Papain-like cysteine proteases prepare plant cyclic peptide precursors for cyclization

Fabian B. H. Rehm[a], Mark A. Jackson[a], Ewout De Geyter[a], Kuok Yap[a], Edward K. Gilding[a], Thomas Durek[a,1], and David J. Craik[a,1]

[a]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

Edited by Stephen J. Benkovic, The Pennsylvania State University, University Park, PA, and approved March 12, 2019 (received for review February 5, 2019)

Cyclotides are plant defense peptides that have been extensively investigated for pharmaceutical and agricultural applications, but key details of their posttranslational biosynthesis have remained elusive. Asparaginyl endopeptidases are crucial in the final stage of the head-to-tail cyclization reaction, but the enzyme(s) involved in the prerequisite steps of N-terminal proteolytic release were unknown until now. Here we use activity-guided fractionation to identify specific members of papain-like cysteine proteases involved in the N-terminal cleavage of cyclotide precursors. Through both characterization of recombinantly produced enzymes and *in planta* peptide cyclization assays, we define the molecular basis of the substrate requirements of these enzymes, including the prototypic member, here termed kalatase A. The findings reported here will pave the way for improving the efficiency of plant biofactory approaches for heterologous production of cyclotide analogs of therapeutic or agricultural value.

plant defensins | cyclotide | kalata B1 | papain-like cysteine protease | RiPP biosynthesis

Ribosomally synthesized and posttranslationally modified peptides are natural products of diverse structure and bioactivity (1). Among them are head-to-tail cyclized peptides and proteins in which the N and C termini are joined via a peptide bond, resulting in enhanced conformational rigidity and metabolic stability over acyclic counterparts (2). Whereas large (35–78 aa) cyclic bacteriocins and pilins are produced by bacteria (3, 4) and cyanobacteria produce smaller (6–11 aa) cyanobactins (5), many higher organisms produce peptide macrocycles that are further constrained by one or more disulfide bonds (6, 7).

Plant-derived cyclotides are the largest such family of disulfide-rich cyclic peptides (8). Their defining structural feature is a cyclic cystine knot in which two disulfide bonds and intervening backbone residues form a ring that is threaded by a third disulfide bond, with the disulfide knot encased within a head-to-tail cyclic peptide backbone. These structural attributes engender cyclotides with exceptional chemical and proteolytic stability (9). The inherent insecticidal, nematicidal, and molluscicidal activities of cyclotides suggest a natural role in plant host defense, highlighting their potential for agricultural applications (10–12), and their exceptional stability has seen cyclotides used for the stabilization of pharmaceutically relevant epitopes (13, 14).

Efforts to cost-effectively produce natural or modified cyclotides in transgenic plants require a detailed understanding of their biosynthesis (15). Cyclotides are produced as gene-encoded precursors comprising an N-terminal signal peptide and propeptides that flank one or multiple cyclotide domains (15) (Fig. 1). In the first step of the prepro-peptide maturation process, cyclotide precursors enter the secretory pathway via the endoplasmic reticulum, where folding of the cyclotide precursor and disulfide bond formation is posited to occur (16). Subsequent proteolytic processing and head-to-tail cyclization are thought to occur in the vacuole (17) to yield mature cyclotides.

While there is uncertainty surrounding the functional genetics affecting maturation, asparaginyl endopeptidases (AEPs) have been convincingly implicated in the final steps of cyclotide processing. Specifically, these cysteine proteases catalyze C-terminal cyclotide processing and concomitant backbone cyclization (18–22). The requirement for AEP-mediated cyclization appears to be conserved across all cyclotide-producing plant families and coincides with the conservation of an asparagine (or aspartic acid) residue at the C terminus of cyclotide precursors. In contrast, details of the proteolytic processing event N-terminally of the cyclotide domain is lacking (20) and presents a major gap in our understanding of cyclotide biosynthesis. As N-terminal processing is considered a prerequisite for AEP-mediated backbone cyclization, elucidation of the enzymes involved will enable strategies to optimize the yield of cyclic peptides in plant biofactory applications.

In this study, we identify specific members of the papain-like cysteine protease family as key cyclotide N-terminal processing enzymes. Through a combination of recombinant enzyme characterization and *in planta* functional assays, we define the substrate specificities for these enzymes and determine the critical residues required for efficient cyclotide precursor processing. Our data demonstrate that efficient N-terminal processing is critical for the biosynthetic production of cyclotides in plants.

Results

Activity-Guided Purification of N-Terminal Processing Activity. We chose the prototypical cyclotide kalata B1 (kB1) as a model peptide to study N-terminal processing enzymes. To specifically assay for N-terminal processing activity, we designed a truncated cyclotide precursor, LQLK-kB1, comprising a short sequence (LQLK) from the flanking N-terminal repeat (NTR) linked to the oxidatively folded kB1 domain. This 33-aa peptide lacked a

Significance

Ribosomally synthesized and post-translationally modified peptides are natural products that hold great promise for a range of medical and biotechnological applications. However, the cost-effective heterologous production of these peptides is hampered by a poor understanding of their biosynthesis. Cyclotides, plant-derived, disulfide-knotted, head-to-tail cyclic peptides, exhibit exceptional stability and great amenability for amino acid substitutions and insertions. Although much effort has been invested toward understanding cyclotide biosynthesis, details of the key proteolytic step before cyclization remained elusive for two decades. We used an activity-guided approach to discover the enzymes involved. Our characterization of these enzymes will enable efficient approaches for heterologous cyclotide production.

Author contributions: T.D. and D.J.C. designed research; F.B.H.R., M.A.J., E.D.G., K.Y., E.K.G., and T.D. performed research; F.B.H.R., M.A.J., E.D.G., K.Y., E.K.G., T.D., and D.J.C. analyzed data; and F.B.H.R., M.A.J., E.K.G., T.D., and D.J.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: t.durek@imb.uq.edu.au or d.craik@imb.uq.edu.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1901807116/-/DCSupplemental.

Published online April 3, 2019.

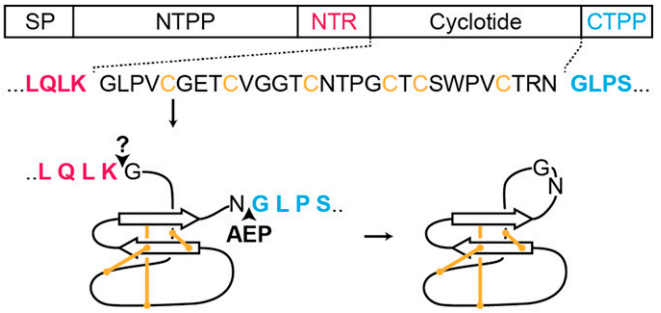

Fig. 1. Schematic representation of kB1 maturation. The gene-encoded precursor peptide is proteolytically processed at both sides of the core peptide. AEPs catalyze C-terminal processing and cyclization, but details of the prerequisite N-terminal processing have remained elusive. CTPP, C-terminal propeptide; NTPP, N-terminal propeptide; NTR, N-terminal repeat sequence; SP, endoplasmic reticulum-targeting signal peptide.

C-terminal propeptide (required for cyclization), to minimize interference with AEP-related activities (20).

As it is known that the addition of a single specialized AEP isoform is sufficient for conversion of the noncyclotide-producing plant *Nicotiana benthamiana* into an efficient cyclotide producer (21), we reasoned that this plant harbors one or more enzymes capable of N-terminally processing cyclotide precursors. On incubation of the LQLK-kB1 probe with crude *N. benthamiana* leaf extract, we detected three processed products—LK-kB1, K-kB1, and acyclic kB1 (*SI Appendix*, Fig. S1)—via MS, with the final and dominant product being acyclic kB1 with the N-terminal LQLK sequence removed. Once fully processed acyclic kB1 was generated, it remained stable to further proteolysis during our assay span (18 h).

To discriminate the enzyme(s) responsible for the observed proteolytic processing, we fractionated the crude *N. benthamiana* leaf extract via consecutive $(NH_4)_2SO_4$ precipitation steps. Ultimately, this approach facilitated the separation of the K-kB1–generating proteolytic event from the LK-kB1– and acyclic kB1-generating events (Fig. 2*A* and *SI Appendix*, Fig. S1). Additional purification steps, consisting of anion exchange followed by hydrophobic interaction chromatography (Fig. 2*B*), did not produce further variation in the generated peptide profiles and yielded a fraction highly enriched in N-terminal processing activity that we subjected to in-solution tryptic digestion. The amino acid sequences of the resultant peptide fragments, as determined by LC-MS/MS analysis, were used to scan for matches in the available *N. benthamiana* transcriptomes (23, 24). This analysis revealed the presence of a range of highly abundant leaf proteins (e.g., RuBisCo) that had remained throughout the purification, along with a single protease, NbCysP6, which was unambiguously identified based on two peptide segments (*SI Appendix*, Fig. S2).

Recombinant NbCysP6 Generates Acyclic kB1 from Synthetic LQLK-kB1. NbCysP6 is a cathepsin L-like papain-like cysteine protease (PLCP) from the RD21 subclass (25, 26). Like most members of this subclass (25), the domain architecture of NbCysP6 consists of an N-terminal signal peptide, an autoinhibitory I29 domain, a C1 protease domain, a proline-rich segment, and a cysteine-rich C-terminal granulin domain. Like AEPs (19, 20), PLCPs are produced as inactive zymogens that autocatalytically mature in low-pH environments, such as the plant vacuole, via cleavage at sites flanking the C1 protease domain.

To assess the cyclotide N-terminal processing activity of NbCysP6, we recombinantly produced the enzyme in *Escherichia coli*. Mature NbCysP6 was purified to electrophoretic purity (Fig. 2*C*), and the identity of the enzyme was confirmed via in-gel and in-solution tryptic digests followed by LC-MS/MS analysis of the resultant peptide fragments (*SI Appendix*, Fig. S3). Incubation of the LQLK-kB1 probe with active NbCysP6 generated peptides matching those generated by the *N. benthamiana* extract (Fig. 2 *A* and *D*) confirming that the observed N-terminal processing activity of the extract is attributable, at least in part, to NbCysP6.

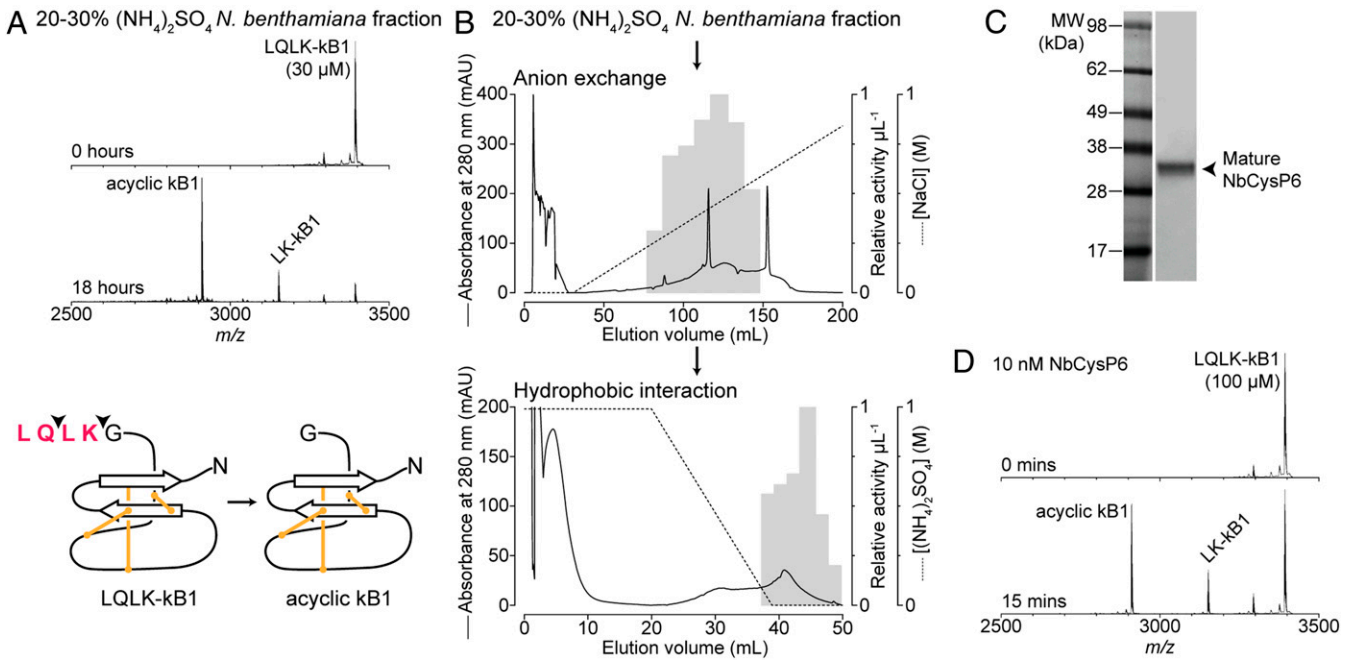

Fig. 2. MS characterization of N-terminal processing activity and its purification. (*A*) MALDI-TOF MS spectra before and after an 18-h incubation of the synthetic peptide probe LQLK-kB1 with the *N. benthamiana* fraction precipitating between 20% and 30% $(NH_4)_2SO_4$ saturation in 50 mM sodium phosphate buffer, pH 6.5, at 37 °C. (*B*) Chromatograms from the purification of N-terminal processing activity, as shown in *A*, from *N. benthamiana* extract. The shaded bars indicate the relative activities of fractions taken for subsequent steps. LC-MS/MS protein identification after hydrophobic interaction chromatography identified the protease NbCysP6. (*C*) Instant blue-stained SDS/PAGE gel of mature recombinant NbCysP6. The migration positions of molecular weight standards are indicated. (*D*) Representative MS spectra before and after incubation of 100 μM LQLK-kB1 with 10 nM NbCysP6 in sodium acetate buffer, pH 5, at 25 °C for 15 min.

Identification and Recombinant Production of Homologous Enzymes. To extend our investigation to the native kB1-producing plant *Oldenlandia affinis*, we assembled a transcriptome from *O. affinis* and interrogated it for PLCPs exhibiting homology to NbCysP6. We identified three full-length contig sequences representing proteases belonging to the RD21 subclass, termed OaRD21A (Fig. 3), OaRD21B, and OaRD21C (*SI Appendix*, Table S2), which share at least 62% identity at the amino acid level, as determined by pairwise alignments. Compared with NbCysP6, 60–71% identity (i.e., 77–81% similarity) was observed (*SI Appendix*, Fig. S4).

To assess whether *O. affinis* PLCPs could perform cyclotide N-terminal processing, we recombinantly produced OaRD21A in *E. coli*. Like NbCysP6 but unlike the previously reported *N. benthamiana* PLCPs NbCysP7 (26) and NbCP14 (27) included for comparison, OaRD21A could be purified only in the mature form, presumably due to rapid autoactivation during purification. The identities of all enzymes were confirmed via in-gel and in-solution tryptic and endoGlu-C digests and subsequent LC-MS/MS analyses (*SI Appendix*, Fig. S3). Peptide fragments that arose from noncanonical cleavage events were used to map autoprocessing sites (Fig. 3*B* and *SI Appendix*, Fig. S3). Like NbCysP6, OaRD21A, as well as NbCysP7 and NbCP14, generated acyclic kB1 on incubation with LQLK-kB1 (*SI Appendix*, Fig. S6). Whereas acyclic kB1 remained stable to further proteolysis, reduced kB1 was rapidly degraded (*SI Appendix*, Fig. S7), suggesting disulfide bond formation to be prerequisite for specific PLCP-mediated processing.

Substrate Specificity. To assess the sequence requirements for N-terminal processing of cyclotides as catalyzed by the recombinantly produced enzymes, a set of alanine-scanned internally quenched fluorescent (IQF) peptides was designed based on the native kB1 N-terminal cleavage site (Gln_{-3}-Leu_{-2}-Lys_{-1}-Gly_1) (*SI Appendix*, Table S1). Whereas mutations to Ala residues in positions −3 and 1 were well tolerated in all cases, mutation of the Leu_{-2} residue resulted in a substantial drop in catalytic efficiency (k_{cat}/K_M). Mutation of Lys_{-1} to Ala also resulted in a significant reduction in k_{cat}/K_M, suggesting that along with a stringent P2 requirement, these enzymes also need a compatible P1 residue to achieve maximum activity. Notably, NbCP14 displayed reduced activity relative to the other tested enzymes, consistent with what was previously reported for this enzyme using dipeptide substrates (27), whereas OaRD21A, NbCysP6, and NbCysP7 had comparable activity. Importantly, an analysis of the putative N-terminal processing sites of other known cyclotide precursors revealed substantial conservation of a hydrophobic aliphatic residue at the P2 position (*SI Appendix*, Fig. S8). Furthermore, a lack of conservation at the other positions aligns with a lack of residue requirements for PLCP activity at those positions.

***In Planta* Mutagenesis of the N-Terminal Processing Site.** We next assessed whether the P2 requirement observed for our recombinantly produced PLCPs extends to efficient cyclotide maturation *in planta*. For this, we used a transient expression approach in *N. benthamiana* to monitor the processing of a series of kB1-encoding Oak1 mutants that had residues in the NTR processing site mutated to Ala (Fig. 4). We simplified this analysis by using a previously characterized transgenic *N. benthamiana* line that constitutively expresses an AEP (21), allowing any differences in cyclotide maturation to be attributed to the changes that we made to the cyclotide precursor sequence. MALDI-TOF MS-based analysis of the resultant peptide extracts (Fig. 4) revealed a substantial increase in kB1 misprocessing when Leu_{-2} was mutated. Importantly, we observed the emergence of putative cyclic (−18 Da) AK-kB1, suggesting miscyclization due to imprecise proteolytic release of the kB1 N terminus. Similarly, mutations of all five residues in the NTR sequence gave rise to putative cyclic AAAA-kB1 and AAA-kB1, aligning with previous reports of broad AEP substrate specificity for the N-terminal amino acids (19, 20, 28). These results suggest the N-terminal sequence requirements for efficient *in planta* production of kB1

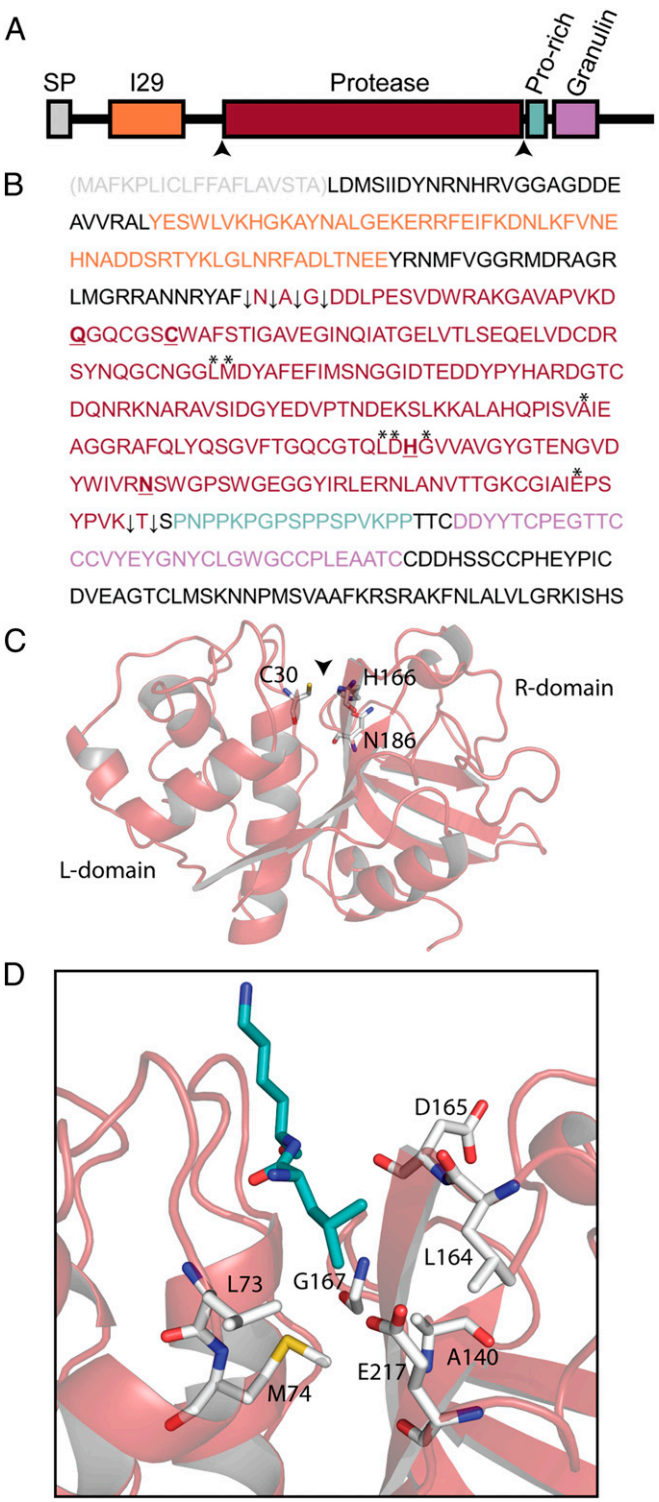

Fig. 3. Domain organization and homology model of active OaRD21A. (*A*) Domain organization with arrows indicating the autoprocessing sites. SP, signal peptide; I29, inhibitor I29 domain; Prorich, proline-rich region. (*B*) Sequence of OaRD21A using the color scheme as in *A*. Active site residues are in bold type and underscored, arrows indicate the identified autoprocessing sites, and asterisks indicate residues that form the S2-binding pocket. (*C*) Homology model of active OaRD21A with catalytic triad residues shown in stick representation; the arrow indicates the active site cleft between the L and R domains. (*D*) View of the S2 binding pocket with a short substrate (cyan) fragment bound. The S2 subsite-defining residues are in stick representation.

to align with the P2 requirement for efficient PLCP-mediated cleavage that we determined with the IQF substrates in vitro.

Assessment of N-Terminal Processing Activity on LQLK-kB1. To investigate how efficiently OaRD21A, as well as NbCysP6, NbCysP7, and NbCP14, could perform N-terminal proteolysis, we quantified acyclic kB1 produced from LQLK-kB1 by these enzymes via MS. Plots of initial enzyme velocities against LQLK-kB1 concentration revealed an initial Michaelis–Menten-like increase in enzyme velocity but a subsequent decrease at higher substrate concentrations, suggesting substrate inhibition (*SI Appendix*, Fig. S9). We determined the kinetic parameters accordingly, revealing that NbCysP6 and NbCysP7 had comparable catalytic efficiencies (k_{cat}/K_M = 13 and 12 mM^{-1} s^{-1}, respectively), whereas OaRD21A displayed marginally enhanced activity (21 mM^{-1} s^{-1}), and NbCP14 was substantially less catalytically efficient (1.4 mM^{-1} s^{-1}), aligning with what we observed for this enzyme using IQF substrates (*SI Appendix*, Table S1). The K_i values that we determined from these plots were more comparable, ranging from 140 to 80 μM.

Reengineering OaRD21A Substrate Specificity. Having established the importance of the P2 Leu specificity for cyclotide processing both in vitro and in our *in planta* assay, we sought to understand the molecular basis on the enzyme level and thus created a homology model of active OaRD21A using the crystal structure of a papain-type cysteine endopeptidase in complex with a chloromethylketone inhibitor (Protein Data Bank ID code 1S4V) as a template (Fig. 3*C*). The papain-like fold (29) consists of two distinct domains divided by the active site cleft. The left (L) domain consists of mostly α-helical segments, and the right (R) domain consists largely of an antiparallel, barrel-like arrangement of β-strands. As is typical of PLCPs and consistent with our observed lack of residue requirements at positions other than P2 for cleavage, only the S2 subsite is a well-defined pocket. Situated at the two-domain interface, it is formed by residues Leu73 and Met74 from the L domain and Ala140, Leu164, Asp165, Gly167, and Glu217 from the R domain (Fig. 3*D*). The sidechains of Leu73 and Glu217, positioned at the side and base of the pocket, respectively, protrude into the binding pocket and constrain its size.

Compared with other well-characterized PLCPs, such as papain and human cathepsins B and K, which have a Tyr residue at the position equivalent to Leu73, OaRD21A has a wider binding pocket more similar to that of human cathepsin L (*SI Appendix*, Table S3). Thus, we hypothesized that although the pocket looks to accommodate P2 Leu well, perhaps larger hydrophobic residues could also be tolerated. Indeed, IQF peptides with a P2 Tyr and, to an approximate 10-fold lesser extent, Trp could be cleaved by OaRD21A, although much less effectively than the P2 Leu substrate (Table 1).

To test whether we could tune the activity of OaRD21A to differentially process these substrates by altering the size of the S2 pocket, we generated a series of OaRD21A mutants. In all cases, any alterations made to S2 subsite residues resulted in substantially decreased processing relative to the wild-type enzyme, but we were able to alter the efficiency of processing Tyr- and Trp-containing substrates relative to the Leu-containing substrate. OaRD21A Glu217Ala, aimed at increasing the depth of the pocket, showed an improved ability to cleave both QYKG and particularly QWKG relative to QLKG. In contrast, mutation of Glu217 to Leu resulted in an enzyme largely incapable of processing the Trp- and Tyr-containing substrates while retaining activity against the Leu-containing substrate. Interestingly, Leu73Tyr, aimed at decreasing the size of the S2-binding pocket, still showed an improved relative ability to process QYKG and QWKG, suggesting that the size constraints imposed by the Tyr residue are insufficient to prevent the access of larger substrates. However, the introduction of a bulkier Trp residue resulted in an enzyme incapable of processing the QWKG substrate while retaining activity against the Leu- and Tyr-containing substrates.

In Vitro Reconstitution of Cyclotide Precursor Proteolytic Processing and Cyclization. To see whether OaAEP1b and OaRD21A could cooperate in a single assay to produce cyclic kB1, we made the oxidatively folded peptide LQLK-kB1-GI. This substrate has both N- and C-terminal propeptide sequences and thus requires proteolytic processing at both ends with subsequent cyclization to produce mature, cyclic kB1. Similar to the N-terminal LQLK sequence, the C-terminal –GI is a shortened version of the full-length kB1 C-terminal propeptide (GLPSLAA) and simplifies chemical synthesis while retaining the ability to be processed by AEPs (*SI Appendix*, Fig. S10) (22). Incubation of LQLK-kB1-GI with OaAEP1b alone generated only C-terminally processed and miscyclized product, and incubation of OaRD21A alone generated N-terminally processed product (Fig. 5). As C-terminal processing before N-terminal processing precludes mature kB1 formation, a greater catalytic contribution from OaRD21A was necessary in the combined assay. Efficient conversion of LQLK-kB1-GI to mature kB1 was achieved at an OaRD21A:OaAEP1b molar ratio of 1:2. These observations suggest that *in planta*, N-terminal processing and AEP-mediated cyclization might be physically separated or may occur in an environment in which N-terminal proteolytic activity predominates over AEP activity.

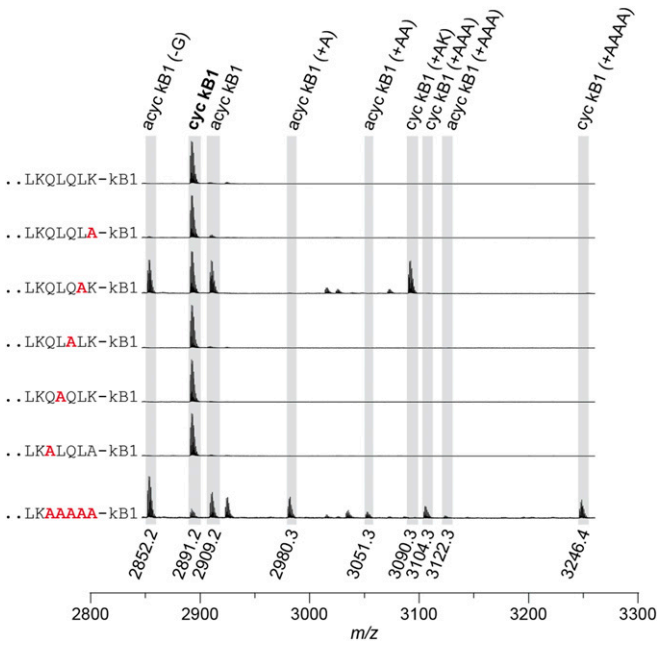

Fig. 4. *In planta* alanine scanning mutagenesis of the NTR processing site. MALDI-TOF MS spectra from *N. benthamiana* (stably expressing OaAEP1b) peptide extracts obtained at 6 d postinfiltration with Oak1 alanine-scanned mutants, as indicated. The observed products are denoted as acyclic (acyc) or cyclic (cyc).

Discussion

The involvement of AEPs in cyclotide biosynthesis was proposed when cyclotide precursor sequences were first reported (10). Despite extensive study of cyclotide biosynthesis (15), details of the prerequisite N-terminal proteolytic release of the cyclotide domain have remained elusive. Here, activity-guided fractionation enabled identification of the RD21-like PLCP NbCysP6, which in its pure, recombinant form, could efficiently generate acyclic kB1 from synthetic LQLK-kB1 precursor. In contrast to recently reported undesirable degradation of recombinant proteins by NbCysP6 (26), we find a favorable biosynthetic role for the enzyme when cyclotides are expressed in *N. benthamiana*. Our discovery of the ability to process cyclotide precursors by this class of PLCPs enabled the identification of OaRD21A from

Table 1. Kinetic parameters of IQF peptide cleavage by wild-type OaRD21A and Glu217Ala, Glu217Leu, Leu73Tyr, and Leu73Trp mutants

Enzyme	Substrate	K_M, μM, mean ± SEM*	k_{cat}, s^{-1}, mean ± SEM*	k_{cat}/K_M, M^{-1} s^{-1}, mean ± SEM*	QLKG:QYKG k_{cat}/K_M ratio	QLKG:QWKG k_{cat}/K_M ratio
OaRD21A	Abz-QLKG-Y(3NO_2)	3.3 ± 0.16	7.0 ± 0.070	2.1 ± 0.13 × 10^6	15:1	175:1
	Abz-QYKG-Y(3NO_2)	39 ± 10	5.5 ± 0.45	1.4 ± 0.48 × 10^5		
	Abz-QWKG-Y(3NO_2)	98 ± 34	1.2 ± 0.18	1.2 ± 0.6 × 10^4		
OaRD21A Glu217Ala	Abz-QLKG-Y(3NO_2)	210 ± 52	0.44 ± 0.056	2.1 ± 0.80 × 10^3	4.8:1	1.9:1
	Abz-QYKG-Y(3NO_2)	170 ± 46	0.072 ± 0.012	4.4 ± 2.0 × 10^2		
	Abz-QWKG-Y(3NO_2)	160 ± 50	0.18 ± 0.025	1.1 ± 0.49 × 10^3		
OaRD21A Glu217Leu	Abz-QLKG-Y(3NO_2)	320 ± 110	0.36 ± 0.079	1.1 ± 0.65 × 10^3	ND	ND
	Abz-QYKG-Y(3NO_2)	>1,000	ND	ND		
	Abz-QWKG-Y(3NO_2)	ND	ND	ND		
OaRD21A Leu73Tyr	Abz-QLKG-Y(3NO_2)	83 ± 34	0.36 ± 0.059	4.3 ± 2.4 × 10^3	3.1:1	3.6:1
	Abz-QYKG-Y(3NO_2)	140 ± 69	0.19 ± 0.045	1.4 ± 1.0 × 10^3		
	Abz-QWKG-Y(3NO_2)	150 ± 80	0.18 ± 0.046	1.2 ± 0.90 × 10^3		
OaRD21A Leu73Trp	Abz-QLKG-Y(3NO_2)	85 ± 31	0.21 ± 0.030	2.5 ± 1.2 × 10^3	7.6:1	ND
	Abz-QYKG-Y(3NO_2)	190 ± 52	0.061 ± 0.0089	3.3 ± 1.4 × 10^2		
	Abz-QWKG-Y(3NO_2)	ND	ND	ND		

ND, not determinable at tested enzyme and substrate concentrations.
*n = 3.

the kB1-producing plant *O. affinis*. To reflect its role in kalata biosynthesis, we propose the name kalatase A for OaRD21A.

In contrast to the conserved Asn or Asp residues at the cyclotide domain C terminus (18), a lack of amino acid conservation N-terminally of the cyclotide domains has inhibited discovery of the proteases involved, with the exception of a subset of precursors that have an N-terminal Asn residue and could be processed N- and C-terminally by AEPs (15). Here we show that the otherwise general lack of conservation N-terminally of the cyclotide domain aligns with a lack of sequence requirements for PLCP-mediated cleavage, with the critical exception of the P2 residue. The Leu residue at this position in the kB1 precursor sequence is essential for PLCP-mediated proteolysis in vitro. When the equivalent residue was mutated to Ala in a transgenically expressed kB1 precursor, we saw the emergence of misprocessed products due to imprecise N-terminal processing *in planta*, whereas mutations of other residues in the sequence had no such consequence.

The finding that Leu and other aliphatic hydrophobic residues are highly conserved at P2 in a large proportion of cyclotide precursors (*SI Appendix*, Fig. S8) supports a widespread role for PLCPs with such a P2 specificity in N-terminally processing cyclotide precursors. The role of the remaining residues in the N-terminal propeptides remains to be determined, but if they do not provide additional cleavage sites, perhaps they play a role in cyclotide folding or trafficking.

Our homology modeling revealed the organization of the S2-binding pocket, providing a molecular understanding for the specificity in cyclotide N-terminal processing. The residues that protrude into the binding pocket, Leu73 and Glu217, are key determinants of pocket size and thus of substrate accommodation. In several well-characterized PLCPs, the equivalently positioned residues (e.g., Tyr and Leu, respectively, in human cathepsin K, or Tyr and Glu in cathepsin B; *SI Appendix*, Table S3) have been implicated as major determinants for substrate processing (30, 31). Our experimental verification of the importance of Leu73 and Glu217 via their mutagenesis enabled us to tune the activity on substrates containing larger hydrophobic residues (Tyr or Trp) relative to the Leu-containing substrate. While the Glu217Ala and Leu73Trp mutants produced the expected correlation between pocket size and the accommodation of bulkier substrates, replacement of Glu217 with a Leu resulted in inability to process the Tyr- and Trp-containing substrates.

Interestingly, the crystal structure of cruzain, which also has a Glu at the equivalent position, bound to a Tyr-containing inhibitor (32) revealed the ability of the Glu sidechain to rotate and increase space in the binding pocket, which the shorter Leu might not be capable of. Similarly, rotation of the Tyr sidechain (about the Cα-Cβ bond) may explain why, unlike the Leu73Trp mutant, the Leu73Tyr mutant did not show substantially decreased activity on bulkier substrates. Notably, all the S2 subsite mutations that we introduced substantially reduced activity relative to the wild-type enzyme, suggesting that the S2 subsite residues cooperate to form an optimal substrate-binding pocket.

PLCPs, like AEPs, mature autocatalytically on reaching a low pH environment. This process involves excision of the mature

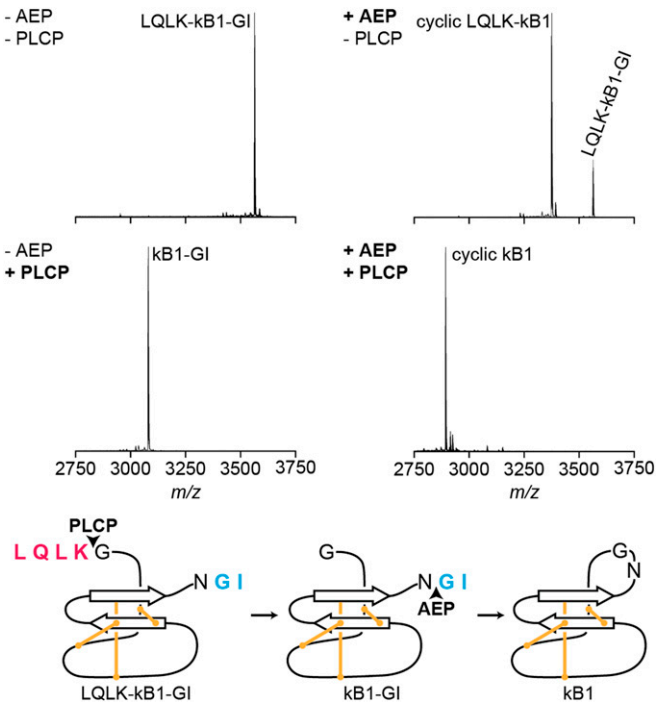

Fig. 5. Formation of mature kB1 from synthetic LQLK-kB1-GI catalyzed by OaRD21A and OaAEP1b. MALDI-TOF MS spectra of 20 μM LQLK-kB1-GI before and after treatment with 20 nM OaRD21A (PLCP) and/or 40 nM OaAEP1b (AEP) (as indicated) in 100 mM sodium phosphate buffer, pH 7, at 25 °C for 60 min.

protease domain via proteolytic cleavage. For several PLCP subclasses, such as RD21-like PLCPs (33), this maturation occurs in vacuoles. Given that cyclotide precursors enter the vacuole as directed by their N-terminal propeptides (17), a role for vacuolar PLCPs in propeptide release supports a controlled, stepwise maturation of cyclotides. Our finding that oxidatively folded but not reduced kB1 resisted PLCP-mediated proteolytic degradation (*SI Appendix*, Fig. S7) suggests that folding and disulfide bond formation of the cyclotide domains, thought to occur in the ER (16), are critical requirements for subsequent maturation steps.

Since AEP-catalyzed head-to-tail cyclization of cyclotides and C-terminal proteolysis occur concomitantly (19, 20), release of the cyclotide N terminus is a prerequisite for cyclotide maturation. Here we show that PLCP and AEP can cooperate in the same environment to mature kB1 precursors in vitro. Notably, we found no cyclization of the LK-kB1-GI intermediate, highlighting the ability of OaRD21A to effectively N-terminally prepare the substrate for correct maturation. While the necessary relative catalytic contributions from OaRD21A and OaAEP1b may differ when the full-length N- and C-terminal propeptides are present, a greater initial catalytic contribution from the PLCP would consistently be required for N-terminal proteolysis to occur first. Conceivably this could be the case in the plant vacuole, but such stepwise processing could also be facilitated by compartmentalization of the different proteolytic activities. However, the increased abundance of acyclic relative to cyclic kB1 that we observed when N-terminal processing was impaired *in planta* suggests that AEP-mediated C-terminal release preceded N-terminal processing, thus precluding cyclization, and is in support of these activities occurring in the same environment.

Interestingly, OaAEP1b incubated with synthetic LQLK-kB1-GI precursor in the absence of N-terminal processing activity resulted in the emergence of putative miscyclized LQLK-kB1. This is in agreement with the reported tolerance for a range of residues at the N terminus in AEP-mediated cyclization (20). Furthermore, we showed that such miscyclization can occur *in planta* due to incomplete N-terminal release, as facilitated by mutation of the P2 Leu residue to an Ala. These findings highlight the importance of N-terminal processing before the AEP-mediated step and suggest that in native cyclotide-producing plants, in which such miscyclized products have not been detected, proteolytic processing at the N terminus strictly occurs before cyclization or an as-yet unknown mechanism for the degradation of miscyclized products has developed.

Kinetic determination of acyclic kB1 formation from the LQLK-kB1 precursor revealed a decrease in enzyme velocities at the higher substrate concentrations. This suggests substrate inhibition, which we found to occur with comparable K_i values ranging from 140 μM to 80 μM, and may indicate a regulatory role. Substrate inhibition of enzymes with similar K_i values has been shown to result in steady-state product formation despite fluctuations in substrate concentration (34). Whether this is relevant to cyclotide biosynthesis and how such inhibition might occur remain to be determined.

In conclusion, this study greatly augments our current model of cyclotide biosynthesis by demonstrating the involvement of PLCPs. Not only is this a major step in understanding the biosynthesis of a class of cyclic peptides of great interest for pharmaceutical and agricultural applications, but it also may provide the first link between endogenous substrate processing by vacuolar PLCPs, a class of enzymes known to be widely involved in plant immunity (35), to the generation of a plant defense response.

Methods

Detailed descriptions of the procedures for peptide synthesis, activity-guided purification, *O. affinis* transcriptome analysis, OaRD21A cloning, recombinant protein production in *E. coli*, homology modeling, and transient expression in *N. benthamiana* are provided in *SI Appendix*.

ACKNOWLEDGMENTS. We thank L. Mach (University of Natural Resources and Life Sciences, Vienna) for DNA encoding NbCysP6, NbCysP7, and NbCP14; A. Jones for expertise and access to MS equipment; and G. Lomonossoff (John Innes Centre) for pEAQ vectors. This work was supported by the Australian Research Council (Grants DP150100443 and FL150100146) and a Clive and Vera Ramaciotti Biomedical Research Award.

1. Arnison PG, et al. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30:108–160.
2. Craik DJ (2006) Seamless proteins tie up their loose ends. *Science* 311:1563–1564.
3. Jack RW, Tagg JR, Ray B (1995) Bacteriocins of Gram-positive bacteria. *Microbiol Rev* 59:171–200.
4. Eisenbrandt R, et al. (1999) Conjugative pili of IncP plasmids and the Ti plasmid T pilus are composed of cyclic subunits. *J Biol Chem* 274:22548–22555.
5. Sivonen K, Leikoski N, Fewer DP, Jokela J (2010) Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Appl Microbiol Biotechnol* 86:1213–1225.
6. Göransson U, Burman R, Gunasekera S, Strömstedt AA, Rosengren KJ (2012) Circular proteins from plants and fungi. *J Biol Chem* 287:27001–27006.
7. Conibear AC, Craik DJ (2014) The chemistry and biology of theta defensins. *Angew Chem Int Ed Engl* 53:10612–10623.
8. Gruber CW, et al. (2008) Distribution and evolution of circular miniproteins in flowering plants. *Plant Cell* 20:2471–2483.
9. Craik DJ, Daly NL, Bond T, Waine C (1999) Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J Mol Biol* 294:1327–1336.
10. Jennings C, West J, Waine C, Craik D, Anderson M (2001) Biosynthesis and insecticidal properties of plant cyclotides: The cyclic knotted proteins from *Oldenlandia affinis*. *Proc Natl Acad Sci USA* 98:10614–10619.
11. Colgrave ML, et al. (2008) Cyclotides: Natural, circular plant peptides that possess significant activity against gastrointestinal nematode parasites of sheep. *Biochemistry* 47:5581–5589.
12. Gilding EK, et al. (2016) Gene coevolution and regulation lock cyclic plant defence peptides to their targets. *New Phytol* 210:717–730.
13. Wang CK, Craik DJ (2018) Designing macrocyclic disulfide-rich peptides for biotechnological applications. *Nat Chem Biol* 14:417–427.
14. Craik DJ, et al. (2018) Ribosomally-synthesised cyclic peptides from plants as drug leads and pharmaceutical scaffolds. *Bioorg Med Chem* 26:2727–2737.
15. Shafee T, Harris K, Anderson M (2015) Biosynthesis of cyclotides. *Adv Bot Res* 76:227–269.
16. Gruber CW, et al. (2007) A novel plant protein-disulfide isomerase involved in the oxidative folding of cystine knot defense proteins. *J Biol Chem* 282:20435–20446.
17. Conlan BF, Gillon AD, Barbeta BL, Anderson MA (2011) Subcellular targeting and biosynthesis of cyclotides in plant cells. *Am J Bot* 98:2018–2026.
18. Saska I, et al. (2007) An asparaginyl endopeptidase mediates in vivo protein backbone cyclization. *J Biol Chem* 282:29721–29728.
19. Nguyen GKT, et al. (2014) Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat Chem Biol* 10:732–738.
20. Harris KS, et al. (2015) Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat Commun* 6:10199.
21. Poon S, et al. (2018) Co-expression of a cyclizing asparaginyl endopeptidase enables efficient production of cyclic peptides in planta. *J Exp Bot* 69:633–641.
22. Jackson MA, et al. (2018) Molecular basis for the production of cyclic peptides by plant asparaginyl endopeptidases. *Nat Commun* 9:2411.
23. Bombarely A, et al. (2012) A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol Plant Microbe Interact* 25:1523–1530.
24. Nakasugi K, Crowhurst R, Bally J, Waterhouse P (2014) Combining transcriptome assemblies from multiple de novo assemblers in the allo-tetraploid plant *Nicotiana benthamiana*. *PLoS One* 9:e91776.
25. Richau KH, et al. (2012) Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics. *Plant Physiol* 158:1583–1599.
26. Paireder M, et al. (2017) The papain-like cysteine proteinases NbCysP6 and NbCysP7 are highly processive enzymes with substrate specificities complementary to *Nicotiana benthamiana* cathepsin B. *Biochim Biophys Acta Proteins Proteomics* 1865:444–452.
27. Paireder M, et al. (2016) The death enzyme CP14 is a unique papain-like cysteine proteinase with a pronounced S2 subsite selectivity. *Arch Biochem Biophys* 603:110–117.
28. Yang R, et al. (2017) Engineering a catalytically efficient recombinant protein ligase. *J Am Chem Soc* 139:5351–5358.
29. Turk V, et al. (2012) Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochim Biophys Acta* 1824:68–88.
30. Lecaille F, Chowdhury S, Purisima E, Brömme D, Lalmanach G (2007) The S2 subsites of cathepsins K and L and their contribution to collagen degradation. *Protein Sci* 16:662–670.
31. Bromme D, Bonneau P, Lachance P, Storer AC (1994) Engineering the S2 subsite specificity of human cathepsin-S to a cathepsin-L-like and cathepsin-B-like specificity. *J Biol Chem* 269:30239–30242.
32. Gillmor SA, Craik CS, Fletterick RJ (1997) Structural determinants of specificity in the cysteine protease cruzain. *Protein Sci* 6:1603–1611.
33. Yamada K, Matsushima R, Nishimura M, Hara-Nishimura I (2001) A slow maturation of a cysteine protease with a granulin domain in the vacuoles of senescing *Arabidopsis* leaves. *Plant Physiol* 127:1626–1634.
34. Reed MC, Lieb A, Nijhout HF (2010) The biological significance of substrate inhibition: A mechanism with diverse functions. *BioEssays* 32:422–429.
35. Misas-Villamil JC, van der Hoorn RAL, Doehlemann G (2016) Papain-like cysteine proteases as hubs in plant immunity. *New Phytol* 212:902–907.