) (Conlan et al., 2012). Here, transient expression of O. affinis kB1 incorporating mutations
around the C- and N-terminal processing sites was carried out in *N. benthamiana*. At the N-terminus, neither of the two residues targeted (Lys-1 and Leu-2) were required for cyclisation. However, substitution of Leu-2 did result in the appearance of additional residues at the N-terminus suggesting that this mutation hinders clean cleavage of the Ntr from the cyclotide domain in *N. benthamiana*. The lack of apparent sequence specificity around this cleavage site suggests that the residue immediately upstream of the cyclotide domain may not be important, at least in this system. Cleavage may instead be driven by surface-exposure and accessibility leading to non-specific targeting by various proteases, or by a single enzyme with limited sequence specificity.

At the C-terminal cleavage site of kB1, a conservative Asn29Gln substitution abolished the production of cyclic peptide in *N. benthamiana*, further supporting a role for VPEs as the mediator of this processing event. At the P2 position, substitution of Arg28 with Ala or Lys did not preclude production of cyclic protein suggesting this position is not important for cyclisation. However, overall expression of these variants was reduced compared to native Oak1, perhaps as a result of misfolded precursor being unable to exit the ER, demonstrating that sequence variation around these processing sites is also limited by factors other than specificity of the processing enzymes. At the P1’ position, substitution of the native Gly with the correspondingly small residues Ala or Ser was well tolerated, whilst mutation to the larger, bulky residue, Phe, resulted in a lack of correctly processed cyclic product. This suggested that the conservation of small residues at this position was attributable to the constraints imposed by enzyme specificity. The presence of Pro at the P3’ position was deemed non-essential since cyclic product was still detected following substitution with Ala. Consistent with this, a series of truncations of the C-terminal propeptide identified the dipeptide Gly-Leu as the minimal C-terminal sequence required for cyclisation. Interestingly, whilst no cyclic product was detected in the P2’ Leu31Ala mutant, an unusual post-translational modification was observed: hydroxylation of the neighbouring proline. When a more conservative substitution of Leu31Ile was introduced, both cyclic kB1 and a small amount of the hydroxylated product were produced. Whilst not directly tested, it is possible that hydroxylation of this Pro could also directly affect cyclisation and a requirement to limit this event could be contributing to sequence conservation in this region.
A limitation of studying processing of cyclotide precursors in transgenic plants is that they are necessarily carried out species that do not natively produce cyclotides and so are unlikely to provide optimal conditions for processing cyclic peptides. This is exemplified by the relatively small proportion of the total Oak1 gene product that is correctly processed to the mature, cyclic peptide in transgenic tobacco (Conlan et al., 2012; Gillon et al., 2008). Small changes in the level of cyclic product may therefore render the product below the detection limit of the assay and drawing absolute conclusions about essential residues for cyclisation should be approached with caution. It remains to be determined if any of the aberrantly processed forms are relevant in native cyclotide producing plants where the high level of cyclotide expression and lack of detection of any linear forms imply a highly efficient maturation process. However, native cyclotide producers are currently poor models for probing cyclisation requirements since they are not readily amenable to transformation and production of transgenic cyclotides is obscured by native cyclotides. Co-expression of cyclotides and VPEs coupled with advances in cyclotide detection could facilitate improvements in this area. Additionally, the recent discovery of cyclotides in the Solanaceae has revealed subtle sequence variation in the regions flanking the mature cyclotide, perhaps reflecting the specificity of processing enzymes within this plant family (Poth et al., 2012). Tailoring cyclotide precursor sequences within model plant systems to reflect the substrate specificity of endogenous VPEs may therefore improve the utility of this technique.

5.3 ATYPICAL CYCLOTIDE MATURATION

Cyclotides have thus far been discovered in the Rubiaceae, Violaceae, Cucurbitaceae, Fabaceae and Solanaceae plant families and are predicted to be present in many more (Craik et al., 2013; Gruber et al., 2008). Most known sequences belong to the Rubiaceae and Violaceae and are reported to be similarly processed from precursor proteins containing one, two or three cyclotide domains (Dutton et al., 2004). However, the biosynthesis of at least some cyclotides from other plant families does not follow this established maturation process and mature cyclotides are produced from precursors with varying gene structures (see Figure 1). Furthermore, the Momordica cochinchinensis trypsin inhibitor II (MCoTI-II) precursor protein contains the less commonly observed Asn at the N-terminal P1 site, implicating a VPE in processing of both the C- and N-terminal propeptides (Mylne et al., 2012). In such cases it
will be of interest to determine if a single VPE has evolved to act as both a protease and cyclase, or if different allelic paralogs carry out the two different functions.

Cyclotides recently discovered in *Clitoria ternatea* (butterfly pea) of the Fabaceae family are also produced from genes embedded within an albumin precursor protein (Poth, Colgrave, Philip, et al., 2011). The *C. ternatea* cyclotides do not contain an N-terminal prodomain or an Ntr region as with typically reported cyclotide precursors; rather, their ER signal sequence is immediately followed by the cyclotide domain. This suggests that, as well as gene structure, the processing events leading to mature cyclotide can vary across plant families. Interestingly, the sequences of cyclotides from *C. ternatea* differ in that the consensus C-terminal processing motif is His-Val-Ile (compared to Gly-Leu-Pro in kB1) (Figure 7). In particular, His is not observed in the P1’ position in the C-terminal propeptide of any other cyclotide precursors and reflects a divergence from the small residues that are typically accommodated here (Poth, Colgrave, Lyons, et al., 2011). Similarly, Glu is found at this site in the recently reported cyclotides from the Solanaceae (Poth et al., 2012). This raises the possibility that the specificity of the enzyme responsible for cyclisation may also differ between plant families, providing support for a more tailored approach to cyclotide production in transgenic plants. Alternatively, the sequence requirements for enzymatic cyclisation may be less stringent than initially suggested by the overall sequence conservation and transgenic data.

**Figure 7 | C. ternatea cyclotide terminal processing sequences**
The cyclotide precursor from *C. ternatea* differs from the standard form. The cyclotide domain is located at the beginning of the construct, before an albumin domain that is also processed into a mature protein. The N-terminal sequence (P1’’ to P3’’) and C-terminal processing sequence (P3 to P1 and P1’ to P3’’) based on an alignment of 11 *C. ternatea* cyclotide precursor sequences. The consensus sequence is shown below, where ‘+’ is Arg/Lys. The bar-chart shows the percentage conservation at each residue position.

Acyclic, linear peptides that share high sequence homology with the cyclotides – including the characteristic cystine knot motif - have also been identified in an increasing number of plant families (Gerlach et al., 2010; Ireland et al., 2006; Giang Kien Truc Nguyen et al., 2011, 2012, 2013; Poth et al., 2012). These peptides, termed “acyclotides” (Poth et al., 2012), can share structural similarity with their cyclic counterparts (Ireland et al., 2006) and retain some native
cyclotide bioactivities (Gerlach et al., 2010; Giang Kien Truc Nguyen et al., 2012, 2013). A common feature of the acyclotides is the absence of Asx at the C-terminus; either as a result of a truncation prior to the terminal Asx or the introduction of an upstream stop codon (Giang Kien Truc Nguyen et al., 2013). This observation further supports the importance of Asx-specific processing in the enzymatic maturation of backbone-cyclised peptides.

5.4 Substrate sequence requirements for VPE-mediated transpeptidation

Recently it has been reported that a VPE, named butelase 1, isolated from C. ternatea is able to cyclise a number of peptides (including kB1) containing an appropriate recognition sequence (Giang Nguyen et al., 2014). The catalytic mechanism of VPEs and their role in cyclisation are discussed in more detail in section 5, however it is worth noting here that this cyclase required the addition of only a small recognition sequence to unrelated, acyclic peptides to render these peptides cyclisable. For example, thanatin, an anti-microbial peptide isolated from insects (Fehlbaum et al., 1996) only required the addition of a C-terminal Asn and the dipeptide His-Val (representing the C-terminal prodomain of the native substrate) for efficient butelase 1-mediated cyclisation. Therefore the addition of three residues (Asn-His-Val) can provide sufficient affinity to the enzyme-substrate interaction for cyclisation to occur, provided the N-terminus can also be accommodated by the active site. This is consistent with the in planta studies showing that only two residues C-terminal to the P1 Asn were required for cyclisation of the prototypic cyclotide kB1.

At the N-terminus of the cyclotides identified in C. tenatea the established tripeptide motif is retained and most commonly exemplified by Gly-Ile-Pro, although His has also been observed in place of Gly (Figure 7) (Poth, Colgrave, Lyons, et al., 2011). Accordingly, the N-terminus of most of the synthetic peptides cyclised by butelase 1 contained a similar sequence (Gly-Val or Gly-Leu), either natively or engineered. However, when the N-terminal specificity of butelase 1 was assessed in the context of model peptides, most amino acids could be accommodated in place of Gly with similar efficiency (Figure 8). This apparent N-terminal flexibility is incongruent with predictions based on sequence conservation and data from transgenic plants. For example, an N-terminal Phe was readily accepted by butelase 1 whereas, in transgenic plants, the same substitution prevented cyclisation of kB1 precursors (Conlan et
This may be due to a greater stringency of in planta over in vitro experiments or indicate a real difference in substrate specificity of VPEs from different sources.

The reported P2” preference of butelase 1 for the hydrophobic amino acids Leu, Ile and Val is more consistent with previous reports, as determined in the context of model peptides (Figure 8b). For example, the presence of Ala at this position resulted in a dramatic reduction in the ligation efficiency of model peptides by butelase 1, as well as precluding the production of cyclic kB1 from the mutagenized Oak1 gene in transgenic plants (Gillon et al., 2008; Giang Nguyen et al., 2014). However, hydrophobicity at this site was not an absolute requirement for butelase 1 activity since cyclisation of thanitin, which contains Ser at this position, was still achieved; albeit at a slower rate.

![Figure 8](image-url)

**Figure 8 | Substrate specificity of butelase 1 for the incoming N-terminus during transpeptidation.**

(a) Butelase 1 cleaves after a P1 Asn (back triangle) in an 8-residue model peptide and ligates the newly exposed C-terminus to the N-terminus of a second peptide. In the second peptide, either the P1” or P2” site is substituted with each of the 20 amino acids and the effect on yield examined. (b) Yields of ligated product from model peptides varied at the P1” and P2” positions. Adapted from reference (Giang Nguyen et al., 2014).

Whilst it remains to be determined if the broad specificity reported for butelase 1 is applicable to a full-length cyclotide, and to cyclisation in planta, the apparent lack of preference for an N-terminal Gly raises the possibility that enzyme specificity may not be driving the high level of conservation at this position. This is unlikely to be driven by functional constraints since Ala and Lys scanning of the prototypic cyclotide kB1 demonstrated that substitution at position 1
did not impede its insecticidal or nematocidal activity (Huang et al., 2010; Simonsen et al., 2008). Alternatively, folding efficiency may be impaired following substitution at this position. Indeed, the yield of chemically synthesised Gly₁Ala was only 50% that of wild type providing support for this hypothesis (Simonsen et al., 2008).

5.5 SUMMARY

Sequence conservation around the C- and N-terminal processing sites provided early insight into the requirements for cyclisation and this was supported by targeted mutagenesis in transgenic plants. However, these motifs may not be applicable to all families of cyclotide producing plants. The recent isolation of butelase 1 demonstrated that cyclisation of unrelated peptides could be achieved following the addition of a small recognition motif and that sequence requirements outside of this appear minimal. It is unknown if the substrate promiscuity displayed by butelase 1 is representative of all VPEs that can achieve cyclisation, or if it uniquely accommodates the sequences expressed in C. ternatea. Isolation of VPEs from other plant species, as well as combining cyclotide and VPE expression to improve the utility of transgenic studies will aid in answering this question and further elucidating the mechanisms of cyclotide maturation.
5 CIRCULARIZATION OF CYCLOTIDES IS MEDIATED BY A VACUOLAR PROCESSING ENZYME (VPE)

VPEs (also known as asparaginyl endopeptidases [AEPs] or legumains) are proposed to mediate peptide cyclisation by a promiscuous transpeptidase activity – a modification of their native hydrolysis mechanism. In this section, we will outline the evidence that led to the identification of VPEs as key players in cyclotide maturation and discuss our current understanding of the mechanisms behind cyclization.

5.1 VPE IS LIKELY THE CYCLOTIDE CYCLASE

Initial attempts to identify potential enzymes that might be responsible for cyclotide cyclisation identified VPEs as likely candidates. VPEs exclusively cleave C-terminal to Asn/Asp (I. Hara-Nishimura et al., 1991; Nagako Hiraiwa et al., 1999) and this substrate specificity matches the high sequence conservation of Asn/Asp at the cyclotide C-terminus (Gillon et al., 2008). Additionally, their vacuolar localisation is consistent with the reported production and storage of cyclotides within this compartment (Conlan et al., 2011). Strengthening this early conclusion was the report that the lectin, concanavalin A (Con A), from Canavalia ensiformis (jackbean) is matured by a VPE (Sheldon et al., 1996). Con A maturation involves VPE-mediated proteolysis and transpeptidation (though not backbone cyclisation), demonstrating the capability of this group of enzymes (Figure 9). It was speculated that the ligation function of VPE was driven by the close proximity of the N- and C-termini of the intermediate Con A processing products to preference transpeptidation (Sheldon et al., 1996).
Evidence *in planta* that cyclotide transpeptidation is mediated by VPEs came from studies where suppression of VPE activity resulted in a decreased amount of cyclic product in transgenic plants (Mylne et al., 2012; Saska et al., 2007). A limitation of this approach (as with the *in planta* mutagenesis studies discussed in section 4) is that the plant species used in these studies do not natively produce cyclotides and consequently the amount of cyclic product produced is relatively low, making it difficult to draw absolute conclusions. However, the finding that butelase 1 from *C. ternatea* could cyclise the *O. affinis* cyclotide kB1 provided the first direct *in vitro* evidence for the role of VPEs in cyclotide maturation.

### 5.2 Cyclisation by a Modified Covalent Proteolysis Mechanism

The mechanisms and sequence requirements underpinning the transpeptidation ability of VPEs remain poorly understood. VPEs are cysteine proteases and as such contain a catalytic triad in which an acidic or polar residue aligns a histidine to deprotonate a cysteine, increasing its nucleophilicity. Catalysis proceeds by nucleophilic attack by the activated cysteine on the carbonyl of the peptide substrate to form a covalent acyl-enzyme intermediate (Figure 10a). The cleaved substrate C-terminus is released and a second nucleophilic attack occurs to resolve the intermediate and release the N-terminal peptide. In the more commonly observed proteolysis, this second nucleophile is a water molecule which is polarised by the histidine to hydrolyse the intermediate causing the final products to be the two peptide fragments (Figure...
However in cyclisation, it is predicted that the N-terminus of the covalently bound peptide circles around and binds in the P’ pockets. The free amine of the peptide then acts as the second nucleophile to resolve the acyl-enzyme intermediate, yielding the mature cyclotide and linear C-terminal propeptide as the final products (Figure 10c).

The ratio of hydrolysis to transpeptidation is dependent on the relative strengths and local concentrations of water versus peptide amine to act as the second nucleophile. Transpeptidation relies on the N-terminal amine outcompeting the vastly more abundant...
water to resolve the acyl-enzyme intermediate. It has been hypothesised that VPE achieves a high local concentration of the cyclotide N-terminus by binding a conserved tripeptide motif (represented by GLP in the prototypic cyclotide kB1) at the N-terminus of the peptide (Figure 11) (Gillon et al., 2008) and that this is aided by the structural constraints imposed by the disulphide bonded structure of cyclotides. In this scenario, the specificity of the enzyme P' pockets for this tripeptide motif not only aids the binding of the initial substrate (Figure 11a), but also subsequent binding of the N-terminus (Figure 11b,c) for cyclisation (Figure 11d). However, the finding that butelase 1 is able to accommodate a broad range of residues at the first position of this motif with little impact on yield of ligated product suggests that this N-terminal motif may not be as critical for cyclisation as first thought (Giang Nguyen et al., 2014).

![Diagram](image.png)

**Figure 11 | Proposed binding specificity for cyclisation by VPE**

(a) VPE specifically preferentially binds the substrate sequence TRNGLP in kB1. (b) Upon formation of the acyl-enzyme intermediate, the cleaved C-terminal peptide is released to expose the GLP binding pocket. (c) If the peptide N-terminus also contains the GLP motif, it binds to the exposed active site pocket. (d) The N-terminal amine outcompetes water as the nucleophile to resolve the acyl-enzyme intermediate into a cyclic product.

Another factor that was previously thought to be important in pushing the equilibrium towards transpeptidation was the proximity of the N- and C-termini of the cyclotide precursor, maintained by the compact, disulphide-bonded structure. However butelase 1 does not require these structural constraints for efficient cyclisation since a reduced and alkylated linear kB1 precursor was cyclised more readily than disulphide bonded linear kB1 (Giang Nguyen et al., 2014). Furthermore, unrelated peptides derived from conotoxin, thanatin and histatin-3 that are not naturally backbone-cyclised were also cyclised following addition of compatible C- and N-termini.

There is evidence that as well as favouring transpeptidation, VPE also disfavours hydrolysis. Indeed, butelase 1 appears to have a low reactivity from water in its active site since, whilst still able to occur, hydrolysis of substrates lacking a free amine at the N-terminus is slow,
indicating a preference for cyclisation over hydrolysis (Giang Nguyen et al., 2014). This relatively low hydrolysis rate indicates that water is either poorly polarised by the active site or hydrophobically excluded from it.
6 VACUOLAR PROCESSING ENZYMES (VPEs)

The preceding sections have detailed the evidence that VPEs can mature cyclotides to their native, backbone-cyclised product. Although VPEs are widely expressed in plants and animals, only three members of this class have been directly demonstrated to be able to ligate peptide bonds and, of these, only butelase 1 performs as a cyclase (Dall et al., 2015; Giang Nguyen et al., 2014; Sheldon et al., 1996). Whilst this number is likely to expand, given the circumstantial evidence of involvement of other VPEs in peptide cyclisation (Mylne et al., 2012; Saska et al., 2007), the more commonly reported role is targeted proteolysis to mature precursor proteins and activate cellular pathways. It is unknown whether this disparity reflects subtle sequence differences that confer a preference for ligation or proteolysis or if dual capabilities are characteristic of all VPEs and dependent on substrate composition or conditions in vivo. In this context, this section will review the biosynthesis, structure and function of this important group of enzymes.

6.1 ISOLATION AND SUBSTRATE SPECIFICITY

The first evidence of VPE activity in plants was detected in cotyledons of *Vicia sativa* (common vetch) seedlings where cleavage C-terminal to Asn residues in the insulin A and B chains was observed (Shutov et al., 1982). VPEs have since been isolated from many plants such as *Ricinus communis* (castor bean) (I. Hara-Nishimura et al., 1991), *Glycine max* (soybean) (Scott et al., 1992), and *A. thaliana* (T. Kinoshita et al., 1995) as well as the protozoa *Schistosoma mansoni*, a parasitic blood fluke (Dalton et al., 1990; Klinkert et al., 1989). Mammalian homologues (legumains), have also been identified in a number of organisms such as pigs (JM Chen et al., 1997), mice (J. M. Chen, Dando, et al., 1998) and humans (Dalton et al., 1990; Klinkert et al., 1989). The proteinase responsible for this Asn-specific cleavage was later purified from maturing seeds of castor bean (I. Hara-Nishimura et al., 1991) and its vacuolar location demonstrated by immunocytochemical analysis (I. Hara-Nishimura et al., 1993). However, cell wall localisation, together with a role in the maturation of cell wall proteins, has also been reported demonstrating that VPEs are not exclusively vacuolar (Linnestad et al., 1998). Similarly, mammalian legumains have been reported primarily in the lysosomes but also in the nucleus (Haugen et al., 2013; Patel et al., 2009) which has a neutral pH, suggesting complex regulation and involvement in diverse cellular processes.
Apart from the characteristic Asn and, to a lesser extent, Asp specificity, there is little consensus in the substrate specificity of VPEs from different sources. For example, soybean VPE had an absolute requirement for a P1 Asn (Jung et al., 1998), whilst other VPEs extend this specificity to Asp (Becker et al., 1995; Nagako Hiraiwa et al., 1999; Giang Nguyen et al., 2014). Furthermore, substitution of the P2’ Leu with Ala in a native soybean VPE substrate appeared to be well tolerated (although it should be noted that no enzyme kinetics values are reported here) whereas the same substitution in the cyclotide precursor kB1 precluded its cyclisation in transgenic plants (discussed in section 4.3) (Conlan et al., 2012; Gillon et al., 2008). At the P3 position, kidney bean (Phaseolus vulgaris) VPE favored an amino acid with a bulky side-chain (Val, Phe, Tyr) (Rotari et al., 2001) and, similarly, sugarcane (Sacharum officinarum) VPE preferred Val at this site (Santos-Silva et al., 2012). However, again this is not applicable to all VPEs since mothbean VPE and a mammalian VPE from pig kidney did not exhibit the same specificity, (Dando et al., 1999; Kembhavi et al., 1993; Rotari et al., 2001). Additionally, butelase 1 could accommodate residues with different properties at this site, although processing efficiency of variants were not directly compared. A comparative study of the specificity of human and Schistosome legumain also revealed a lack of consensus in the optimal substrate sequences at the P2 and P3 positions (Mathieu et al., 2002). The precise substrate specificity VPEs from different sources is therefore variable, perhaps reflecting the multifunctional nature of VPEs and arguing for co-evolution of substrate and enzyme, as has been reported for other substrate/enzyme combinations (Kolli et al., 2006). Interestingly, the finding that VPE-mediated peptide circularization occurs in distantly related plant lineages suggests that cyclisation ability may have independently evolved multiple times, perhaps reflecting an enzyme that is particularly suited to being recruited to this function (Mylne et al., 2012).

6.2 BIOSYNTHESIS AND EXPRESSION OF PLANT VPEs

Plant VPEs have been isolated from both the maturing seeds and vegetative organs and share high sequence identity with each other and with mammalian homologues. They are synthesized as inactive precursors that have an ER signal peptide as well as short N-terminal and longer C-terminal prodomains that flank the mature endoproteinase domain (Figure 12). They also have a conserved tetra-peptide at the C-terminus that is essential for targeting to the vacuole (Kuroyanagi et al., 2002). The VPE precursors are trafficked through the secretory pathway to either the vacuole (Fischer et al., 2000; Nagako Hiraiwa et al., 1999; Schlereth et al., 2001) or the cell wall (Linnestad et al., 1998). Post-translational processing and activation of the enzyme occurs immediately after the precursor has entered an acidic environment such as the vacuole or the cell wall (pH 5-6). This occurs by sequential self-processing C-terminal to
selected Asn residues that yields the mature active enzyme (N. Hiraiwa et al., 1997; Nagako Hiraiwa et al., 1999; Müntz et al., 2002). Inactivity of the VPE precursor enables it to be packaged together with its substrates (including cyclotide precursors) into transport vesicles or ER bodies for transport to and activation in the vacuoles (Ikuko Hara-Nishimura et al., 1998; Jung et al., 1998). Maintaining inactivity of the proteinase during sub-cellular trafficking is attributed to the C-terminal prodomain that acts as an auto-inhibitory domain by masking the active site and regulating the final processing step in the activation cascade (Dall et al., 2013; Kuroyanagi et al., 2002). Interestingly, recent evidence has indicated that non-canonical self-processing of human legumain at a less acidic pH can impart a carboxypeptidase function, suggesting the complexity of regulation of this group of proteases is only beginning to be understood (Dall et al., 2013).

![Figure 12](https://doi.org/10.1016/bs.abr.2015.08.005)

**Figure 12 | Maturation and activation of VPEs**

A schematic representation of a VPE precursor. The signal peptide is removed post-translationally in the ER. Once it reaches the acidic environment of the vacuole the C-terminal prodomain is removed and the enzyme becomes active. Subsequent cleavage of the N-terminal prodomain produces the fully mature active form.

Typically, different plant species contain multiple VPE homologues. These were initially classified as either seed-type, vegetative-type or uncharacterised according to sequence and expression patterns (Yamada et al., 2005). However, functional redundancies evident from selective VPE knock-outs and the finding that vegetative-type VPEs can be detected in developing seeds and vice-versa: seed-type VPEs have been detected in vegetative tissues (D. F. Gruis et al., 2002; D. Gruis et al., 2004; Julián et al., 2013) suggests that this tissue-specific expression is not absolute. This functional overlap is unlikely to be complete since the loss of
a given homologue was not always fully compensated by other VPEs implying that subtle sequence differences may dictate preferred substrates and, consequently, function. Whilst this may also be due to inadequate expression levels of the compensating homologue, it raises the possibility that the cyclising ability recently attributed to VPEs is homologue specific rather than a capability intrinsic to all VPEs under the right conditions. Indeed, in extracts from the butelase 1 producing plant, C. ternatea, protein fractions that were active against the generic VPE substrate Z-AAN-MCA did not contain the cyclising enzyme and, conversely, the butelase 1 containing fraction did not display activity against this generic substrate, providing support for this argument (Giang Nguyen et al., 2014). To better understand the factors driving activity of VPEs as cyclases or proteases, the isolation and functional analysis of multiple VPE sequence variants from cyclotide producing plants will be required.

### 6.3 Structural Features of VPEs

VPEs are part of the CD clan of cysteine proteases and so share the same structure and mechanism as caspases and separases, despite their limited overall sequence identity (Noriyuki Hatsugai et al., 2015; Rawlings et al., 2012a). The CD clan of cysteine proteases consists of a conserved core of secondary structural elements, with a set of variable loops at the bottom, and another at the top containing the binding and catalytic sites (Figure 13a). The active site contains an absolutely conserved catalytic dyad of Cys and His, with some members additionally using a third residue (Asn) or backbone carbonyl to form a catalytic triad (Giang Nguyen et al., 2014; Rawlings et al., 2012b; Watt et al., 1999). The VPEs, like most CD clan protease families, are synthesised as zymogens that undergo self-catalysis to release the mature active proteinase (Cohen, 1997; Nagako Hiraiwa et al., 1999; Kuroyanagi et al., 2002; Rawlings et al., 2012a). Furthermore, VPEs can hydrolyse some caspase substrates, and their activity is blocked by some caspase inhibitors (N. Hatsugai et al., 2004; Kuroyanagi et al., 2005).

Within the CD clan, VPEs form a sister family (C13) to the caspase family (C14) due to a number of fundamental differences. Firstly, VPEs and caspases share limited sequence identity (N. Hatsugai et al., 2006). Secondly, VPEs are typically vacuolar and function optimally at acidic pH (Tetsu Kinoshita et al., 1999) whereas caspases are cytosolic with an optimum pH of 7.5 (Yamin et al., 1996). Thirdly, the definition of a caspase is the absolute requirement for an Asp at the P1 position whereas VPEs prefer cleavage after an Asn residue although they can cleave next to Asp residues (N. Hatsugai et al., 2006; Woltering et al., 2002).
The sequence and structural features of VPE that enable it to act as an efficient cyclase are currently unknown. Although the structure of a plant VPE is yet to be solved, the crystal structure of the related human legumain has been determined with and without its prodomains (Dall et al., 2013), allowing a homology model to be built for the known cyclase butelase 1 (Figure 13b).

The main difference between human legumain and the model VPE structure is an extended poly-proline loop (present in all plant VPEs) close to the P binding sites of the active site. However, because this is not present in the template structure, its exact structure and orientation is unknown. As observed in the crystal structure of human legumain, the C-terminal region of butelase 1 is predicted to be helical. This is reminiscent of the capping helices that impart cyclisation ability on the unrelated cyclase PatG (discussed in section 7.2) (Agarwal et al., 2012; Koehnke et al., 2012) and it has been hypothesised that this region of butelase 1 may therefore be important for favouring transpeptidation over hydrolysis (Giang Nguyen et al., 2014). However, the presence of these helices in the human enzyme, which is not reported to be a native cyclase, suggests that simply the presence or absence of this structural feature is unlikely to be an accurate predictor of cyclisation ability. The C-terminal auto-processing site of butelase 1 is yet to be experimentally defined (Giang Nguyen et al., 2014) and the precise nature of this could also affect the overall structure of this region. A further potentially important feature is the different binding pocket surface charges of butelase 1 and human legumain. Whereas legumain lacks any apparent specificity at its P’
sites and has a charged surface, the region thought to comprise the P’ binding sites in butelase 1 is more hydrophobic in agreement with the apparent preference for substrate hydrophobicity in at least some of these positions. Ultimately, the importance of any or all of these features cannot be known without a combination of structural and functional analyses of wt and mutant enzymes.

6.4 FUNCTIONS

In addition to the recently confirmed role in cyclisation (discussed in section 5), plant VPEs also exhibit targeted proteolysis to control a variety of cellular processes. Early reports described the processing of proprotein precursors from developing seeds to their mature forms by VPEs (I. Hara-Nishimura et al., 1987, 1991; Scott et al., 1992; Shimada et al., 1994) and the activation of other proteases to effect downstream outcomes such as protein degradation in senescing cells (Okamoto et al., 1999; Rojo et al., 2003). Subsequently, a role for VPE in programmed cell death (PCD) was observed in VPE-deficient transgenic plants challenged with the tobacco mosaic virus (TMV) and the fungal toxin fumonisin B1 (N. Hatsugai et al., 2004; Kuroyanagi et al., 2005). In each case, only the wild type plants displayed evidence of PCD at the infection site. VPEs have since been implicated in the PCD response to other stimuli including elicitors (Gauthier et al., 2007; Zhang et al., 2010), ER stress (Qiang et al., 2012) and heat stress (Li et al., 2012) as well as developmental (Noriyuki Hatsugai et al., 2015; Nakaune et al., 2005; Teper-Bamnolker et al., 2012) Defining the functional profile of specific VPE homologues will provide insight into the mechanisms of activity and evolution of both the protease and cyclase activities attributable to these enzymes in vivo.

In animals, limited proteolytic processing by legumains has also been implicated in a variety of cellular processes. For example, legumain has been reported to play a role in host defence via regulation of antigen processing in the class II MHC pathway (Manoury et al., 1998; Matthews et al., 2010) and processing of Toll-like receptors, which are involved in pathogen detection and mobilisation of the immune response (Ewald et al., 2011; Sepulveda et al., 2009). Recently, it has been reported that human legumain is capable of acting as a ligase via a mechanism independent of the catalytic cysteine (discussed in section 7.3) that requires a C-terminal Asn (Dall et al., 2015). Interestingly, butelase 1 reportedly produced a small amount of cyclic product from substrates lacking any residues C-terminal to the P1 Asn suggesting this novel mechanism could also be applicable to butelase 1 (Giang Nguyen et al., 2014). Given that this would require Asn-specific hydrolysis of cyclotide precursors in planta, followed by a separate cyclisation step, it seems unlikely to be of biological relevance. The functional significance of this capability of human legumain remains unknown, however production of
alternately spliced peptides for class II MHC presentation could be one possibility, as has been reported in the context of class I MHC presentation (Dalet et al., 2010; Hanada et al., 2004; Vigneron et al., 2004; Warren et al., 2006).

Mammalian legumain has also been associated with some types of cancer since overexpression of legumain has been reported in colorectal and breast cancer tissue and have also been correlated with a more invasive phenotype (Cheng Liu et al., 2003). As reported in section 6.2, in addition to the typical lysosomal localisation, active mammalian legumain has been observed in the nucleus in two colorectal cell lines (Haugen et al., 2013). The difference in pH in these two compartments (acidic vacuole and neutral nucleus) is inconsistent with reports that active legumain was irreversibly denatured at neutral pH (J. M. Chen, Rawlings, et al., 1998; JM Chen et al., 1997). However, this may be reconciled by the recent finding that a non-canonical processing event takes place at near neutral pH, imparting an asparaginyl-specific carboxypeptidase function (Dall et al., 2013). Binding of legumain to integrin αVβIII has also been reported (Y. Liu et al., 2012) and this was shown to render legumain more stable at higher pH revealing multiple layers of regulation on this versatile enzyme (Dall et al., 2013).

It is possible that compartmentalisation and the involvement of cofactors could also offer some level of control over the specific functions of plant VPEs as opposed to primary sequence differences. However the vacuolar localisation of cyclotides and cyclising ability of purified butelase 1 suggest that these factors do not play a role in controlling cyclisation.

6.5 SUMMARY

VPEs are present in a wide range of plants and animals and are involved in a range of cellular processes. They are self-processed from inactive zymogens and display characteristic Asx-specificity, although sequence requirements outside of this region are not well-conserved across species. They share structural homology with caspase-1, and exhibit a high degree of sequence conservation. VPEs display an array of functions that involve both protease and cyclase activity. Whilst not yet well understood, these distinct functions may be unique to specific VPE homologues. Further work will be required to fully elucidate the function and regulation of this interesting group of enzymes.
Cyclotides are one of many unrelated families of cyclic peptides that are produced ribosomally and enzymatically cyclised. There are several common features amongst the biosynthesis of these different families, indicating that similar cyclisation mechanisms have evolved repeatedly. Cyclic proteins so far identified are all produced with propeptide regions that are cleaved off during cyclisation to energetically pay for the transpeptidation reaction. Transpeptidation is typically performed by a cysteine or serine protease cyclising the peptide via an acyl-enzyme intermediate.

### 7.1 VPE INVOLVEMENT IN CYCLISATION OF OTHER CYCLIC PROTEIN FAMILIES

In addition to the cyclotides, there is evidence for VPE involvement in the maturation of another class of backbone cyclised peptides: the Paw-S derived family. This group includes the sunflower trypsin inhibitor SFTI-I, a 14 residue cyclic peptide comprised of two β-sheets cross-linked by a disulphide bond (Colgrave, Korsinczky, et al., 2010). Like the cyclotides, it is produced as linear propeptide, however the SFTI-I encoding domain is buried within a functional preproalbumin, Paw-S1 (Mylne et al., 2011). This is particularly surprising given that sunflower trypsin inhibitors show remarkable sequence, structure and functional convergence with the inhibitory loop of the Bowman-Birk Inhibitors (BBIs) which are over 100 amino acids long (Luckett et al., 1999). It is likely that these cyclic peptides recently evolved de novo within the genus *Helianthus* since several seed preproalbumins encode unique extra proteins that are released during maturation of the albumin (Yamada et al., 1999). Like cyclotides, conserved features of the precursor sequence led to the identification of a VPE as the likely cyclase (Mulvenna et al., 2005). This peptide also contains an Asn at the N-terminal P1 site and studies in VPE-null transgenic plants have demonstrated that proteolytic cleavage at both the C- and N-termiini is mediated by a VPE (Mylne et al., 2011). Interestingly, additional protease-mediated modification of SFTI occurs when it is inhibiting trypsin. In this case, the protease both cleaves the SFTI inhibitory loop and re-cyclises it (Colgrave, Korsinczky, et al., 2010; Marx et al., 2003).
7.2 Cyclisation by other enzymes

Segetalin A, a representative of another group of plant backbone cyclised peptides, the orbitides, is not cyclised by a VPE but by the serine protease PCY-1. This represents a broader example of convergent evolution since both serine and cysteine proteases act via a similar covalent mechanism. Cyclisation by PCY-1 is also largely independent of length and composition of the target peptide outside of a C-terminal recognition sequence (Barber et al., 2013).

Similarly, a serine protease has also been implicated in the cyclisation of the Cyanobactins; a family of short cyclic peptides from bacteria with particularly well characterised biosynthesis (Agarwal et al., 2012). Understanding their biosynthesis has been aided by the organization of bacterial genomes, whereby genes that act in a pathway are commonly clustered into operons. For example, the cyclotide Patellamide A (from Prochloron didemi) is encoded by the PatE gene in an operon that also contains a processing protease (PatA), the cyclising protease (PatG) and several post-translational modification enzymes (e.g. PatD).

The cyclising protease PatG is a subtilisin-like serine protease from the SB clan. Structural analysis revealed the presence of an additional two helices above the active site compared with other subtilisin-like proteases (Agarwal et al., 2012; Koehnke et al., 2012). The proximity of these helices to the catalytic serine led to speculation that this addition imparts the cyclisation ability of PatG by dictating the spatial position of the N- and C-termini. Indeed, when these helices are lacking, cyclisation ability is lost (Koehnke et al., 2012). It is possible that the C-terminus of a VPE may control access to the active site in a similar manner. As with butelase 1 ligation of cyclotides, Pat G does not require that substrates be prearranged in a near cyclical conformation for cyclisation to occur, although it is postulated that this may help offset the entropic costs of constraining an otherwise flexible substrate within the active site (Agarwal et al., 2012; McIntosh et al., 2010). PatG exhibits limited sequence specificity outside of the recognition motif and can cyclise a broad range of substrate sequences and lengths following addition of a short C-terminal sequence (McIntosh et al., 2010). Interestingly, there appears to be little N-terminal sequence requirements for this group of cyclic peptides suggesting that, provided access to the active site is not sterically hindered, the nature of the N-terminal sequence is unlikely to be crucial.

There is little mechanistic information available on how the enzyme achieves the active site desolvation presumably required for cyclisation, however the apparent promiscuity of butelase 1, PCY1 and Pat G suggest that this is mediated primarily by the enzyme and the short C-terminal recognition sequence in the substrate. In the case of PatG, it is hypothesised
that the C-terminal sequence motif remains bound to the active site post-cleavage from the main peptide, unable to be displaced by water, and is only disrupted by the incoming N-terminus of the peptide, leading to cyclisation to resolve the acyl intermediate (Koehnke et al., 2012). This is consistent with the C-terminal sequence requirements reported for cyclotides.

Threonine proteases also perform proteolysis by a covalent mechanism and can perform transpeptidation (F. Conlan et al., 2011). For example, the proteasome (a multi-subunit threonine protease) excises and ligates peptides together from the fibroblast growth factor 5 precursor protein (Dalet et al., 2010; Vigneron et al., 2004). Although there are currently no examples of cyclisation by a threonine protease, it is conceivable that one may yet be discovered.

7.3 A UNIQUE LIGATION MECHANISM

As a counterexample to the transpeptidation-based peptide ligation mechanisms described in Figure 10, there is evidence of an entirely different mechanism performed by human legumain. In addition to its canonical cysteine protease mechanism, legumain may be able to perform condensation ligation by a different mechanism using residues in the same active site (Figure 14). Adjacent to the His of the catalytic triad is an Asp residue (Asp$^{147}$) which forms a high-energy succinimide ring (Suc$^{147}$) by intramolecular condensation (PDB:4N6N). It has been proposed that this succinimide ring can act as an electrophile in a mechanism that involves conversion of the substrate C-terminal asparagine (Asn$_{39i}$) to a reactive succinimide (Suc$_{39i}$) which is then by the amino terminus of the second substrate (Dall et al., 2015).

A key feature of this succinimide mechanism is that it is not transpeptidation and so does not require the simultaneous cleavage of a propeptide. The energy to achieve ligation by condensation comes from the reactive succinimide ring. Additionally, the mechanism proceeds through a succinimide electrophile covalent intermediate which is entirely different from the canonical cysteine nucleophile covalent intermediate. It therefore represents an alternative possible specialised method of peptide ligation available to enzymes.
Figure 14 | Proposed condensation ligation mechanism for Legumain

The enzyme (black) contains a high-energy succinimide (Suc$_{147}$) ring in its active site. This unnatural residue is the result of an intramolecular condensation of an aspartate residue. The carboxylic acid of the first substrate peptide (red) performs a nucleophilic attack on the succinimide ring to form an acyl-enzyme intermediate. This intermediate is then attacked by the substrate asparagine (Asn$_{39i}$) to release the substrate. The enzyme now contains an iso-aspartic acid (iso-Asp$_{147}$) and the substrate contains a high-energy succinimide ring (Suc$_{39i}$). Finally, the substrate’s high-energy succinimide is attacked by the amine of the second substrate (blue) to form a new peptide bond and the enzyme regenerates its succinimide by intramolecular condensation. Overall, this condensation reaction yields a ligated peptide product.

7.4 Enzymatic Cyclisation in Biotechnology

Proteins that are naturally linear can be enzymatically cyclised using biotechnological tools to improve stability or increase activity. For example cyclised conotoxins (in development for
pain relief) are less susceptible to proteolysis (Clark et al., 2012), and cyclised histatin 1 promotes wound healing more efficiently (Oudhoff et al., 2009). Both nature and humans have repeatedly recruited proteases and run their native proteolysis reaction in reverse. Transpeptidation is achieved by coupling the cleavage of a C-terminal propeptide or activating group to intramolecular peptide bond formation to the N-terminus via an acyl-enzyme intermediate.

Sortase A is a bacterial transpeptidase (protease CL clan) which natively ligates proteins to the bacterial cell wall. It has been used in vitro to cyclise proteins whose N and C-termini are spatially close (Bolscher et al., n.d.; Wu et al., 2011). Analogous to the processing enzymes of naturally cyclic proteins, it performs transpeptidation via covalent catalysis using a cysteine and requires a small C-terminal propeptide of a single Gly residue which is cleaved during transpeptidation.

Similarly, subtiligase is an engineered mutant of the protease subtilisin (protease SB clan) that has been optimised to favour peptide ligation over hydrolysis. It can perform cyclisation of activated peptides, however instead of a propeptide, the peptide is activated by a chemically added ester group which removes the need for a C-terminal propeptide (Jackson et al., 1995). Once again, catalysis is covalent and the activating ester group is cleaved off during the transpeptidation reaction.
The biosynthesis of cyclotides is a multi-stage process occurring through several organelles and mediated by a number of enzymes. Expression, folding and disulphide formation occur in the ER followed by proteolytic processing in the vacuole. The final step, cleavage of the C-terminal propeptide, is coupled to protease-mediated intramolecular transpeptidation to form the mature cyclic product.

Characterisation of the cyclisation step has identified VPEs as the likely cyclase. The VPEs are members of the CD clan of cysteine proteases and are typically involved in maturation of proteins by proteolysis. Evidence for their role as cyclases comes both from genetic studies of mutants in vivo and from biochemical characterisation of the enzymes in vitro. The cyclisation mechanism is likely to be a modification of covalent proteolysis in which the N-terminus of the peptide binds into the VPE active site and performs the final nucleophilic attack instead of water. Characterisation of cyclotide cyclisation holds both evolutionary interest and potential as a biotechnological tool.

Cyclisation by VPEs is an example of convergent evolution on several levels. Firstly, VPEs have also been evolutionarily recruited to cyclise other circular peptides such as SFTI. This may indicate that VPEs have features that make them particularly suitable proteases for conversion to cyclases such as their localisation, abundance, activity or specificity. Secondly, recruitment of VPEs parallels similar evolution of serine proteases to act as cyclases for orbitides and a threonine protease performing linear transpeptidation on fibroblast growth factor 5. It is likely that the two-step, mechanism of the covalent, nucleophilic proteases makes them particularly suitable for the evolutionary switch from hydrolysis to transpeptidation.

Production of cyclotides both in vivo and in vitro is of interest since they represent important candidates for production of transgenic plants as well as potential lead compounds for drug design. Development of transgenic plants using cyclotides to improve pathogen resistance requires introduction the cyclotide precursor gene as well as genes for the processing enzymes if the target is not a native cyclotide producer. Likewise, the recombinant production of cyclotides for pharmaceutical applications requires an efficient method of cyclisation, and the native cyclase is likely the most efficient option. Finally, thorough characterisation of the cyclase may allow its use in cyclisation and transpeptidation of non-native substrates and so add it to the growing toolkit of protein biotechnology.
REFERENCES


Barber, C. J. S., Pujara, P. T., Reed, D. W., Chiwocha, S., Agarwal, V., Pierce, E., McIntos...974-228.


https://doi.org/10.1016/bs.abr.2015.08.005

CC-BY 4.0 Postprint


Gillon, A. D.,aska, J., Jennings, C. V., Guarino, R. F., Craik, D. J., Anderson, M. A., ... Anderson, M. A.


Ireland, D. C., Colgrave, M. L., Nguyencong, P., Daly, N. L., & Craik, D. J. (2006). Discovery and


Kinoshita, T., Nishimura, M., & Hara-Nishimura, I. (1995). Homologues of a vacuolar processing enzyme that are expressed in different organs in Arabidopsis thaliana. Plant Molecular Biology, 29(1), 81–89.


Yamada, K., Shimada, T., & Nishimura, M. (1999). PROTEIN CHEMISTRY AND STRUCTURE:
Multiple Functional Proteins Are Produced by Cleaving Asn-Gln Bonds of a Single Precursor by Vacuolar Processing Enzyme

Multiple Functional Proteins Are Produced by Cleaving Asn-Gln Bonds of a Single Precursor by Vacuolar.

http://doi.org/10.1074/jbc.274.4.2563


http://doi.org/10.1074/jbc.271.22.13273