Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase

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

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 \textit{Oldenlandia affinis}. We recombinantly express \textit{OaAEP\textsubscript{1b}} and find it functions preferably as a cyclase by coupling C-terminal cleavage of propeptide substrates with backbone cyclization. Interestingly, \textit{OaAEP\textsubscript{1b}} cannot cleave at the N-terminal site of \textit{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 \textit{OaAEP\textsubscript{1b}} is a powerful tool for use in peptide engineering applications where increased stability of peptide products is desired.
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, peptide presentation by major histocompatibility complex class I molecules, and anchoring of bacterial proteins to the cell wall. Recently, this enzymatic transpeptidation has also been implicated in the backbone cyclization of ribosomally synthesized cyclic peptides.

Cyclotides are a well-studied class of gene-encoded cyclic peptides that are expressed in plants and exhibit a range of bioactivities including insecticidal, nematicidal and molluscidal activity against agricultural pests. 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. Backbone cyclization can also endow peptides with oral bioavailability, suggesting that this modification might find broad application in peptide drug engineering. 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. 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.

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. Only four native cyclotides have been identified to date and the best characterized of these is the serine protease PatG, which cyclizes the bacterial cyanobactin. In plants, the serine protease PCY1 cyclizes the segetalins; cyclic peptides from the Caryophyllaceae. 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 some AEPs has been reported, the production of an active form of 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.

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. Only one AEP with any cyclizing ability has been produced recombinantly, and this enzyme was highly inefficient, producing mainly hydrolysed substrate. 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.

**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 Glu47→Val variant), was identified from genomic DNA (OaAEP1g) (Table 2; 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%.

OaAEP1b was expressed in Escherichia coli as a His6-ubiquitin-AEP1 fusion protein (Supplementary Fig. 2a). AEPs are usually produced as zymogens that are self-processed at low pH to their mature, active form. Consistent with this processing, activity of rOaAEP1b 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 rOaAEP1b, by Western blotting (Fig. 2c; Supplementary Fig. 2b). The average total protein yield from two independent experiments was ~1.8 mg l⁻¹ after activation and purification, however batch to batch variation in purity was observed. Although glycosylation of some AEPs has been reported, the production of an active form in E. coli confirms that this is not a requirement for activity of O. affinis AEP1b.

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 rOaAEP1b 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, rOaAEP1b 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
inhibit AEPs\(^{27}\), 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\(_b\) 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\(^{28}\), rOaAEP1\(_b\) exhibited strong P2' selectivity since after Leu\(_{31}\)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\(^{21}\). Kinetic parameters (\(V_{\text{max}}\) \(K_m\) and \(k_{\text{cat}}\)) are reported where applicable (Table 1). The turnover rates (\(k_{\text{cat}}\)) reported here (\(~0.06–1.6\) \(\text{min}^{-1}\)) 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\(_b\) prefers to carry out cyclization, rather than the hydrolysis being measured here. Supporting the observed P2' selectivity, rOaAEP1\(_b\) was unable to cleave the generic AEP substrate Z-AAN-MCA (Supplementary Fig. 5).

The substrate specificity of rOaAEP1\(_b\) was compared with that of recombinant human legumain (rhuLEG; Supplementary Fig. 4b)\(^{29}\). A stringent P1 Asx requirement was again observed; however, in contrast to rOaAEP1\(_b\), rhuLEG cleaved the Leu\(_{31}\)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\(_b\), 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 (\(\pm 0.1\)) \(s^{-1}\) for \(k_{\text{cat}}\), 212 (\(\pm 76\)) \(\mu\text{M}\) for \(K_m\) and 2,500 \(M^{-1}\) \(s^{-1}\) 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\(_b\).**

| IQF peptide | Sequence* | \(V_{\text{max}}\) (nmol min\(^{-1}\) mg\(^{-1}\) protein) (\(\pm\)s.e.m.)\(^{†}\) | \(K_m\) (\(\mu\text{M}\)) (\(\pm\)s.e.m.)\(^{†}\) | \(k_{\text{cat}}\) (min\(^{-1}\)) (\(\pm\)s.e.m.)\(^{†}\) |
|-------------|------------|-------------------------------------------------|-----------------|-----------------|
| wt | Abz-STR\(_b\) GLPS-Y(3NO\(_2\)) | 5.13 (\(\pm 0.58\)) | 55.0 (\(\pm 6.4\)) | 1.6 (\(\pm 0.2\)) |
| R\(_b28\)A | Abz-STR\(_b\) GLPS-Y(3NO\(_2\)) | 6.9 (\(\pm 0.6\)) | 13.0 (\(\pm 2.4\)) | 0.2 (\(\pm 0.02\)) |
| R\(_b28\)K | Abz-STR\(_b\) GLPS-Y(3NO\(_2\)) | 29.3 (\(\pm 3.5\)) | 42.0 (\(\pm 4.0\)) | 0.9 (\(\pm 0.1\)) |
| N\(_b29\)A | Abz-STR\(_b\) GLPS-Y(3NO\(_2\)) | NA\(^{§}\) | – | – |
| N\(_b29\)Q | Abz-STR\(_b\) GLPS-Y(3NO\(_2\)) | NA\(^{§}\) | – | – |
| N\(_b29\)D | Abz-STR\(_b\) GLPS-Y(3NO\(_2\)) | \(\sim 21\)\(^{||}\) | ND\(^{||}\) | \(\sim 0.06\)\(^{||}\) |
| G\(_b30\)A | Abz-STR\(_b\) ALPS-Y(3NO\(_2\)) | 51.0 (\(\pm 2.0\)) | 29.0 (\(\pm 1.4\)) | 1.6 (\(\pm 0.07\)) |
| G\(_b30\)D | Abz-STR\(_b\) GLPS-Y(3NO\(_2\)) | 35.5 (\(\pm 2.7\)) | 31.4 (\(\pm 2.5\)) | 1.1 (\(\pm 0.08\)) |
| L\(_b28\)A | Abz-STR\(_b\) GLPS-Y(3NO\(_2\)) | NA\(^{§}\) | – | ND\(^{§}\) |
| L\(_b28\)I | Abz-STR\(_b\) GLPS-Y(3NO\(_2\)) | ND\(^{§}\) | ND\(^{§}\) | ND\(^{§}\) |

*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\(_5\)-Asp\(_{22}\) native Cys\(_{26}\) is substituted with Ser to avoid unpaired Cys residues.

\(^{§}\)W28, 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.

\(^{||}\)No activity detected under the conditions tested (up to 80 \(\mu\text{M}\) substrate; up to 6 h incubation).

\(^{\sim\sim}\)Low \(V_{\text{max}}\) preclude accurate estimation of kinetic parameters.
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 rOoAEP1b 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 OaAEP1b. 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 rOoAEP1b-mediated processing of modified kB1 and kB2 precursors over time (Fig. 4). When presented within an IQF peptide, the Leu₃₁Ala substrate analogue was not hydrolysed by rOoAEP1b (Table 1). However, the same substitution within the kB1 precursor did not preclude cyclization (Fig. 4b). 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.

**Water is excluded during cyclization.** An alternative ligation mechanism, distinct from transpeptidation, was recently proposed for huLEG31. 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 rOoAEP1b, reactions were carried out in the presence of 18O-labelled water and the products were analysed by high-resolution MS. An isotopic shift consistent with the incorporation of 18O 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 rOoAEP1b (Fig. 5a).
rOaAEP1b can cyclise an unrelated peptide. We also investigated cyclization of other substrates structurally unrelated to cyclotides by rOaAEP1b, 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, rOaAEP1b processing produced cyclic peptide (Supplementary Fig. 10).

**Figure 4** | Modified linear kB1 and kB2 precursors are cyclized at different rates. MALDI MS spectra of (a) kB1wt, (b) kB1L31A, (c) kB16xS, (d) kB26xS and (e) kB1acetyl cyclotide precursors at 1, 5 and 22 h post-enzyme addition. Data are representative of three technical replicates. * denotes rOaAEP1b cleavage site. Observed monoisotopic masses (Da; [M + H]+) 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) kB1acetyl and (b) kB1wt linear precursors in the presence and absence of 18O-labelled water. An isotope shift indicative of 18O incorporation only occurs during hydrolysis. Observed masses of two isotopic peaks (Da; [M + H]+) are indicated. Data are representative of three technical replicates.

rOaAEP1b can cyclise an unrelated peptide. We also investigated cyclization of other substrates structurally unrelated to cyclotides by rOaAEP1b, 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, rOaAEP1b processing produced cyclic peptide (Supplementary Fig. 10). Evidence of an additional, linear, rOaAEP1b-generated cleavage product was only observed for the R1 variant without any N-terminal flanking residues (Fig. 6d; Supplementary Fig. 10c).

| 1h |
|---|
| 3,518.7 |
| 2,891.2 |

| 5h |
|---|
| 3,476.9 |
| 2,810.5 |

| 22h |
|---|
| 3,428.9 |
| 2,864.5 |
| 3,560.9 |

| G1 LPVCGETCVGGTCTNPCTGCTCSPVCTRNGLPSLA kB1wt |
|---|
| G1 LPVCGETCVGGTCTNPCTGCTCSPVCTRNAGPSLA kB1L31A |
| G1 LPVSGGETSVGGTSTNPCTGCTCSPVCTRNGLPSLA kB16xS |
| G1 LPVSGGETSVGGTSTNPCTGCTCSPVCTRNGLPSLA kB1acetyl |
| AcG1 LPVCGETCVGGTCTNPCTGCTCSPVCTRNGLPSLA kB1acetyl |

**Figure 4** | Modified linear kB1 and kB2 precursors are cyclized at different rates. MALDI MS spectra of (a) kB1wt, (b) kB1L31A, (c) kB16xS, (d) kB26xS and (e) kB1acetyl cyclotide precursors at 1, 5 and 22 h post-enzyme addition. Data are representative of three technical replicates. * denotes rOaAEP1b cleavage site. Observed monoisotopic masses (Da; [M + H]+) 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) kB1acetyl and (b) kB1wt linear precursors in the presence and absence of 18O-labelled water. An isotope shift indicative of 18O incorporation only occurs during hydrolysis. Observed masses of two isotopic peaks (Da; [M + H]+) are indicated. Data are representative of three technical replicates.
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, rOaAEP1b 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 AEP34, butelase 1 (ref. 5) and human legumain28,35 (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 rhuLEG28,35. 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 rOaAEP1b 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). rOaAEP1b could also hydrolyse precursors lacking a free N-terminal amine to produce linear products, albeit at a slower rate. Although it is unknown if...
**Methods**

**Peptide substrates and inhibitors.** IQF peptides containing an N-terminal o-aminobenzoic acid (Abz) group and a C-terminal 3-nitrotrotyrosine (Y[3NO2])
were synthesized by Genscript at > 90% purity. Control IQP peptides representing the predicted cleavage products of the wt peptide (Abz-STRN; GLPS-Y(NCO)) were also synthesized by Genscript at >90% purity. All IQP peptides were solubilized in 25% (v/v) acetonitrile water. The fluorogenic peptide substrate Z-AEP-NH2 (where AEP, Aspartyl glutathione peptidase) was 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, 30% isopropanol, 2 mM reduced glutathione, 1 mM oxidized glutathione, pH 8.2) for 3 days.46 The products were isolated by RP-HPLC at >95% purity and characterized by high-resolution MS and NMR spectroscopy. Peptides KB135 and KB245, 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. 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 1:2,000 vol isopropanol. The pellets were dissolved in 0.5 ml water and RNA was precipitated overnight at 4°C. The RNA was isolated by RP-HPLC at >95% purity and characterized by high-resolution MS and NMR spectroscopy. Peptides KB135 and KB245, 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.

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 CAT GCC GCG G3’-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 TTC ATG GAA AGC-3’; OaAEP2-R, 5’-TTA TGC ACT GAA TCC TTT ATG GAG G3’-3’; OaAEP3-R 5’-TTA TGC ACT GAA TCC TTC ATC G-3’-3’) were designed with the aid of Primer 3 (ref. 46). Each primer set successfully amplified an AEP sequence.

Identification of the auto-processing sites of rOaAEP1b. Aliquots of rOaAEP1b (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 each protease. Aliquots of rOaAEP1b (5 μl) were added to a 1 ml Tris-HCl buffer (50 mM Tris-HCl, pH 8.0) and the mixture was incubated for 5 h at 37°C. Protein precipitation after 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 in Genbank (accession codes: OaAEP1b against both internally quenched and other fluorescent peptides, rOaAEP1b). 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−1 for reversed-phase separation on a C18 Zorbax column (Agilent 300SB-C18, 3.5 μm particle, 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.

Assaying protease activity against fluorescent peptides. To assay activity of rOaAEP1 against both internally quenched and other fluorescent peptides, substrate and enzyme were diluted as appropriate in activity buffer (50 mM sodium acetate, 0.1 mg ml−1 BSA, 1 mM EDTA, 50 mM NaCl, 1 mM β-galactosidase, 0.4 mM 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 lysodeoxycholic acid (hen egg white; 0.4 μg ml−1). DNase (bovine pancreas; 40 μg ml−1) and MgCl2 (0.4 μM) were also added. Cellular debris was removed by centrifugation and the lysate was stored at ~80°C until required.

Recombinant expression of O. affinis AEP1b, (OaAEP1b). Initial trials to produce active O. affinis AEP1b, based on predicted N- and C-terminal processing sites (residues D135-245) were unsuccessful and subsequent expression attempts incorporated both N- and C-terminal polytags. DNA encoding full-length O. affinis AEP1b, without the putative signaling domain (residues A2-P474) was inserted into the pHIEU vector57 to give a Hist-ubiquitin-OaAEP1b fusion protein construct. (Culture supernatant was injected into T7 shuffler cell line (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 lysodeoxycholic acid (hen egg white; 0.4 μg ml−1). DNase (bovine pancreas; 40 μg ml−1) and MgCl2 (0.4 μM) were also added. Cell debris was removed by centrifugation and the lysate was stored at ~80°C until required.

To determine the kinetics of rOaAEP1b activity against IQP 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−1. It was not possible to precisely determine the concentration of active enzyme due to the opaqueness of the preparation and the absence of an inhibitor appropriate for active site titration. However, a conservative turnover rate (kcat) 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\textsuperscript{15}. This was achieved as described previously\textsuperscript{15}. The output generated by the fluorescent hydrolysis product (Abs-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, rOAeEP\textsubscript{1}(4.4 μg ml\textsuperscript{-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 μM). Enzyme activity against the wt IQF peptide was then assayed as described above.

**Cyclization assay.** Linear target peptides (280 μM) were incubated with rOAeEP\textsubscript{1} (12 μg ml\textsuperscript{-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, RT derivatives processed by rOAeEP\textsubscript{1}, were subsequently digested with endoGlu-C (25 μg ml\textsuperscript{-1}) in reaction buffer (10 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 %\textsuperscript{2}H\textsubscript{2}O) was used in place of unlabelled water. Linear target peptides (70 μM) were incubated with rOAeEP\textsubscript{1} (0.5–0.75 μg ml\textsuperscript{-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\textsuperscript{2}O concentration in the assay was 81%.

**M5 to track cyclization of linear peptides.** Cyclization of linear target peptides was monitored by MALDI MS. The reaction mixture (10–20 μl) was desalted using C\textsubscript{18} zip tips and eluted in 4 mM sodium acetate, 50 mM NaCl, 1 mM EDTA, pH 5). A volume of 700 μl 1:4 mixed with the MALDI matrix, spotted onto a MALDI plate and analysed by an Ultraflex III (TOF/TOF (Bruker) in positive reflector mode.

To investigate if the slowed rate of cyclization was 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**

1. Sheldon, P. S., Keen, J. N. & Bowles, D. J. Post-translational peptide bond formation during cononavalin A processing in vitro. Biochem. J. 320, 865–870 (1996).
2. 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).
3. 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).
4. 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).
5. Nguyen, G. K. T. et al. Butelase 1 is an Axx-specific lissage enabling peptide macrolization and synthesis. Nat. Chem. Biol. 10, 732–738 (2014).
6. Luo, H. et al. Peptide macrolization catalyzed by a prolyl oligopeptidase involved in γ-aminitin biosynthesis. Chem. Biol. 21, 1610–1617 (2014).
7. Lee, J., McIntosh, J., Hathaway, R. J. & Schmidt, E. W. Using marine natural products to discover a protease that catalyzes peptide macrolization of diverse substrates. J. Am. Chem. Soc. 131, 2122–2124 (2009).
8. Gillon, A. D. et al. Biosynthesis of circular proteins in plants. Plant J. 53, 505–515 (2008).
9. Sasaki, I. et al. An asparaginyl endopeptidase mediates in vivo protein backbone cyclization. J. Biol. Chem. 282, 29721–29728 (2007).
10. Bernath-Levin, K. et al. Peptide macrolization by a bifunctional endopeptidase. Chem. Biol. 22, 571–582 (2015).
11. Jennings, C., West, J., Waine, C., Craik, D. & Anderson, M. Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from \textit{Oenothera affinis}. Proc. Natl Acad. Sci. USA 98, 10614–10619 (2001).
12. Plan, M. R., Sasaki, I., Cagauan, A. G. & Craik, D. J. Backbone cyclised peptides from plants show molluscicidal activity against the rice pest Poromaea canaliculata (golden apple snail). J. Agric. Food Chem. 56, 5237–5241 (2008).
13. 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).
14. Colgrave, M. L. et al. Anthelmintic activity of cyclotides: In vitro studies with canine and human hookworms. Acta Trop. 109, 163–166 (2009).
15. 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).
16. 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).
17. 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).
18. Chan, L. Y. et al. Cyclization of the antimicrobial peptide gomesin with native \textit{Oldenlandia affinis} Sortase A. \textit{ChemBioChem} 14, 617–624 (2013).
19. Craik, D. J. Host-defense activities of cyclotides. Toxins (Basel) 4, 139–156 (2012).
20. Oldenberger, K. et al. Backbone cyclization of a recombinant cystine-knot peptide by engineered Sortase A. FERS Lett. 588, 4487–4494 (2016).
21. 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).
22. Mylne, J. S. et al. Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. Nat. Chem. Biol. 7, 257–259 (2011).
23. Hiraisha, N., Nishimura, M. & Hara-Nishimura, I. Expression and activation of the vacuolar processing enzyme in \textit{Saccharomyces cerevisiae}. Plant J. 12, 819–829 (1997).
24. Hiraiwa, N., Nishimura, M. & Harra-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 cytoste endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. Arch. Biochem. Biophys. 303, 208–213 (1993).

27. Hatsuigui, 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. Koehne, 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, 345–357 (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 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/