

ARTICLE

Received 4 Aug 2015 | Accepted 16 Nov 2015 | Published 18 Dec 2015

DOI: 10.1038/ncomms10199

OPEN

# Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase

Karen S. Harris<sup>1,\*</sup>, Thomas Durek<sup>2,\*</sup>, Quentin Kaas<sup>2</sup>, Aaron G. Poth<sup>2</sup>, Edward K. Gilding<sup>2</sup>, Brendon F. Conlan<sup>1,†</sup>, Ivana Saska<sup>2</sup>, Norelle L. Daly<sup>2,†</sup>, Nicole L. van der Weerden<sup>1</sup>, David J. Craik<sup>2</sup> & Marilyn A. Anderson<sup>1</sup>

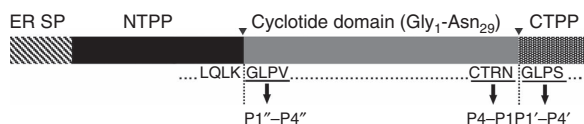
Cyclotides are diverse plant backbone cyclized peptides that have attracted interest as pharmaceutical scaffolds, but fundamentals of their biosynthetic origin remain elusive. Backbone cyclization is a key enzyme-mediated step of cyclotide biosynthesis and confers a measure of stability on the resultant cyclotide. Furthermore, cyclization would be desirable for engineered peptides. Here we report the identification of four asparaginyl endopeptidases (AEPs), proteases implicated in cyclization, from the cyclotide-producing plant *Oldenlandia affinis*. We recombinantly express *OaAEP1<sub>b</sub>* and find it functions preferably as a cyclase by coupling C-terminal cleavage of propeptide substrates with backbone cyclization. Interestingly, *OaAEP1<sub>b</sub>* cannot cleave at the N-terminal site of *O. affinis* cyclotide precursors, implicating additional proteases in cyclotide biosynthesis. Finally, we demonstrate the broad utility of this enzyme by cyclization of peptides unrelated to cyclotides. We propose that recombinant *OaAEP1<sub>b</sub>* is a powerful tool for use in peptide engineering applications where increased stability of peptide products is desired.

<sup>1</sup>Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia. <sup>2</sup>Division of Chemistry and Structural Biology, Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia. \*These authors contributed equally to this work. † Present addresses: Research School of Biology, Australian National University, Canberra, Australian Capital Territory 0200, Australia (B.F.C.); Centre for Biodiscovery and Molecular Development of Therapeutics, AIITHM, James Cook University, Cairns, Queensland 4878, Australia (N.L.D.). Correspondence and requests for materials should be addressed to M.A. (email: m.anderson@latrobe.edu.au)

Proteases are abundant throughout nature and are essential for a wide range of cellular processes. They typically serve to hydrolyse polypeptide chains, resulting in either degradation of the target sequence or maturation to a biologically active form. Less frequently, proteases can also ligate polypeptides, producing new or alternatively spliced variants. This unusual function has been reported for processes such as the maturation of the lectin concanavalin A<sup>1</sup>, peptide presentation by major histocompatibility complex class I molecules<sup>2</sup>, and anchoring of bacterial proteins to the cell wall<sup>3</sup>. Recently, this enzymatic transpeptidation has also been implicated in the backbone cyclization of ribosomally synthesized cyclic peptides<sup>4–10</sup>.

Cyclotides are a well-studied class of gene-encoded cyclic peptides that are expressed in plants and exhibit a range of bioactivities including insecticidal, nematocidal and molluscicidal activity against agricultural pests<sup>11–14</sup>. Structurally, they are characterized by a cyclic cystine knot motif that confers exceptional stability. Importantly, this stable framework can be used as a pharmaceutical scaffold, and bioactive sequences have been successfully grafted into cyclotides<sup>15</sup>. Backbone cyclization can also endow peptides with oral bioavailability, suggesting that this modification might find broad application in peptide drug engineering<sup>16–18</sup>. However, *in vitro* cyclization of synthetic peptides is challenging and the limited availability of enzymes capable of this process is a hurdle to large-scale production<sup>19,20</sup>. Furthermore, expression yields of cyclotides in transgenic plants that are not native cyclotide producers is poor, impeding transfer of agriculturally relevant bioactivities to other plants<sup>8,21</sup>. The mechanism of enzymatic cyclization intrinsic to cyclotide biosynthesis is poorly understood. Elucidating it will be important for the realization of the pharmaceutical and agricultural potential of cyclotides and for increasing the cyclization efficiency of unrelated ‘designed’ bioactive peptides.

Cyclotides are produced as precursors in which the cyclotide sequence is flanked by N- and C-terminal propeptides (Fig. 1). It is thought that enzymatic removal of the N-terminal propeptide precedes the final maturation step of C-terminal propeptide cleavage and ligation of the free N- and C-termini<sup>8,21</sup>. Only four native cyclases have been identified to date and the best characterized of these is the serine protease PatG, which cyclizes the bacterial cyanobactins<sup>4–7</sup>. In plants, the serine protease PCY1 cyclizes the segetalins; cyclic peptides from the Caryophyllaceae<sup>4</sup>. However, in the two other classes of plant-derived cyclic peptides (cyclotides and the PawS-derived cyclic peptides), strong Asx sequence conservation at the C-terminal P1 site implicates as possible cyclases the asparaginyl endopeptidases (AEPs), a group of cysteine proteases also known as vacuolar processing enzymes or Legumains, and this hypothesis is supported by studies in transgenic plants<sup>8,9,21,22</sup>.



**Figure 1 | Schematic representation of the *Oak1* gene.** The precursor protein encoded by the *Oak1* gene is proteolytically processed to mature kB1. ▼ indicates the N- and C-terminal processing sites. rOaAEP1<sub>b</sub> targets the C-terminal processing site. The C-terminal P1/P1' - P4/P4' sites are indicated. P1''-P4'' denote the N-terminal residues that replace the P1'-P4' residues on release of the C-terminal propeptide. CTPP, C-terminal propeptide; ER SP, endoplasmic reticulum signal peptide; NTPP, N-terminal propeptide.

Recently, an AEP (butelase 1) was isolated from the cyclotide-producing plant *Clitoria ternatea* and shown to cyclize a modified precursor of the prototypical cyclotide, kalata B1 (kB1) from *Oldenlandia affinis*, however, recombinant expression of functionally active butelase 1 has not been achieved, limiting its application<sup>5</sup>. Only one AEP with any cyclizing ability has been produced recombinantly, and this enzyme was highly inefficient, producing mainly hydrolysed substrate<sup>10</sup>. Here we report the identification, recombinant production and characterization of an *O. affinis* AEP that preferentially functions as a cyclase. The enzyme can cyclize native kalata substrate precursors and the unrelated anti-malarial peptide, R1, at close to 100% efficiency. This AEP releases the C-terminal propeptide of kB1, but it does not mediate the N-terminal processing event, which must occur first if efficient cyclization is to take place. Moreover, its specificity for model peptides mirrors the sequence requirements for cyclization of kB1 in transgenic plants, supporting a native function in the maturation of *O. affinis* cyclotides<sup>8,21</sup>.

## Results

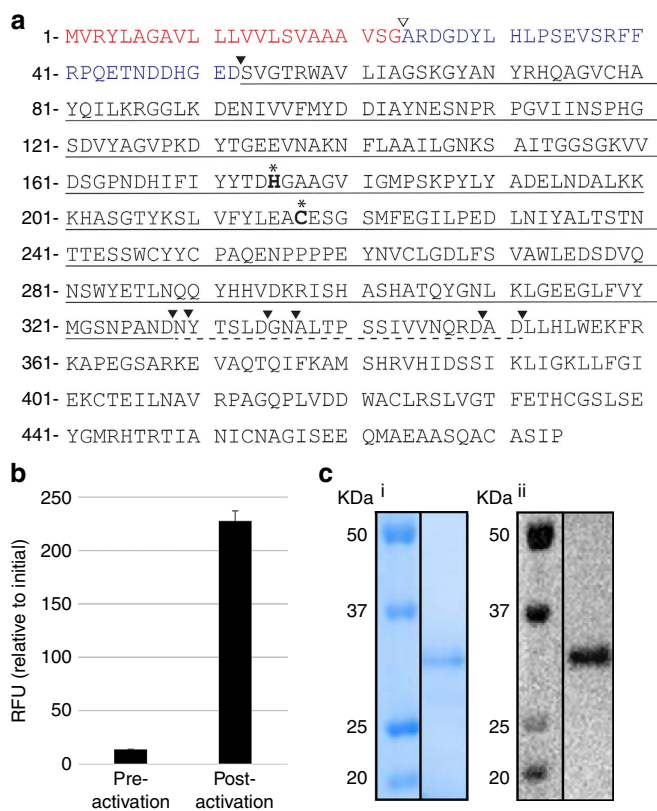
### Identification and recombinant expression of *O. affinis* AEPs.

Three expressed AEP isoforms were identified in an *O. affinis* complementary DNA library (OaAEP1-3) and a fourth sequence, with a single nucleotide change from OaAEP1 (resulting in a Glu<sub>371</sub>Val variant), was identified from genomic DNA (OaAEP1<sub>b</sub>) (Fig. 2a; Supplementary Fig. 1). The four isoforms share at least 77% identity at the protein level, as determined by pairwise protein alignments. When compared with butelase 1, 64–69% identity was observed, whereas identity with human legumain was 49–53%.

OaAEP1<sub>b</sub> was expressed in *Escherichia coli* as a His6-ubiquitin-AEP1<sub>b</sub> fusion protein (Supplementary Fig. 2a). AEPs are usually produced as zymogens that are self-processed at low pH to their mature, active form<sup>23–25</sup>. Consistent with this processing, activity of rOaAEP1<sub>b</sub> against an internally quenched fluorescent (IQF) peptide representing the native C-terminal processing site in kB1 (Table 1; wildtype (wt)) was markedly increased following incubation at pH 4.5 (Fig. 2b). After purification, a dominant band of ~32 kDa was evident by reducing SDS-polyacrylamide gel electrophoresis (PAGE) and confirmed to be rOaAEP1<sub>b</sub> by Western blotting (Fig. 2c; Supplementary Fig. 2b). The average total protein yield from two independent experiments was ~1.8 mg l<sup>-1</sup> after activation and purification, however batch to batch variation in purity was observed. Although glycosylation of some AEPs has been reported<sup>26</sup>, the production of an active form in *E. coli* confirms that this is not a requirement for activity of *O. affinis* AEP1<sub>b</sub>.

Mass spectrometry (MS)/MS sequencing of peptide fragments generated from tryptic, chymotryptic or endoGlu-C digestion of the activated enzyme identified several peptide fragments with non-canonical cleavage sites, suggesting that they may be derived from rOaAEP1<sub>b</sub> auto-processing events (Supplementary Fig. 2). This allowed Asp52 to be assigned as the likely N-terminal auto-processing site and Asp328, Asn329, Asp334, Asn336, Asp349 and/or Asp351 as potential C-terminal processing sites (Fig. 2a; Supplementary Fig. 2). No peptides downstream of Asp351 were identified, indicating that the activation was essentially complete and that the C-terminal domain (Leu352–Pro474) was removed during the post-activation purification step. The theoretical mass of the processed forms (30.4–32.8 kDa) is in good agreement with that determined by SDS-PAGE/Western blotting (Fig. 2c).

Consistent with cysteine proteases of this class, rOaAEP1<sub>b</sub> was inhibited by iodoacetamide (1 mM), but was not affected by E64 (250 μM) or pepstatin A (10 μM) (Supplementary Fig. 3). Ac-YVAD-CHO (500 μM), a caspase-1 inhibitor reported to also



**Figure 2 | Expression of active *rOaAEP1<sub>b</sub>* in *E. coli*.** (a) Sequence of *OaAEP1<sub>b</sub>* predicted from *O. affinis* genomic DNA. Predicted ER signal sequence shown in red; N-terminal propeptide shown in blue; the putative signal peptidase cleavage site is indicated by  $\nabla$  and autocatalytic processing sites by  $\blacktriangledown$ . The mature *OaAEP1* cyclase domain is underlined and the C-terminal auto-processing region is indicated with broken underline. The putative catalytic dyad is shown in bold and labelled with \*. (b) An *rOaAEP1*-containing anion exchange fraction pre- and post-activation at pH 4.5 (5 h, 37 °C) was tested for activity against the wt IQF peptide (14  $\mu$ M). Baseline fluorescence from a no substrate control has been subtracted and the relative fluorescence intensity (RFU) at  $t = 90$  min is reported. The average of two technical replicates is shown and error bars report the range. (c) Activated *rOaAEP1<sub>b</sub>* purified by cation exchange was analysed by SDS-PAGE and (i) Instant blue staining or (ii) Western blotting with anti-*OaAEP1<sub>b</sub>* (residues D<sub>47</sub>-P<sub>474</sub>) polyclonal rabbit serum.

inhibit AEPs<sup>27</sup>, was a poor inhibitor of the recombinant enzyme suggesting that at least some P' residues are important for active site targeting.

**Substrate specificity.** The activity of *rOaAEP1<sub>b</sub>* against IQF peptides representing wt and mutant versions of the native kB1 C-terminal cleavage site was determined (Table 1; Supplementary Fig. 4a). Along with the strict P1 Asx specificity characteristic of AEPs<sup>24</sup>, *rOaAEP1<sub>b</sub>* exhibited strong P2' selectivity since after Leu<sub>31</sub>Ala substitution within the IQF peptide barely any hydrolysis was observed. This observation is consistent with the lack of cyclic product generated when the corresponding mutation was introduced *in planta*<sup>21</sup>. Kinetic parameters ( $V_{\max}$ ,  $K_m$  and  $k_{\text{cat}}$ ) are reported where applicable (Table 1). The turnover rates ( $k_{\text{cat}}$ ) reported here ( $\sim 0.06$ – $1.6 \text{ min}^{-1}$ ) are much slower than that reported for recombinant human legumain assayed against a small substrate ( $\sim 8 \text{ s}^{-1}$ ) (ref. 28). This is not unexpected given that *rOaAEP1<sub>b</sub>* prefers to carry out cyclization, rather than the hydrolysis being measured here. Supporting the observed P2' selectivity, *rOaAEP1<sub>b</sub>* was unable to cleave the generic AEP substrate Z-AAN-MCA (Supplementary Fig. 5).

The substrate specificity of *rOaAEP1<sub>b</sub>* was compared with that of recombinant human legumain (rhuLEG; Supplementary Fig. 4b)<sup>29</sup>. A stringent P1 Asx requirement was again observed; however, in contrast to *rOaAEP1<sub>b</sub>*, rhuLEG cleaved the Leu<sub>31</sub>Ala substrate at a rate similar to the wt substrate, demonstrating that P2' specificity is not a feature of all AEPs.

**Cyclization of kB1 precursors.** To explore the cyclization ability of *rOaAEP1<sub>b</sub>*, processing of correctly folded (as determined by NMR) synthetic kB1 precursors was assessed by MS. When incubated with the wt kB1 precursor carrying the native C-terminal pro-hepta-peptide (GLPSLAA), the active enzyme produced a peptide of 2,891.2 Da (monoisotopic,  $[M+H]^+$ ), consistent with the expected mass of mature, cyclic kB1 (Fig. 3a). This product was confirmed to be identical to native kB1 by reversed phase-high performance liquid chromatography (RP-HPLC) co-elution (Supplementary Fig. 6) and one- and two-dimensional-NMR experiments (Supplementary Fig. 7). Kinetic parameters ( $\pm$  s.e.m.) for the processing of the wt kB1 precursor were  $0.53 (\pm 0.1) \text{ s}^{-1}$  for  $k_{\text{cat}}$ ,  $212 (\pm 76) \mu\text{M}$  for  $K_m$  and  $2,500 \text{ M}^{-1} \text{ s}^{-1}$  for  $k_{\text{cat}}/K_m$  as determined from a Michaelis–Menten plot (Supplementary Fig. 8). While the turnover rate ( $k_{\text{cat}}$ ) is lower than that reported for the plant-derived

**Table 1 | Kinetic parameters of IQF peptide cleavage by *rOaVPE1<sub>b</sub>*.**

IQF peptide	Sequence*	$V_{\max}$ (nmoles $\text{min}^{-1} \text{mg}^{-1}$ protein) ( $\pm$ s.e.m.) <sup>†</sup>	$K_m$ ( $\mu\text{M}$ ) ( $\pm$ s.e.m.) <sup>†</sup>	$k_{\text{cat}}$ ( $\text{min}^{-1}$ ) ( $\pm$ s.e.m.) <sup>†,‡</sup>
wt	Abz-STRN $\downarrow$ GLPS-Y(3NO <sub>2</sub> )	51.3 ( $\pm$ 5.8)	55.0 ( $\pm$ 6.4)	1.6 ( $\pm$ 0.2)
R <sub>28</sub> A	Abz-STAN $\downarrow$ GLPS-Y(3NO <sub>2</sub> )	6.9 ( $\pm$ 0.6)	13.0 ( $\pm$ 2.4)	0.2 ( $\pm$ 0.02)
R <sub>28</sub> K	Abz-STKN $\downarrow$ GLPS-Y(3NO <sub>2</sub> )	29.3 ( $\pm$ 3.5)	42.0 ( $\pm$ 4.0)	0.9 ( $\pm$ 0.1)
N <sub>29</sub> A	Abz-STRA $\downarrow$ GLPS-Y(3NO <sub>2</sub> )	NA <sup>§</sup>	–	–
N <sub>29</sub> Q	Abz-STRQ $\downarrow$ GLPS-Y(3NO <sub>2</sub> )	NA <sup>§</sup>	–	–
N <sub>29</sub> D	Abz-STRD $\downarrow$ GLPS-Y(3NO <sub>2</sub> )	$\sim 2$ <sup>  </sup>	ND <sup>  </sup>	$\sim 0.06$ <sup>  </sup>
G <sub>30</sub> A	Abz-STRN $\downarrow$ ALPS-Y(3NO <sub>2</sub> )	51.0 ( $\pm$ 2.0)	29.0 ( $\pm$ 1.4)	1.6 ( $\pm$ 0.07)
G <sub>30</sub> S	Abz-STRN $\downarrow$ SLPS-Y(3NO <sub>2</sub> )	35.5 ( $\pm$ 2.7)	31.4 ( $\pm$ 2.5)	1.1 ( $\pm$ 0.08)
L <sub>31</sub> A	Abz-STRN $\downarrow$ GAPS-Y(3NO <sub>2</sub> )	NA <sup>§</sup>	–	–
L <sub>31</sub> I	Abz-STRN $\downarrow$ GIPS-Y(3NO <sub>2</sub> )	ND <sup>¶</sup>	ND <sup>¶</sup>	ND <sup>¶</sup>

IQF, internally quenched fluorescent; NA, no activity; ND, not determined.

\*IQF peptide residues are numbered according to their position within the native kB1 precursor, where the mature cyclotide incorporates Gly<sub>1</sub>-Asn<sub>29</sub>; native Cys<sub>26</sub> is substituted with Ser to avoid unpaired Cys residues.

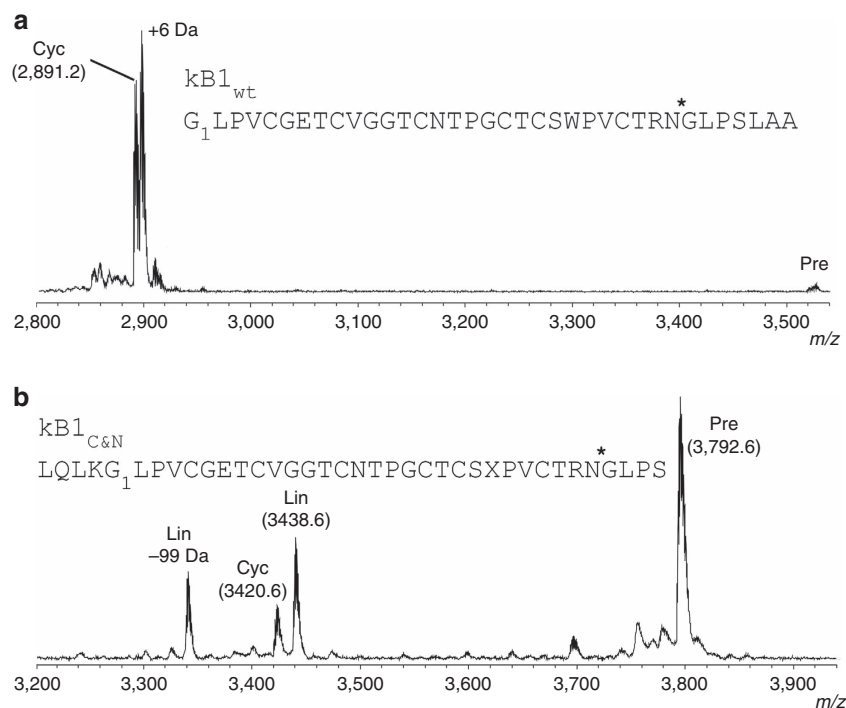
<sup>†</sup> $N \geq 3$ ;  $\pm$  standard error of the mean (s.e.m.)

<sup>‡</sup> $k_{\text{cat}}$  is a conservative estimate assuming that the total concentration of active enzyme is equal to the total protein concentration in the enzyme preparation and an enzyme mass of 32 kDa.

<sup>§</sup>No activity detected under the conditions tested (up to 80  $\mu\text{M}$  substrate; up to 6 h incubation).

<sup>||</sup>Low  $V_{\max}$  precluded accurate estimation of kinetic parameters.

<sup>¶</sup> $K_m$  above the range of the substrate concentrations used in this analysis precluded accurate estimation of kinetic parameters.



**Figure 3 | Enzymatic processing products of linear kB1 precursors.** (a) MALDI MS profile of a kB1 precursor ( $kB1_{wt}$ ) containing the C-terminal propeptide in the presence of  $rOaAEP1_b$  (22 h incubation). The +6 Da peak was observed only in the presence of reducing agent and corresponds to the reduced form of cyclic kB1. (b) MALDI MS profile of a kB1 precursor ( $kB1_{C\&N}$ ) containing four C-terminal propeptide residues and four N-terminal propeptide residues in the presence of  $rOaAEP1_b$  (20.5 h incubation). A side product originating from chemical synthesis likely represents a Val deletion (-99 Da). Data are representative of at least two technical replicates. X, benzoylphenylalanine. \* denotes  $rOaAEP1_b$  cleavage site. Observed monoisotopic masses (Da;  $[M + H]^+$ ) for dominant peaks are listed. Cyc, cyclic product; Lin, linear product; Pre, linear precursor.

butelase 1 ( $17.08 s^{-1}$ ; ref. 5), it is far higher than that of the recombinantly expressed cyclase PatG (1 per day; ref. 7).

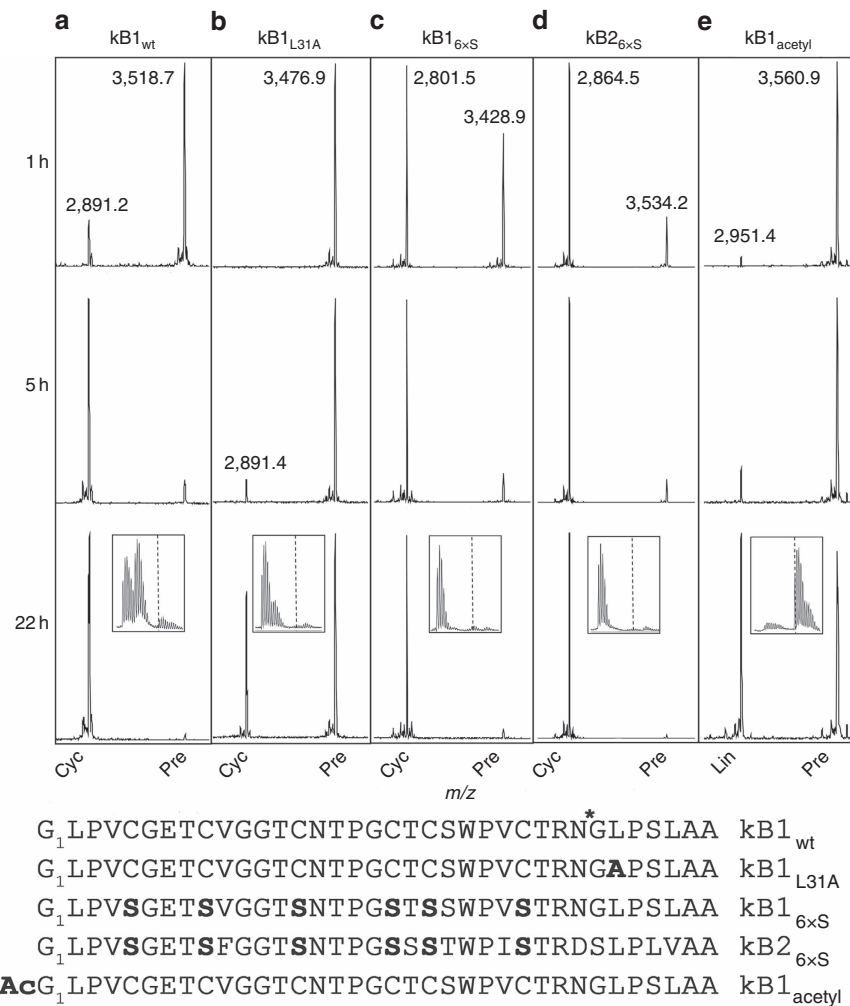
To determine if  $rOaAEP1_b$  could also carry out the N-terminal processing required for cyclotide maturation, a kB1 precursor was tested that contained the folded cyclotide domain flanked by four residues from each of the N- and C-terminal propeptides (Fig. 3b). No N-terminal processing was observed, indicating that this processing is conducted by an enzyme other than  $OaAEP1_b$ . Although the bulk of the precursor remained intact after 20.5 h, the predominant processing product was a linear peptide lacking the C-terminal propeptide, demonstrating that correct N-terminal processing must occur before cyclization. Interestingly, a mass corresponding to a cyclized version of the C-terminally processed peptide (that is, C-terminal propeptide residues released, N-terminal propeptide residues remaining) was also observed, although this was the least abundant product.

**Processing of modified cyclotide precursors.** To further probe cyclization requirements, we tested  $rOaAEP1_b$ -mediated processing of modified kB1 and kB2 precursors over time (Fig. 4). When presented within an IQF peptide, the  $Leu_{31}Ala$  substrate analogue was not hydrolysed by  $rOaAEP1_b$  (Table 1). However, the same substitution within the kB1 precursor did not preclude cyclization by  $rOaAEP1_b$ , although this version was cyclized far more slowly than the wt precursor (Fig. 4a,b). Surprisingly, the presence of disulfide bonds in the cyclotide precursors is not a requirement for cyclization since a kB1 substrate analogue in which all six cysteines were substituted with serines was also efficiently cyclized (Fig. 4c). The absence of a defined kB1-like structure in the  $kB1_{6xS}$  mutant was confirmed by NMR spectroscopy (Supplementary Fig. 7c). Similarly, a kB2 linear

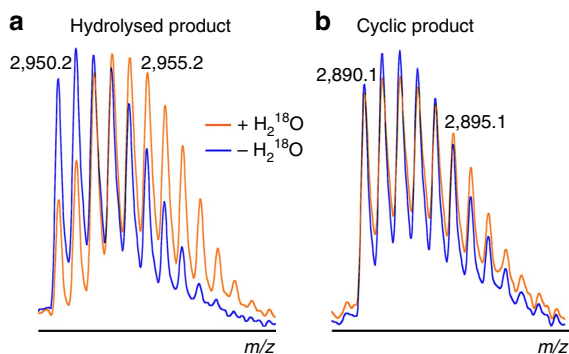
precursor with the same Cys→Ser substitutions was also efficiently cyclized, confirming that enzyme activity is not specific to individual cyclotides (Fig. 4d).

In cysteine protease-mediated peptide bond hydrolysis, nucleophilic attack of a water molecule is required to resolve the acyl-enzyme thioester intermediate. However, during peptide cyclization (or transpeptidation) the substrate's N-terminal amine is postulated to function as a competing nucleophile, facilitating aminolysis of the reactive thioester intermediate<sup>30</sup>. Accordingly, a kB1 precursor with an acetyl-capped N-terminal amine was processed only to a linear peptide lacking the C-terminal propeptide (Fig. 4e). This hydrolysis occurred at a slower rate than cyclization of the wt precursor (compare with Fig. 4a). Water can therefore access the active site of  $rOaAEP1_b$ , but cyclization is favoured over hydrolysis in the presence of an appropriately positioned nucleophile.

**Water is excluded during cyclization.** An alternative ligation mechanism, distinct from transpeptidation, was recently proposed for huLEG<sup>31</sup>. In that mechanism, initial hydrolysis of the C-terminal propeptide is followed by a separate ligation event requiring a C-terminal Asn residue in the substrate. To distinguish between these mechanisms in the case of  $rOaAEP1_b$ , reactions were carried out in the presence of  $^{18}O$ -labelled water and the products were analysed by high-resolution MS. An isotopic shift consistent with the incorporation of  $^{18}O$  was evident following enzymatic hydrolysis of the N-terminal acetylated kB1 precursor to give a linear product (Fig. 5a). However, there was no isotopic shift after processing of the wt precursor to a cyclic product, suggesting that hydrolysis is unlikely to play a role in cyclization by  $rOaAEP1_b$  (Fig. 5a).

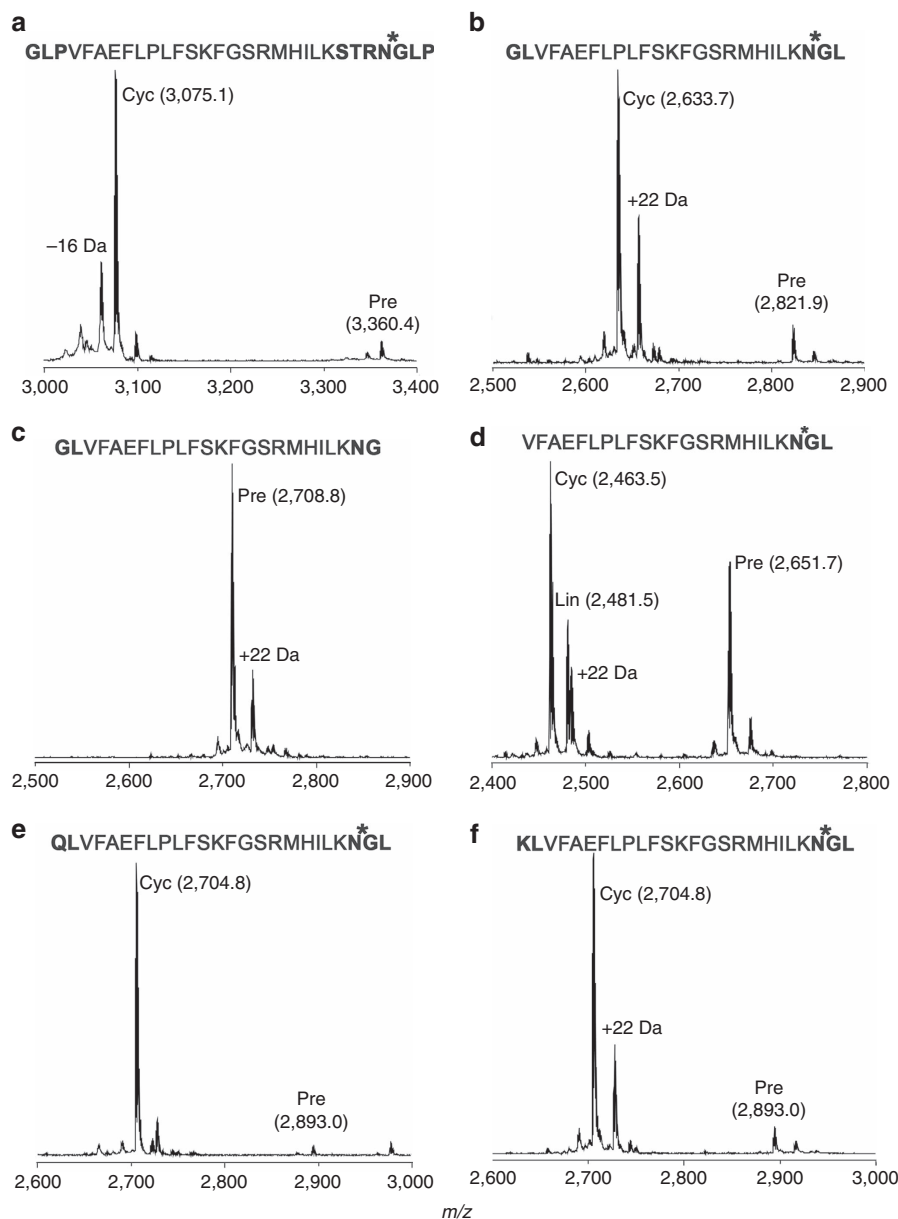


**Figure 4 | Modified linear kB1 and kB2 precursors are cyclized at different rates.** MALDI MS spectra of (a) kB1<sub>wt</sub>, (b) kB1<sub>L31A</sub>, (c) kB1<sub>6xS</sub>, (d) kB2<sub>6xS</sub> and (e) kB1<sub>acetyl</sub> cyclotide precursors at 1, 5 and 22 h post-enzyme addition. Data are representative of three technical replicates. \* denotes rOaAEP1<sub>b</sub> cleavage site. Observed monoisotopic masses (Da; [M + H]<sup>+</sup>) for dominant peaks are listed. Boxed inset at the 22 h time point zooms in on the region containing the processing product. Approximate positions of the monoisotopic mass of processed products is indicated by ∴. Cyc, cyclic product; Lin, linear product; Pre, linear precursor.



**Figure 5 | Enzymatic cyclization excludes water.** MALDI MS profile of the enzymatic processing products of (a) kB1<sub>acetyl</sub> and (b) kB1<sub>wt</sub> linear precursors in the presence and absence of <sup>18</sup>O-labelled water. An isotope shift indicative of <sup>18</sup>O incorporation only occurs during hydrolysis. Observed masses of two isotopic peaks (Da; [M + H]<sup>+</sup>) are indicated. Data are representative of three technical replicates.

**rOaAEP1<sub>b</sub> can cyclise an unrelated peptide.** We also investigated cyclization of other substrates structurally unrelated to cyclotides by rOaAEP1<sub>b</sub>, focussing on the anti-malarial peptide R1 (refs 32,33). This peptide was efficiently cyclized following the addition of N- and C-terminal AEP recognition sequences (Fig. 6a). Sequential trimming of the added recognition residues revealed that cyclization could be achieved following the addition of only a C-terminal Asn–Gly–Leu motif (although some linear product was also produced from this precursor) (Figs 6a–d). Lys and Gln were also accepted in place of Gly at the N terminus (Figs 6e–f) with little impact on yield at the time point tested. No processing of either the native R1 peptide or a modified R1 carrying the N-terminal Gly–Leu motif with only an Asn at the C terminus was observed (Supplementary Fig. 9). Subsequent digestion with endoGlu-C confirmed that, in all cases, rOaAEP1<sub>b</sub> processing produced cyclic peptide (Supplementary Fig. 10). Evidence of an additional, linear, rOaAEP1<sub>b</sub>-generated cleavage product was only observed for the R1 variant without any N-terminal flanking residues (Fig. 6d; Supplementary Fig. 10c).



**Figure 6 | Flanking sequence requirements for cyclization of a model peptide by rOaAEP1.** (a–f). MALDI MS spectra of the R1 peptide (VFAEFLPLFSKFGSRMHILK) with various flanking sequences 22 h post addition of rOaAEP1<sub>b</sub>. Bold residues, flanking sequences. \*denotes rOaAEP1<sub>b</sub> cleavage site. Observed monoisotopic masses (Da; [M+H]<sup>+</sup>) are listed. +22 Da and –16 Da peaks present in some precursor and product spectra are likely to represent Na<sup>+</sup> adducts and a synthesis-derived modification respectively. Data are representative of three technical replicates. Cyc, cyclic product; Lin, linear product; Pre, linear precursor.

## Discussion

This study reports the cloning of four AEPs from the cyclotide-producing plant *O. affinis*; one of which was recombinantly expressed. The recombinant enzyme required self-processing to produce the active product: a cyclase that preferentially and efficiently couples C-terminal processing with C- and N-terminal ligation of linear *O. affinis* cyclotide precursors. Furthermore, this cyclizing ability was highly efficient when transferred to an unrelated anti-malarial peptide, demonstrating broad applicability in peptide engineering.

Consistent with other auto-inhibited proteases, rOaAEP1<sub>b</sub> required proteolytic activation to achieve maximum activity (Fig. 2b). The observed N-terminal auto-processing site (Asp52) is consistent with other experimentally validated N-terminal auto-processing sites identified in jack bean AEP<sup>34</sup>, butelase 1

(ref. 5) and human legumain<sup>28,35</sup> (Fig. 2a, Supplementary Fig. 1). In contrast, six potential C-terminal auto-processing sites (Asp328/334/349/351, Asn329/336) were observed within a region particularly rich in Asn/Asp residues (324–351). This finding is in agreement with the multiple C-terminal maturation steps recently described for rhuLEG<sup>28,35</sup>. Regardless of which of these sites is relevant *in planta*, the instability of active AEPs above pH 6 (refs 28,29,36) will likely preclude direct production of active enzyme in *E. coli*. Activated rOaAEP1<sub>b</sub> proteolytically removed the C-terminal (but not N-terminal) propeptide of a kB1 precursor and resolved the acyl-intermediate in a hydrolysis-independent manner, generating a backbone cyclized product (Figs 3 and 5). rOaAEP1<sub>b</sub> could also hydrolyse precursors lacking a free N-terminal amine to produce linear products, albeit at a slower rate. Although it is unknown if

*OaAEP1<sub>b</sub>* performs both these cyclase and protease activities *in vivo*, the observed preference for cyclization over hydrolysis suggests that it probably functions predominantly as a cyclase.

Dual protease/ligase capabilities have been reported for PatG<sup>37</sup>, a serine protease involved in cyclic peptide production in cyanobacteria, and more recently for human legumain<sup>31</sup>. Separate mechanistic pathways for hydrolysis and ligation have been proposed for human legumain: proteolysis proceeds via hydrolysis of a cysteinyl-thioester enzyme intermediate, whereas peptide ligation occurs via activation of the free peptidyl- $\alpha$ -carboxylate through transient formation of an enzyme-linked anhydride intermediate that is subsequently resolved via aminolysis<sup>31</sup>. The catalytically critical residue for this mechanism of ligation, Asp188, is conserved in *OaAEP1<sub>b</sub>*, but several lines of evidence preclude a role for this pathway in *OaAEP1<sub>b</sub>*-mediated peptide cyclization. First, *rOaAEP1<sub>b</sub>* cannot cyclize peptides carrying a free C-terminal Asn, the minimal proposed substrate requirement in the alternative pathway (Supplementary Fig. 9). Second, our H<sub>2</sub><sup>18</sup>O experiments demonstrate the absence of <sup>18</sup>O incorporation into the cyclized product, which strongly indicates that cyclization does not follow a hydrolysis/ligation mechanism as proposed in the alternative pathway (Fig. 5). Third, our MS/MS data for *rOaAEP1<sub>b</sub>* show no evidence of a reactive succinimide enzyme intermediate required for formation of the substrate–enzyme anhydrides<sup>31</sup>. Hence, our mechanistic data are in agreement with the traditional concerted mechanism, in which some of the energy from the (exergonic) cleavage of the C-terminal Asn-propeptide bond is preserved in the form of a thioester intermediate and used to overcome the energetically unfavourable (endergonic) peptide bond formation (cyclization) in the second step. However, alternative mechanisms may still play a role in cyclization mediated by other AEPs or with alternative substrates.

In the context of this established mechanism, C- and N-terminal proximity was thought to be crucial for cyclization to be favoured over hydrolysis<sup>8</sup>. Here we show that pre-organization of C- and N-termini in the substrate is not required by *rOaAEP1<sub>b</sub>* since unconstrained cyclotide precursors lacking the characteristic disulfide-bonded structure are efficiently cyclized (Fig. 4). Furthermore, *rOaAEP1<sub>b</sub>* can cyclize an anti-malarial peptide that is structurally and functionally unrelated to cyclotides following the addition of short flanking sequences (Fig. 6). These findings are consistent with the limited structural and/or sequence requirements imposed by other native cyclases on their substrates<sup>4,5,7</sup>. Conceivably, polypeptides of diverse composition and length may be cyclized by *rOaAEP1<sub>b</sub>*, provided that association of the C- and N-termini is not sterically hindered.

The application of this technology is limited to peptides that can retain activity following incorporation of the additional residues required for AEP-mediated processing. While *rOaAEP1<sub>b</sub>* can cyclize a model peptide with only a single non-native residue incorporated to the mature peptide, this is at the cost of cyclization efficiency (Fig. 6d). Understanding the interplay between the sequence requirements for efficient cyclization and retention of bioactivity for a given target peptide will be crucial to realizing the potential of AEP-mediated cyclization. Importantly, the estimated turnover rate of *rOaAEP1<sub>b</sub>* ( $k_{\text{cat}}$ , 0.53 s<sup>-1</sup>; Supplementary Fig. 8) is multiple orders of magnitude higher than the recombinantly produced cyclase PatG (1 d<sup>-1</sup>; ref. 7), supporting its widespread application in peptide engineering.

In a previous study, a conserved tripeptide motif C terminal to both the N- and C-terminal cyclotide processing sites was identified through cyclotide sequence analysis<sup>8</sup>. This motif is Gly–Leu–Pro in the kB1 sequence, and its importance for efficient cyclization is supported by mutagenesis studies in transgenic

plants<sup>8,21</sup>. The protease specificity of *rOaAEP1<sub>b</sub>* reported here against IQF peptides mirrors these requirements (Table 1). At the C-terminal processing site, the P1 Asn and P2' Leu are particularly well-conserved, and both were crucial for both *in planta* cyclization and *in vitro* cleavage of model peptides. However, over the longer incubation period of the cyclization assays, a kB1 precursor with a Leu<sub>31</sub>Ala mutation was still enzymatically processed to a cyclic product (Fig. 4). It was initially proposed that the conserved Leu residue at the P2' position of cyclotides was important for preventing water from accessing the active site during cyclization. However the observed cyclization of the kB1 Leu<sub>31</sub>Ala mutant suggests that the role of a conserved bulky hydrophobic residue at the P2' position is only to promote appropriate enzyme–substrate interaction. Congruent with this hypothesis, the absence of a P2' residue renders substrates poor targets of both *rOaAEP1<sub>b</sub>* (cyclization and hydrolysis; Fig. 6c; Supplementary Fig. 5) and butelase 1 (at least for hydrolysis)<sup>5</sup>. In summary, our results suggest a cyclization model in which the cleaved C-terminal propeptide retains sufficient affinity to remain bound to the active site until it is displaced by the incoming N terminus of the peptide, finally leading to cyclization by resolving the acyl intermediate<sup>8,38</sup>.

This P2' requirement is not characteristic of all AEPs<sup>9,34,39</sup> and might be a predictor of cyclization ability within this protease family. Indeed, in extracts from *C. ternatea*, protein fractions that were active against the generic AEP substrate Z-AAN-MCA (which does not contain a P2' residue) did not contain the cyclizing enzyme and, conversely, the butelase 1 containing fraction did not display activity against this substrate<sup>5</sup>. Here we report the presence of four unique AEP sequences in *O. affinis* (Fig. 2a; Supplementary Fig. 1), and demonstrate that one (*OaAEP1<sub>b</sub>*) is capable of cyclizing *O. affinis* cyclotide precursors. Further work will investigate whether all, or a subset of, *O. affinis* AEPs (at least two more of which should exist to explain all the AEP contig sequences observed) exhibit this function and whether their substrate specificity is an accurate predictor of cyclization ability.

The constraints on the sequence of the incoming N terminus may not be very stringent. At least two residues with different properties are accepted in place of Gly at the P1' position by *rOaAEP1<sub>b</sub>* with comparable yields of cyclic product under the conditions tested (Fig. 6e,f). Butelase 1 and Pat G also exhibit promiscuity in this region<sup>5,37</sup>. Interestingly, Gly<sub>1</sub> is highly conserved across cyclotides from different plant species, raising the possibility that selection at this position is not driven by AEP cyclase specificity. In transgenic plants, more stringent requisites were observed and no cyclic product was made from kB1 precursors with a conservative Gly<sub>1</sub>Ala mutation<sup>8</sup>. Because AEP cannot liberate the N-terminal propeptide, we hypothesize that selection at Gly<sub>1</sub> might be driven by the putative N-terminal processing enzyme. Further analysis will be necessary to determine if this reflects differences in the enzyme homologues being assayed (that is, AEPs from the model plant *Nicotiana benthamiana* compared with AEPs from native cyclotide producers) or experimental conditions.

In conclusion, this study unequivocally demonstrates the involvement of an AEP in maturation of native *O. affinis* cyclotides, advancing our understanding of the biosynthesis of this important class of cyclic peptides. Furthermore, the promiscuous yet highly efficient nature of this recombinantly produced enzyme highlights its exciting potential value as a biological tool for cyclization of a range of bioactive peptides.

## Methods

**Peptide substrates and inhibitors.** IQF peptides containing an N-terminal *o*-aminobenzoic acid (Abz) group and a C-terminal 3-nitrotyrosine (Y[3NO<sub>2</sub>])

were synthesized by Genscript at >90% purity. Control IQF peptides representing the predicted cleavage products of the wt peptide (Abz-STRN; GLPS-Y(3NO<sub>2</sub>)) were also synthesized by Genscript at >90% purity. All IQF peptides were solubilized in 25% (v/v) acetonitrile:water. The fluorogenic peptide substrate Z-AAN-MCA (where Z is carboxybenzyl; MCA is 7-amido-4-methylcoumarin) and the caspase inhibitor Ac-YVAD-CHO (where Ac, acetyl; CHO, aldehyde) were supplied by the Peptide Institute and solubilized in dimethyl sulfoxide. The linear cyclotide precursor peptides kB1<sub>wt</sub>, kB1<sub>C8N</sub>, kB1<sub>L31A</sub> and kB1<sub>acetyl</sub> were chemically synthesized in-house by standard Fmoc solid-phase peptide synthesis. Folding and disulfide formation was carried out by incubating the reduced peptides in folding buffer (100 mM ammonium bicarbonate, 50% isopropanol, 2 mM reduced glutathione, 1 mM oxidized glutathione, pH 8.2) for 3 days<sup>40</sup>. The products were isolated by RP-HPLC at >95% purity and characterized by high-resolution MS and NMR spectroscopy. Peptides kB1<sub>6xS</sub> and kB2<sub>6xS</sub> as well as R1 and its derivatives were supplied by Genscript at >85% purity, as determined by RP-HPLC and MS. Peptides were dissolved in ultrapure water before analysis.

**O. affinis transcriptome.** Total RNA was extracted from *O. affinis* root, leaf and seedling tissues using a phenol extraction method. Plant material was frozen in liquid nitrogen and ground to a fine powder, which was then resuspended in buffer (0.1 M Tris-HCl pH 8.0, 5 mM EDTA, 0.1 M NaCl, 0.5% SDS, 1% 2-mercaptoethanol), extracted twice with 1:1 phenol:chloroform and precipitated by addition of isopropanol. The pellets were dissolved in 0.5 ml water and RNA was precipitated overnight at 4 °C by addition of 4 M lithium chloride. The extracted RNA of each tissue was analysed by GeneWorks using the Illumina GAIIX platform. In total, 69.3 million 75 bp paired-end reads were generated. Reads were filtered with a phred confidence value of Q37 and assembled into contigs using Oases<sup>41</sup> with k-mer ranging from 41–67. The assemblies were merged using cd-hit-est<sup>42</sup>, resulting in 270,000 contigs. Statistics on the depth of sequencing were made by aligning the reads of each tissue on the contigs using BWA<sup>43</sup>. All the sequences, including one AEP, previously identified from an EST library of *O. affinis* were present among the contigs<sup>44</sup>. Homologues of this AEP sequence were searched using BLAST<sup>45</sup> in the contig library using a maximum E-value of 1e-20, resulting in the identification of 371 putative AEP transcripts. These sequences could then be clustered in 13 groups sharing at least 90% sequence identity using cd-hit<sup>42</sup>.

**OaAEP1-3 cloning.** Full-length AEP transcripts from the *O. affinis* transcriptome assembly were used to design a set of primers. A single degenerate forward primer (OaAEPdegen-F, 5'-ATG GTT CGA TAT CYC GCC GG-3') was sufficient to amplify all sequences since variability within the extreme 5' region of each full-length transcript was limited to a single nucleotide position. Three reverse primers (OaAEP1-R, 5'-TCA TGA ACT AAA TCC TCC ATG GAA AGA GC-3'; OaAEP2-R, 5'-TTA TGC ACT GAA TCC TTT ATG GAG GG-3'; OaAEP3-R, 5'-TTA TGC ACT GAA TCC TCC ATC G-3') were designed with the aid of Primer 3 (ref. 46). Each primer set successfully amplified an AEP sequence.

To clone expressed OaAEPs, total RNA was extracted from *O. affinis* leaves and shoots using TRIzol (Life Technologies) and was reverse transcribed with SuperScript III reverse transcriptase (Life Technologies) according to the manufacturer's instructions<sup>11</sup>. Target sequences were amplified from the resulting complementary DNA using Phusion High Fidelity Polymerase (New England BioLabs) and the primers described above under the recommended reaction conditions. Gel extracted PCR products were dA-tailed by incubation with Invitrogen Taq Polymerase (Life Technologies) and 0.5 μl 10 mM dA in the supplied buffer. The processed products were cloned into pCR8-TOPO (Life Technologies) and transformed into *E. coli*. Purified DNA from clones that were PCR positive for an AEP insert were sent for Sanger sequencing at the Australian Genome Research Facility (www.agrf.org). Coding sequences have been deposited in Genbank (accession codes: OaAEP1 (KR259377), OaAEP2 (KR259378), OaAEP3 (KR259379)).

In an alternative approach, genomic DNA was extracted from *O. affinis* leaf tissue using a DNeasy Plant Mini Kit according to the manufacturer's instructions. PCR amplification from this DNA used primers specifically targeting the OaAEP1 nucleotide sequence. Gel extracted product was dA-tailed as above, cloned into the TOPO vector and transformed into *E. coli*. Sequencing of PCR-positive clones identified a fourth sequence with a single amino acid change from OaAEP1 (OaAEP1<sub>b</sub>).

**Antibodies.** Polyclonal anti-OaAEP1<sub>b</sub> rabbit serum was generated by immunizing a New Zealand White rabbit with a denatured, inactive form of OaAEP1<sub>b</sub> (residues D<sub>47</sub>–P<sub>474</sub>) that was produced recombinantly in *E. coli*. The rabbit received three doses, four weeks apart, of 150 μg of antigen in 50% (v/v) PBS and Freund's incomplete adjuvant. Serum was obtained 2 weeks after the final dose and used at a 1:2,000 dilution for Western blotting.

**Recombinant expression of *O. affinis* AEP1<sub>b</sub> (rOaAEP1<sub>b</sub>).** Initial trials to produce active *O. affinis* AEP1<sub>b</sub> based on predicted N- and C-terminal processing sites (residues D<sub>47</sub>–D<sub>420</sub>) were unsuccessful and subsequent expression attempts incorporated both N- and C-terminal prodomains. DNA encoding full-length

*O. affinis* AEP1<sub>b</sub> without the putative signalling domain (residues A<sub>24</sub>–P<sub>474</sub>) was inserted into the pHUE vector<sup>47</sup> to give a His6-ubiquitin-OaAEP1<sub>b</sub> fusion protein construct (Supplementary Fig. 2) and introduced into T7 shuffle *E. coli* cells (New England BioLabs). Transformed cells were grown at 30 °C in superbroth (3.5% tryptone (w/v), 2% yeast extract (w/v), 1% glucose (w/v), 90 mM NaCl, 5 mM NaOH) to mid-log phase; the temperature was then reduced to 16 °C and expression was induced with isopropyl β-D-1-thiogalactopyranoside (0.4 mM) for ~20 h. Cells were harvested by centrifugation and resuspended in non-denaturing lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% Triton X 100, 1 mM EDTA, pH 7). Lysis was promoted by a total of five freeze/thaw cycles and the addition of lysozyme (hen egg white; 0.4 mg ml<sup>-1</sup>). DNase (bovine pancreas; 40 μg ml<sup>-1</sup>) and MgCl<sub>2</sub> (0.4 M) were also added. Cellular debris was removed by centrifugation and the lysate was stored at -80 °C until required.

**Purification and activation of rOaAEP1<sub>b</sub>.** Lysate containing expressed rOaAEP1<sub>b</sub> was filtered through a 0.1 μm glass fibre filter before being diluted 1:8 in buffer A (20 mM bis-Tris, 0.2 M NaCl, pH 7) and loaded onto two 5 ml HiTrap Q Sepharose high performance columns connected in series (GE Healthcare; 1.6–3.1 ml undiluted lysate per millilitre resin). Bound proteins were eluted with a continuous salt gradient (0–30% buffer B (20 mM bis-Tris, 2 M NaCl, pH 7); 15 column volumes) and AEP-positive fractions identified by Western blotting (anti-OaAEP1<sub>b</sub> rabbit serum (1:2,000); peroxidase-conjugated anti-rabbit IgG (GE Healthcare NA934; 1:5,000)). To self-activate rOaAEP1<sub>b</sub>, EDTA (1 mM) and Tris(2-carboxyethyl)phosphine hydrochloride (0.5 mM) were added, the pH was adjusted to 4.5 with glacial acetic acid and the protein pool was incubated for 5 h at 37 °C. Protein precipitation at this pH allowed removal of the bulk of the contaminating proteins by centrifugation. The remaining protein was filtered (0.22 μm), diluted 1:8 in buffer A2 (50 mM acetate, pH 4) then captured on a 1 ml HiTrap SP Sepharose high performance column (GE Healthcare). Bound proteins were eluted with a salt gradient (0–100% buffer B2 (50 mM acetate, 1 M NaCl, pH 4); 10 column volumes) and AEP-positive fractions were pooled. The final product was analysed by MS and reducing SDS-PAGE followed by Western blotting and staining with Instant blue (Expedeon). The total concentration of protein was estimated by bicinchoninic acid assay according to the manufacturer's instructions.

**Identification of the auto-processing sites of rOaAEP1<sub>b</sub>.** Aliquots of rOaAEP1<sub>b</sub> (5 μl) were diluted 1:1 with either 100 mM ammonium bicarbonate pH 8.0 (trypsin, chymotrypsin) or 100 mM ammonium phosphate pH 8.0 (endoGlu-C) and enzymatically digested with endoGlu-C, trypsin or chymotrypsin (100 ng μl<sup>-1</sup>). Cleavages were conducted over 16 h at 37 °C (endoGlu-C, trypsin) or 30 °C (chymotrypsin). Injections of each digest (5 μl) were introduced to a Shimadzu nanoLC delivering a linear acetonitrile gradient at a flow rate of 500 nl min<sup>-1</sup> for reversed-phase separation on a C18 Zorbax column (Agilent 300SB-C18, 3.5 μm particle size, 150 mm × 100 μm). Column eluate was interfaced directly with a 5600 TripleTOF LC-MS/MS instrument (AB SCIEX, Canada) equipped with a nanoelectrospray ionization source.

Tandem MS data were generated in Information Dependent Acquisition experiments, wherein full-scan TOF-MS spectra were acquired for 250 ms over *m/z* 350–1,800, and the 20 most intense signals with charge state +2 to +5 were selected for product ion scans of duration 50 ms over *m/z* 80–1,400. Data were acquired and processed using the Analyst TF 1.6 software (AB SCIEX, Canada). Automated protein identification was performed in ProteinPilot 4.0 using the Paragon algorithm with 'no enzyme' settings to interrogate an *E. coli* K12 proteome database (Uniprot) concatenated with sequences of rOaAEP1<sub>b</sub> and those of the three cleavage enzymes, with false discovery rate set at *P* < 0.05. Spectral data for peptides software matched to rOaAEP1<sub>b</sub> were confirmed via manual interpretation in Analyst TF 1.6.

**Assaying protease activity against fluorescent peptides.** To assay activity of rOaAEP1<sub>b</sub> against both internally quenched and other fluorescent peptides, substrate and enzyme were diluted as appropriate in activity buffer (50 mM sodium acetate, 50 mM NaCl, 1 mM EDTA, 0.5 μM Tris(2-carboxyethyl)phosphine hydrochloride, pH 5). To assay activity of rhuLEG (R&D systems) against the same substrates, the enzyme was first activated by incubation in 50 mM sodium acetate, 100 mM NaCl, pH 4 (4 μl buffer/1 μl enzyme) for 2 h at 37 °C. Substrates and activated rhuLEG were diluted in 50 mM MES, 250 mM NaCl, pH 5 as required. Diluted enzyme and substrate were added to black, flat bottomed microtiter plates in a total assay volume of 110–200 μl. The change in fluorescence intensity over time was monitored on a SpectraMax M2 (Molecular Devices) using excitation/emission wavelengths of 320/420 nm (IQF peptides) or 360/460 nm (other fluorescent peptides). Substrate and enzyme concentrations in each assay and the time point presented are as indicated in the figure legends.

To determine the kinetics of rOaAEP1<sub>b</sub> activity against IQF peptides, each substrate was assayed at a range of concentrations between 2.5 and 80 μM in a total volume of 200 μl. The total protein concentration of the enzyme preparation used in the kinetic assays was 3.5 μg ml<sup>-1</sup>. It was not possible to precisely determine the concentration of active enzyme due to impurities remaining in the preparation and the absence of an inhibitor appropriate for active site titration. However, a conservative turnover rate (*k*<sub>cat</sub>) was estimated based on a mass of 32 kDa and



the assumption that the total protein concentration reflected active enzyme. At each substrate concentration, initial velocities were calculated from the linear portion of the progress curve.  $K_m$  and  $V_{max}$  were estimated using the Michaelis-Menten equation and the curve-fitting program GraphPad Prism (GraphPad Software, San Diego).

The high peptide concentrations required for estimating kinetic parameters necessitated the use of a correction factor to account for the inner filter effect; a phenomenon where high relative concentrations of the quenching group impede detection of the signal from the fluorescent donor even after substrate hydrolysis<sup>48</sup>. This was achieved as described previously<sup>48</sup>. The output generated by the fluorescent hydrolysis product (Abz-STRN) was measured in the presence of each concentration of non-hydrolysed substrate. The correction factor was the ratio between the expected and observed fluorescence signal at each substrate concentration. The corrected signal for each data point was then converted to amount of product by comparison to a standard curve of the fluorescent hydrolysis product.

**Inhibition assays.** To investigate the impact of inhibitors on enzyme activity against the wt IQF peptide, rOaAEP1<sub>b</sub> (4.4  $\mu\text{g ml}^{-1}$  total protein) was incubated with the indicated concentration of E64, Ac-YVAD-CHO, pepstatin A and iodoacetamide for 40 min before addition to the substrate (11  $\mu\text{M}$ ). Enzyme activity against the wt IQF peptide was then assessed as described above.

**Cyclization assay.** Linear target peptides (280  $\mu\text{M}$ ) were incubated with rOaAEP1<sub>b</sub> (12  $\mu\text{g ml}^{-1}$  total protein unless otherwise indicated) in activity buffer. The reaction was allowed to proceed for up to 22 h at room temperature and was analysed by matrix-assisted laser desorption/ionization MS (MALDI MS), RP-HPLC or NMR as appropriate.

To confirm the presence of cyclic product, R1 derivatives processed by rOaAEP1<sub>b</sub> were subsequently digested with endoGlu-C (25  $\mu\text{g ml}^{-1}$ ) in reaction buffer (50 mM Tris-HCl, 0.5 mM Glu-Glu, pH 8) such that the final dilution of the cyclization mix was 1:4. The reaction was allowed to proceed for 18 h at 37 °C before analysis by MALDI MS.

In heavy water experiments, isotopically labelled water (97 atom % <sup>18</sup>O) was used in place of unlabelled water. Linear target peptides (70  $\mu\text{M}$ ) were incubated with rOaAEP1<sub>b</sub> (6  $\mu\text{g ml}^{-1}$  total protein) in a non-reducing activity buffer (50 mM sodium acetate, 50 mM NaCl, 1 mM EDTA, pH 5) for 22 h at room temperature. The final H<sub>2</sub><sup>18</sup>O concentration in the assay was 81%.

**MS to track cyclization of linear peptides.** Cyclization of linear target peptides was monitored by MALDI MS. The reaction mixture (10–20  $\mu\text{l}$ ) was desalted using C18 zip tips and eluted in 4  $\mu\text{l}$  50% acetonitrile, 0.1% trifluoroacetic acid (TFA). A saturated MALDI matrix solution ( $\alpha$ -cyano-4-hydroxycinnamic acid) prepared in 95% acetonitrile, 0.1% TFA was diluted 1:2.2 such that the final matrix solution comprised 90% acetonitrile, 0.1% TFA and 1 mM NH<sub>4</sub>H<sub>2</sub>PO<sub>4</sub>. Eluted samples were mixed 1:4 with the MALDI matrix, spotted onto a MALDI plate and analysed by an Ultraflex III TOF/TOF (Bruker) in positive reflector mode.

**Purification of kB1 following in vitro cyclization.** The crude cyclization mixture was loaded to an Agilent Zorbax C18 reversed-phase column (4.6  $\times$  250 mm, 300 Å) and separated on a Shimadzu Prominence system using a linear gradient of 5–55% buffer B (90% acetonitrile, 10% H<sub>2</sub>O, 0.05% TFA) in buffer A (0.05% TFA/H<sub>2</sub>O) over 60 min. Fractions were collected manually, analysed by MALDI MS essentially as described above and lyophilized. Analytical HPLC and co-elution studies with chemically synthesized kB1 were carried out as described above.

**Nuclear magnetic resonance spectroscopy.** All peptides were dissolved in 90% H<sub>2</sub>O/10% D<sub>2</sub>O at a concentration of ~2.0  $\text{mg ml}^{-1}$  (0.5–0.75 mM). kB1 obtained from *in vitro* cyclizations was dissolved at 0.3  $\text{mg ml}^{-1}$  (~0.1 mM). Spectra were recorded on a Bruker Avance 600 MHz spectrometer equipped with a cryoprobe at 298 K. Phase-sensitive mode using time-proportional phase incrementation for quadrature detection in the  $t_1$  dimension was used for all two-dimensional spectra. Excitation sculpting with gradients was used to achieve water suppression. NMR experiments included TOCSY using a MLEV-17 spin lock sequence with an 80 ms mixing time, and NOESY with a 200 ms mixing time. Spectra were recorded with 4,096 data points in the  $F_2$  dimension and 512 increments in the  $F_1$  dimension. The  $t_1$  dimension was zero-filled to 1,024 real data points, and the  $F_1$  and  $F_2$  dimensions were multiplied by a sine-squared function before Fourier transformation. The spectra were referenced to the water signal at 4.77 p.p.m. at 298 K. All spectra were processed using TopSpin (Bruker) and manually assigned with CCPNMR using the sequential assignment protocol.

**Cyclization kinetics.** To determine the kinetics of rOaAEP1<sub>b</sub> activity against the wt kB1 precursor, the substrate was assayed at room temperature at a range of concentrations between 75 and 250  $\mu\text{M}$  in a total volume of 20–160  $\mu\text{l}$  of activity buffer. The total protein concentration of the enzyme preparation added to the kinetic assays was 19.7  $\mu\text{g ml}^{-1}$ . The reaction was quenched after 5 min with 0.1% TFA and the volume adjusted to 800  $\mu\text{l}$ . A volume of 700  $\mu\text{l}$  was loaded onto a

reversed-phase C18 analytical column (Agilent Eclipse C18, 5  $\mu\text{m}$ , 4.6  $\times$  150 mm) and peptides were separated by HPLC (19 min linear gradient of 12–60% acetonitrile, 0.1% TFA at 1  $\text{ml min}^{-1}$ ). The identity of eluted peaks was confirmed using MALDI MS. The area under the curve corresponding to the precursor peptide was quantitated by comparison to a standard curve and initial velocities were calculated by converting this to  $\mu\text{moles product formed}$ . Kinetic parameters were estimated using the Michaelis-Menten equation and the curve-fitting program GraphPad Prism (GraphPad Software, San Diego). As for the IQF peptides, it was not possible to precisely determine the concentration of active enzyme due to impurities remaining in the preparation and the absence of an inhibitor appropriate for active site titration. However, a conservative turnover rate ( $k_{cat}$ ) was estimated based on a mass of 32 kDa and the assumption that the total protein concentration reflected active enzyme. Differences in enzyme preparations means these parameters are not directly comparable to those determined for the IQF peptides.

## References

- Sheldon, P. S., Keen, J. N. & Bowles, D. J. Post-translational peptide bond formation during concanavalin A processing *in vitro*. *Biochem. J.* **320**, 865–870 (1996).
- Hanada, K., Yewdell, J. W. & Yang, J. C. Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* **427**, 1–5 (2004).
- Mazmanian, S. K., Liu, G., Ton-That, H. & Schneewind, O. Staphylococcus aureus Sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760–763 (1999).
- Barber, C. J. S. *et al.* 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 (2013).
- Nguyen, G. K. T. *et al.* Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol.* **10**, 732–738 (2014).
- Luo, H. *et al.* Peptide macrocyclization catalyzed by a prolyl oligopeptidase involved in  $\alpha$ -amanitin biosynthesis. *Chem. Biol.* **21**, 1610–1617 (2014).
- Lee, J., McIntosh, J., Hathaway, B. J. & Schmidt, E. W. Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *J. Am. Chem. Soc.* **131**, 2122–2124 (2009).
- Gillon, A. D. *et al.* Biosynthesis of circular proteins in plants. *Plant J.* **53**, 505–515 (2008).
- Saska, I. *et al.* An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclization. *J. Biol. Chem.* **282**, 29721–29728 (2007).
- Bernath-Levin, K. *et al.* Peptide macrocyclization by a bifunctional endopeptidase. *Chem. Biol.* **22**, 571–582 (2015).
- Jennings, C., West, J., Waine, C., Craik, D. & Anderson, M. Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis*. *Proc. Natl Acad. Sci. USA* **98**, 10614–10619 (2001).
- Plan, M. R., Saska, I., Cagauan, A. G. & Craik, D. J. Backbone cyclised peptides from plants show molluscicidal activity against the rice pest *Pomacea canaliculata* (golden apple snail). *J. Agric. Food Chem.* **56**, 5237–5241 (2008).
- Colgrave, M. L. *et al.* Cyclotides: natural, circular plant peptides that possess significant activity against gastrointestinal nematode parasites of sheep. *Biochemistry* **47**, 5581–5589 (2008).
- Colgrave, M. L. *et al.* Anthelmintic activity of cyclotides: *In vitro* studies with canine and human hookworms. *Acta Trop.* **109**, 163–166 (2009).
- Poth, A. G., Chan, L. Y. & Craik, D. J. Cyclotides as grafting frameworks for protein engineering and drug design applications. *Biopolymers* **100**, 480–491 (2013).
- Clark, R. J. *et al.* Engineering stable peptide toxins by means of backbone cyclization: stabilization of the alpha-conotoxin MII. *Proc. Natl Acad. Sci. USA* **102**, 2–7 (2005).
- Clark, R. J. *et al.* The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angew. Chem. Int. Ed. Engl.* **49**, 6545–6548 (2010).
- Chan, L. Y. *et al.* Cyclization of the antimicrobial peptide gomesin with native chemical ligation: influences on stability and bioactivity. *ChemBiochem* **14**, 617–624 (2013).
- Craik, D. J. Host-defense activities of cyclotides. *Toxins (Basel)* **4**, 139–156 (2012).
- Stanger, K. *et al.* Backbone cyclization of a recombinant cystine-knot peptide by engineered Sortase A. *FEBS Lett.* **588**, 4487–4496 (2014).
- Conlan, B. F. *et al.* Insights into processing and cyclization events associated with biosynthesis of the cyclic Peptide kalata B1. *J. Biol. Chem.* **287**, 28037–28046 (2012).
- Mylne, J. S. *et al.* Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nat. Chem. Biol.* **7**, 257–259 (2011).
- Hiraiwa, N., Nishimura, M. & Hara-Nishimura, I. Expression and activation of the vacuolar processing enzyme in *Saccharomyces cerevisiae*. *Plant J.* **12**, 819–829 (1997).

24. Hiraiwa, N., Nishimura, M. & Hara-Nishimura, I. Vacuolar processing enzyme is self-catalytically activated by sequential removal of the C-terminal and N-terminal propeptides. *FEBS Lett.* **447**, 213–216 (1999).
25. Kuroyanagi, M. *et al.* 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 (2002).
26. Kembhavi, A. A., Buttle, D. J., Knight, G. & Barrett, A. J. The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* **303**, 208–213 (1993).
27. Hatsugai, N. *et al.* A plant vacuolar protease, VPE, mediates virus-induced hypersensitive cell death. *Science* **305**, 855–858 (2004).
28. Dall, E. & Brandstetter, H. Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proc. Natl Acad. Sci. USA* **110**, 10940–10945 (2013).
29. Chen, J. *et al.* Cloning, Isolation, and Characterization of Mammalian Legumain, an Asparaginyl Endopeptidase. *J. Biol. Chem.* **272**, 8090–8098 (1997).
30. Conlan, B. F. *et al.* Circular proteins and mechanisms of cyclization. *Biopolymers* **94**, 573–583 (2010).
31. Dall, E., Fegg, J. C., Briza, P. & Brandstetter, H. Structure and mechanism of an aspartimide-dependent peptide ligase in human legumain. *Angew. Chem. Int. Ed. Engl.* **54**, 2917–2921 (2015).
32. Harris, K. S. *et al.* Binding hot spot for invasion inhibitory molecules on plasmodium falciparum apical membrane antigen 1. *Infect. Immun.* **73**, 6981–6989 (2005).
33. Harris, K. S. *et al.* Rapid optimization of a peptide inhibitor of malaria parasite invasion by comprehensive N-methyl scanning. *J. Biol. Chem.* **284**, 9361–9371 (2009).
34. Abe, Y. *et al.* Asparaginyl Endopeptidase of Jack Bean Seeds. *J. Biol. Chem.* **268**, 3525–3529 (1993).
35. Dall, E. & Brandstetter, H. 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**, 24–31 (2012).
36. Chen, J. M., Dando, P. M., Stevens, R. A., Fortunato, M. & Barrett, A. J. Cloning and expression of mouse legumain, a lysosomal endopeptidase. *Biochem. J.* **335**(Pt 1): 111–117 (1998).
37. McIntosh, J. A. *et al.* Circular logic: nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst. *J. Am. Chem. Soc.* **132**, 15499–15501 (2010).
38. Koehnke, J., Bent, A., Houssen, W. E., Zollman, D. & Morawitz, F. The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nat. Struct. Mol. Biol.* **19**, 767–772 (2012).
39. Jung, R. *et al.* The role of proteolysis in the processing and assembly of 11S seed globulins. *Plant Cell* **10**, 343–357 (1998).
40. Simonsen, S. M., Daly, N. L. & Craik, D. J. Capped acyclic permutants of the circular protein kalata B1. *FEBS Lett.* **577**, 399–402 (2004).
41. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092 (2012).
42. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
44. Qin, Q. *et al.* Identification of candidates for cyclotide biosynthesis and cyclisation by expressed sequence tag analysis of *Oldenlandia affinis*. *BMC Genomics* **11**, 111 (2010).
45. Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
46. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291 (2007).
47. Catanzariti, A., Soboleva, T. A., Jans, D. A., Board, P. G. & Baker, R. T. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.* **13**, 1331–1339 (2004).
48. Liu, Y. *et al.* Use of a fluorescence plate reader for measuring kinetic parameters with inner filter effect correction. *Anal. Biochem.* **267**, 331–335 (1999).

## Acknowledgements

We thank Ms Rosemary Guarino for assistance with expressions, Dr Pedro Quimbar and Dr Thomas Shafee for helpful discussions, Mr Owen McCorkelle and Dr Suresh Mathivanan for preliminary MS analysis, Mr James Brown for isolation of the AEP sequence from genomic DNA, Professor Michael Foley for providing the native R1 peptide, Mr Alun Jones and the Institute for Molecular Bioscience Mass Spectrometry Facility for expertise and access to MS equipment and Dr Joshua Mylne for assistance with transcriptomics. This work was supported by The Australian Research Council grants DP0984390 and DP150100443 and Hexima Ltd.

## Author contributions

K.S.H. produced the recombinant enzyme and carried out cyclization assays and kinetic analysis. T.D. and N.L.D. synthesized cyclotide precursors. T.D. carried out NMR and HPLC co-elution studies. T.D., Q.K., E.K.G. and I.S. contributed to the isolation of AEP sequences. T.D. and A.G.P. carried out mass spectrometry analysis of the recombinant AEP. B.F.C. produced rabbit antiserum. K.S.H., T.D., N.L.v.d.W., D.J.C. and M.A.A. contributed to study design and data analysis. All authors contributed to the writing and/or review of the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Harris, K. S. *et al.* Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat. Commun.* **6**:10199 doi: 10.1038/ncomms10199 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>