ACS chemical biology

ARTICLES

pubs.acs.org/acschemicalbiology

Discovery of Cyclotides in the Fabaceae Plant Family Provides New Insights into the Cyclization, Evolution, and Distribution of Circular Proteins

Aaron G. Poth,[†,‡] Michelle L. Colgrave,[‡] Reynold Philip,[§] Bomai Kerenga,[§] Norelle L. Daly,[†] Marilyn A. Anderson,[∥] and David J. Craik[†,*]

[†]Institute for Molecular Bioscience, The University of Queensland, St. Lucia, QLD, Australia

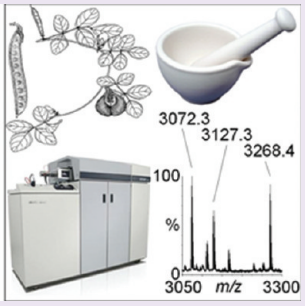
[‡]CSIRO Livestock Industries, St. Lucia, QLD, Australia

[§]Department of Chemistry, University of Papua New Guinea, Port Moresby, Papua New Guinea

[∥]Department of Biochemistry, La Trobe University, Melbourne, VIC, Australia

Supporting Information

ABSTRACT: Cyclotides are plant proteins whose defining structural features are a head-to-tail cyclized backbone and three interlocking disulfide bonds, which in combination are known as a cyclic cystine knot. This unique structural motif confers cyclotides with exceptional resistance to proteolysis. Their endogenous function is thought to be as plant defense agents, associated with their insecticidal and larval growth-inhibitory properties. However, in addition, an array of pharmaceutically relevant biological activities has been ascribed to cyclotides, including anti-HIV, anthelmintic, uterotonic, and antimicrobial effects. So far, >150 cyclotides have been elucidated from members of the Rubiaceae, Violaceae, and Cucurbitaceae plant families, but their wider distribution among other plant families remains unclear. *Clitoria ternatea* (Butterfly pea) is a member of plant family Fabaceae and through its usage in traditional medicine to aid childbirth bears similarity to *Oldenlandia affinis*, from which many cyclotides have been isolated. Using a combination of nanospray and matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) analyses, we examined seed extracts of *C. ternatea* and discovered cyclotides in the Fabaceae, the third-largest family of flowering plants. We characterized 12 novel cyclotides, thus expanding knowledge of cyclotide distribution and evolution within the plant kingdom. The discovery of cyclotides containing novel sequence motifs near the *in planta* cyclization site has provided new insights into cyclotide biosynthesis. In particular, MS analyses of the novel cyclotides from *C. ternatea* suggest that Asn to Asp variants at the cyclization site are more common than previously recognized. Moreover, this study provides impetus for the examination of other economically and agriculturally significant species within Fabaceae, now the largest plant family from which cyclotides have been described.



Cyclotides are a large family of macrocyclic plant proteins of 28−37 amino acids, featuring three intramolecular disulfide bonds. These bonds form a cyclic cystine knot (CCK) arrangement in which a ring formed by opposing peptide backbone segments and two disulfide bonds is penetrated by the third disulfide bond.[1] This motif imbues the cyclotides with structural rigidity and remarkable stability to thermal, chemical, and enzymatic degradation.[2] Their initial discovery resulted from examination of plant material used in the preparation of "kalata−kalata", a Congolese traditional medicine with uterotonic properties.[3] Their heat-stable nature was apparent from the outset, as the decoction consisted of a boiled leaf extract of *Oldenlandia affinis* (plant family Rubiaceae). The complete structural characterization (by NMR) of the major uterotonic component, kalata B1, was reported by Saether *et al.* in 1995,[4] uncovering the knotted disulfide connectivity and the macrocyclic nature of the peptide backbone.

Several pharmaceutically relevant bioactivities have since been ascribed to the cyclotides, including uterotonic,[3−5] anti-HIV,[6−8] antineurotensin,[9] antimicrobial,[10,11] and hemolytic [12,13] activities. Other bioactivities such as antifouling,[14] insecticidal,[15−17] anthelmintic,[18−20] and molluscidal [21] properties are consistent with their primary function as plant defense agents. Since the initial structural characterization of kalata B1, the potential of the cyclotides as scaffolds in the design of target-specific drugs has been recognized.[22] The traditional medicinal (ethnobotanical) use of plant extracts containing kalata B1 demonstrated the bioavailability of cyclotides as well as their heat-stable nature, and subsequent studies highlighting their enzymatic and chemical stability have broadened their pharmaceutical appeal. Recent studies[23−26] have demonstrated the increased stability afforded

Received: November 23, 2010
Accepted: December 31, 2010
Published: December 31, 2010

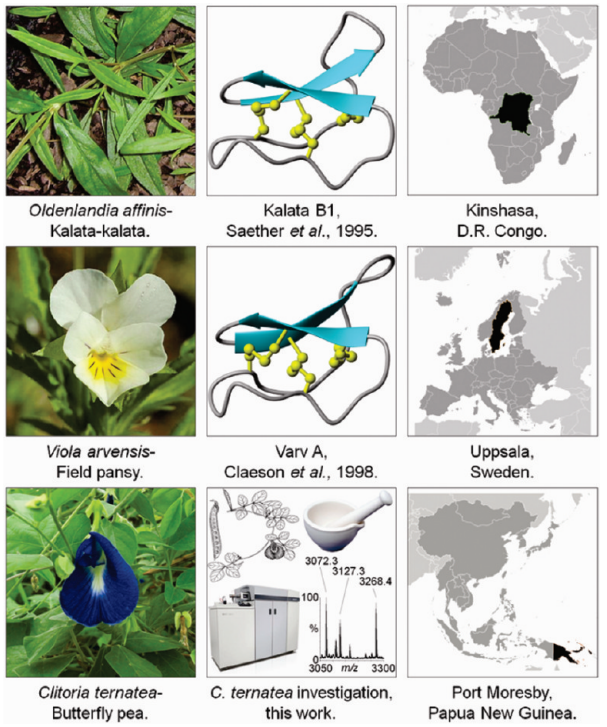


Figure 1. Botanical and geographical origins of the first cyclotides described from Rubiaceae, Violaceae, and Fabaceae plant families.

to xeno-bioactive linear epitopes when grafted to a cyclotide scaffold, as well as their potential as scaffolds for the development of therapeutically useful protease inhibitors. In recent efforts to evaluate the structure−activity relationships of cyclotides and establish novel synthetic routes to them, a variety of *in vivo* expression methodologies have been successfully employed.[27,28]

To date, more than 150 cyclotides have been characterized from members of the Rubiaceae, Violaceae, and Cucurbitaceae plant families,[29] and a recent investigation of Apocynaceae species[30] suggests that they might also be present in this plant family. In addition to the known cyclotide sequences, gene-sequencing (EST-searching) studies have revealed predicted proteins with sequence homology to cyclotides in numerous species within the Poaceae (grass) plant family.[31] Although cyclotides have been found in every species of Violaceae studied, their prevalence among species from other plant families is sporadic, and ultimately their global distribution and evolutionary beginnings remain unclear. Figure 1 illustrates the plant species from which the first cyclotides from each plant family were isolated, along with representations of their 3D structures and the sampling locations used in these studies.

Clitoria ternatea, an ornamental perennial climber also known as the Butterfly pea, is a member of plant family Fabaceae, originally from Africa but now also distributed among equatorial Asiatic countries and the Americas. Preparations of *C. ternatea* are utilized in a variety of indigenous medicines throughout these regions, with anecdotal evidence of their use in the traditional medicines of the Philippines, Cuba, and Indochina to promote uterine contractions and expedite childbirth.[32,33] These uses have some parallels to the cyclotide-containing Congolese preparation known as kalata−kalata from *Oldenlandia affinis*.[34] Furthermore, the reported bioactivities of cyclotides match closely those for crude plant extracts of *C. ternatea*, including antimicrobial,

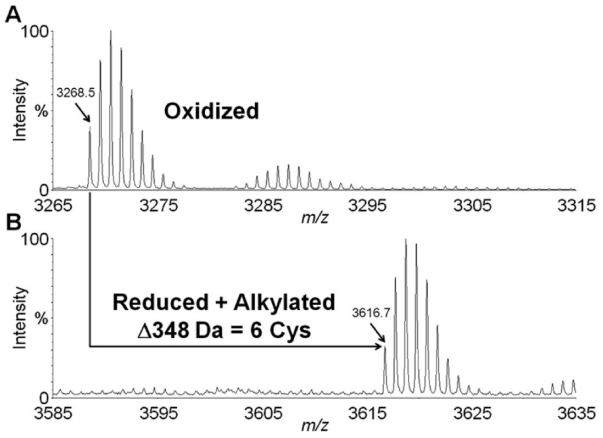


Figure 2. Evidence for the presence of cyclotides in *C. ternatea* seed extract. Offset-aligned MALDI-TOF spectra of (A) "native" and (B) "reduced and carbamidomethylated" putative cyclotide species, Cter A.

anthelmintic, and insecticidal activities,[35−39] as well as a reported effect on isolated rat uterus.[40]

In this study, we investigated seed extracts of *C. ternatea* for the presence of cyclotides using matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) and nanospray techniques. We report the sequences of 12 novel cyclotides, some of which contain previously unseen sequence motifs near the *in planta* cyclization site and provide new insights into the biosynthesis of cyclotides. As the Fabaceae is host to numerous economically and agriculturally important plant species, the discovery of cyclotides within this family elevates their interest among plant-derived defense peptides.

The hallmark of the Fabaceae is their unique ability to fix nitrogen, underpinning a long history of use by humans both as a food source[41] and for the enhancement of agricultural processes. In particular, Fabaceae species are a critical source of protein in human and animal nutrition, grain legumes being cultivated as a direct source of food and forage legumes used as livestock feed. In terms of nutrition, leguminous plants are considered second only to the grasses in their importance to humans.[42] The major agricultural application of Fabaceae plants is their rotation with other crops, resulting in enhanced soil nitrogen levels, which improve crop yields. Leguminous species are also an important source of various phytochemicals, including compounds with pharmaceutical and pesticidal significance, as well as those with industrial uses, including dyes, gums, fibers, and vegetable oil (see reviews by Morris[43] and Graham *et al.*[44]). The Fabaceae represent 27% of worldwide primary crop production.[44] Overall, this study is expected to precipitate the wider search for cyclotides among the Fabaceae, now the largest plant family from which cyclotides have been isolated.

RESULTS AND DISCUSSION

In this study we report the discovery of cyclotides in *C. ternatea*, a member of the Fabaceae plant family. Although an array of pharmaceutically relevant bioactivities has been ascribed to cyclotides, their key purpose in plants is presumed to be as antibiotic agents to fend off plant pests and pathogens.[45] The common usage of kalata−kalata and *C. ternatea* extracts as reproductive medications, as well as literature describing other cyclotide-like activities in *C. ternatea* extracts, including antimicrobial, anthelmintic, and insecticidal *in vitro* activities,[35−39]

Table 1. Sequence Alignment of Novel Cyclotides from *C. ternatea*

peptide	amino acid sequence[a]	expt *m/z*	mass (Da) expt	mass (Da) theory	error Δ (ppm)	subfamily
Cter A	GVIPCGESCVFIPC-ISTVIGCSCKNKVCYRN	1090.07	3267.19	3267.49	−91.8	bracelet
Cter B	G-VPCAESCVWIPCTVTALLGCSCKDKVCYLN	1084.58	3250.75	3250.45	92.6	bracelet
Cter C	G-VPCAESCVWIPCTVTALLGCSCKDKVCYLD	1084.93	3251.76	3251.43	99.2	bracelet
Cter D	G-IPCAESCVWIPCTVTALLGCSCKDKVCYLN	1089.26	3264.76	3264.46	91.0	bracelet
Cter E	G-IPCAESCVWIPCTVTALLGCSCKDKVCYLD	1089.61	3265.79	3265.45	105.2	bracelet
Cter F	G-IPCGESCVFIPC-ISSVVGCSCKSKVCYLD	1536.48	3070.94	3071.34	−132.7	bracelet
Cter G	G-LPCGESCVFIPC-ITTVVGCSCKNKVCYNN	1043.15	3126.42	3126.36	19.0	bracelet
Cter H	G-LPCGESCVFIPC-ITTVVGCSCKNKVCYND	1043.48	3127.43	3127.34	26.8	bracelet
Cter I	GTVPCGESCVFIPC-ITGIAGCSCKNKVCYIN	1052.33	3153.96	3154.39	−135.7	bracelet
Cter J	GTVPCGESCVFIPC-ITGIAGCSCKNKVCYID	1052.67	3154.99	3155.58	−122.2	bracelet
Cter K	H-EPCGESCVFIPC-ITTVVGCSCKNKVCY-N	1037.14	3108.39	3108.31	24.4	bracelet
Cter L	H-EPCGESCVFIPC-ITTVVGCSCKNKVCY-D	1037.47	3109.39	3109.30	31.3	bracelet

[a] Ile and Leu were determined by amino acid analysis where sufficient material was available or assigned on the basis of homology with published cyclotide sequences.

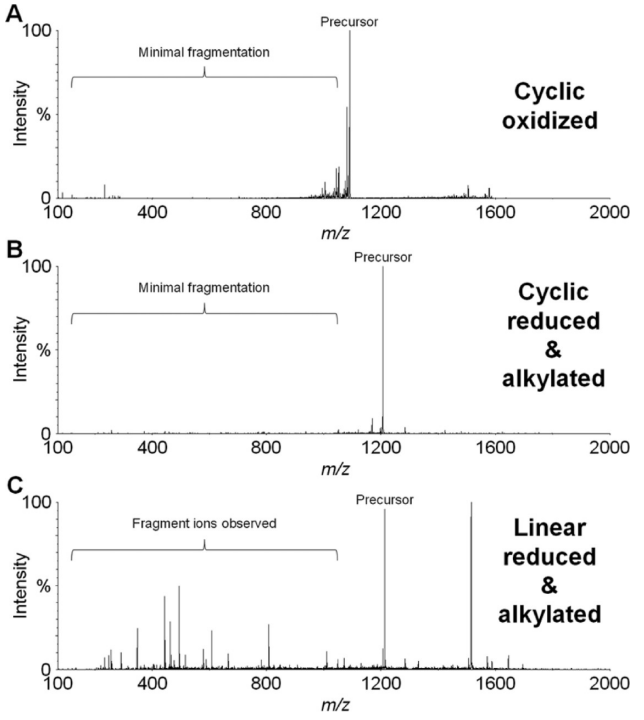


Figure 3. Nanospray tandem MS fragmentation patterns for "native" *versus* chemically modified Cter A at a collision energy setting of 50 V. (A) "Native" (cyclic oxidized) cyclotide precursor *m/z* 1090.1. (B) Cyclic reduced and alkylated precursor *m/z* 1206.1. (C) Digested/linearized reduced and alkylated precursor *m/z* 1212.1. These apparent triply charged fragment ions correspond to species of molecular masses 3267, 3615, and 3633 Da, respectively.

raised our interest in this species and thus we investigated *C. ternatea* for the presence of cyclotides.

Identification of Cyclotides in *C. ternatea*. Initial screening of crude seed extracts of *C. ternatea* revealed the presence of proteins with masses in the range 2500−4000 Da, consistent with those of known cyclotides.[30] Following preparative RP-HPLC of the crude extract, the putative cyclotides were detected in late-eluting fractions *via* MALDI-TOF MS (Figure 2A); the masses of 12 of these putative cyclotides are reported in Table 1. In accordance with established diagnostic methodology for cyclotides,[30] purified peptides were lyophilized, reduced and carbamidomethylated, and reanalyzed *via* MALDI-TOF MS. Mass increases of 348 Da were observed following this process (Figure 2B), indicating the presence of three intramolecular disulfide bonds in the corresponding proteins. Thus, the 12 peptides complied with all three diagnostic criteria previously identified for cyclotides[30] of mass profile, hydrophobicity profile, and disulfide content.

Tandem MS Enables Differentiation of Cyclotides from Linear Peptides. There are several examples of linear proteins, including knottins and also some defensins, that are of similar size to cyclotides, possess three disulfide bonds, and display hydrophobic properties. Therefore, we sought to extend the diagnostic criteria for the detection of cyclotides by including an additional step to distinguish between peptides with cyclic or linear backbones. This additional step is illustrated for the putative cyclotide from *C. ternatea* seed extract, Cter A, with a "native" mass of 3267.3 Da that increases by 348 Da after reduction and carbamidomethylation and a further 18 Da after enzymatic digestion of the peptide backbone with endoproteinase Glu-C (Figure 3). The determination of peptide sequence *via* tandem MS relies in part upon the ability of the N- and C-termini to retain charge. The absence of termini in cyclotides, brought about by their macrocyclic peptide backbone, therefore prevents their efficient fragmentation in tandem MS analyses, either as fully folded CCK-containing "native" proteins or as reduced and alkylated cyclic proteins, as illustrated for Cter A in panels A and B of Figure 3, respectively. Only after enzymatic cleavage of reduced (or reduced and alkylated) *C. ternatea* peptides into their linear forms were the various fragment ions detected during tandem MS analyses (Figure 3C). Hence, we propose that the characteristic lack of fragmentation observed in tandem MS analyses of reduced and/or reduced and alkylated cyclotides is a suitable determinant of their cyclic nature and should be added to previously proposed criteria[30] as an indicator for the presence of cyclotides in a given plant.

In combination, the newly defined criteria proposed here for the positive identification of cyclotides are late-eluting properties *via* RP-HPLC, a mass of 2500−4000 Da, an increase in mass of

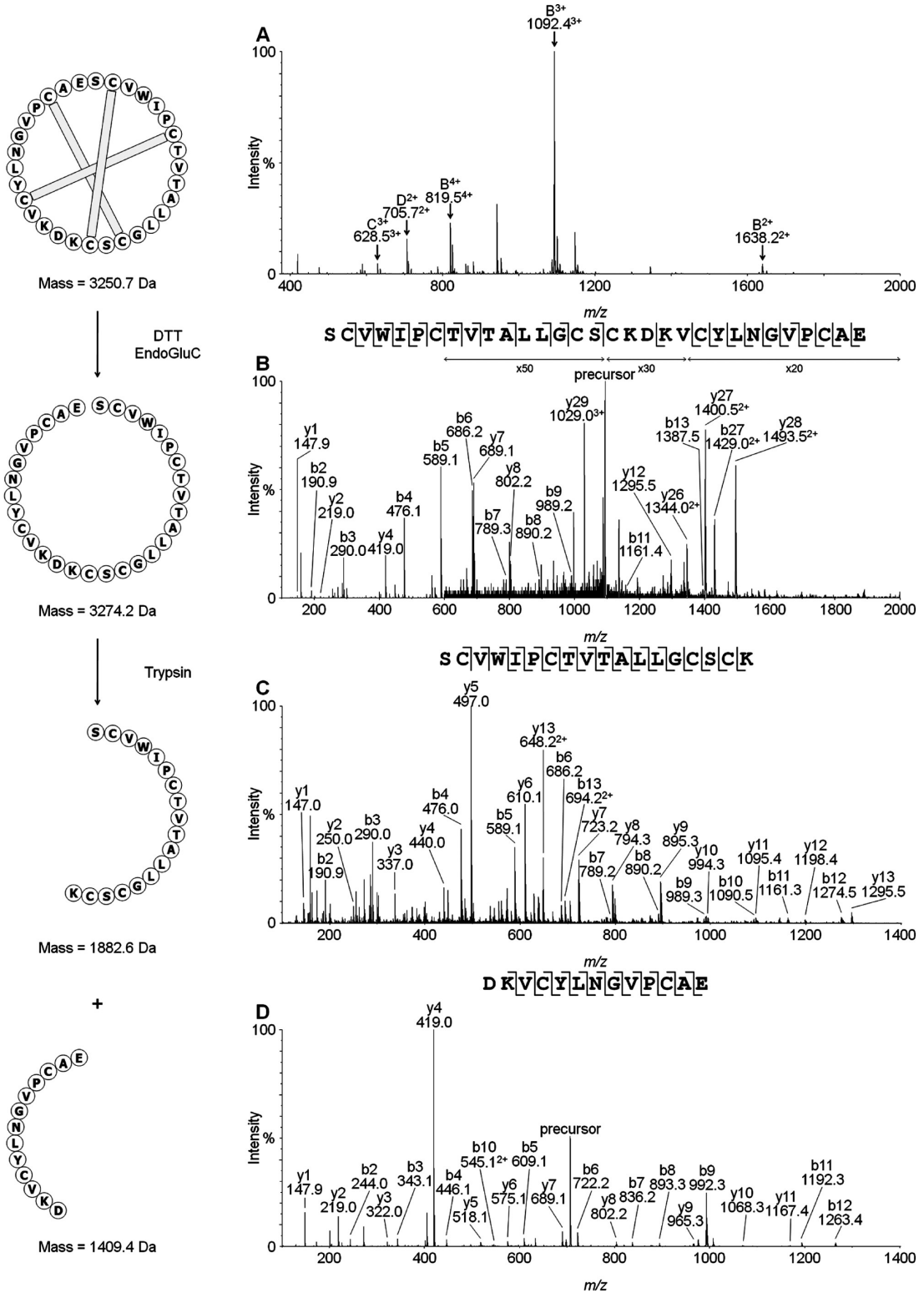


Figure 4. Nanospray sequencing of Cter B. Left panel: Digestion scheme and mass of proteolytic fragments. (A) TOF-MS spectrum of combined trypsin and endoproteinase Glu-C digest. The peaks are labeled according to their charge state, where B^{2+}, B^{3+}, and B^{4+} correspond to the full-length linearized Cter B, and C^{3+} and D^{2+} signify smaller fragments produced through cleavage of the cyclic precursor at two points along the peptide backbone. (B) MS/MS of precursor 1092.4^{3+} (3274.2 Da). (C) MS/MS of precursor 628.5^{3+} (1882.6 Da). (D) MS/MS of precursor 705.7^{2+} (1409.4 Da).

348 Da following reduction and alkylation with iodoacetamide, and inefficient fragmentation in MS/MS analyses of "native" or reduced and alkylated forms. Although yet to be described from plants, cyclic peptides with three intramolecular disulfide bonds not forming a cystine knot arrangement, similar to rhesus θ-defensin-1,[46] could also meet these criteria. However, judging

dx.doi.org/10.1021/cb100388j |ACS Chem. Biol. 2011, 6, 345–355

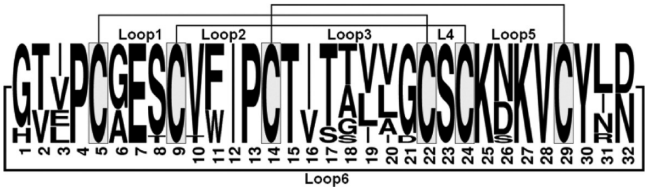


Figure 5. Sequence logo relative frequency plot of *C. ternatea* cyclotides identified in this study. Conserved residues among sequences include Pro4, Cys5, Glu7, Cys9, Ile12, Pro13, Cys14, Thr15, Cys22, Ser23, Cys24, Lys25, Lys27, Val28, Cys29, and Tyr30.

from the size and hydrophobicity of described θ-defensins, false positives are unlikely.

***De Novo* Sequencing of Cyclotides.** To illustrate the sequencing of the new cyclotides the step-by-step MS/MS analysis of Cter B is shown in Figure 4. The linearized peptide resulting from endoproteinase Glu-C digestion of the reduced form of Cter B was analyzed *via* nanospray MS/MS. *De novo* sequencing yielded a tentative identification of SCVWIPCTVTALLGCS-CKDKVCYLNGVPCAE. As indicated in Figure 4B, sequence ion coverage permitted definitive assignment of the sequence near the termini of the peptide but presented incomplete evidence for sequence close to the middle of the peptide, a feature observed in the analyses of many full-length linearized cyclotides. Combined trypsin and endoproteinase Glu-C digestion of reduced Cter B produced peptide fragments with complementary molecular weights of 1882.6 and 1409.4 Da. Complete sequence coverage for both precursors was attained in tandem MS analyses (Figure 4C and D), verifying the initial sequence assignment for the full-length linearized cyclotide. Using this approach, 12 novel cyclotides from *C. ternatea* were sequenced (Table 1). Amino acid analyses were conducted to confirm the MS/MS determined sequences and to discriminate between Ile and Leu for a representative set of cyclotides, including Cter A, Cter B and C, Cter D and E, Cter F, and Cter G and H; these data are provided as Supplementary Tables S1−S5. UniProtKB accession numbers are provided for the Cter peptides in Supplementary Table S6. Using a Basic Local Alignment Search Tool (TBLASTN) search, the deduced cyclotide sequences were used to query Fabaceae entries in the GenBank nonredundant database. No matches with significant sequence similarity were found.

Cyclotides are classified mainly into two subfamilies, Möbius or bracelet, based upon the presence or absence of a *cis*-Pro amide bond in loop 5. Cyclotides belonging to the bracelet subfamily are the most widely represented in the literature, at approximately 3-fold greater incidence than cyclotides belonging to the Möbius subfamily. Consistent with this prominence, the sequences discovered in the current study all belong to the bracelet subfamily. However, several of them have unusual residues at key processing sites, making them of interest for understanding processing mechanisms of cyclotides.

An efficient way to describe and compare the features of cyclotides is by referring to the intercysteine loops, illustrated in Figure 5 as an amino acid incidence plot for the 12 new sequences in sequence logo format.[47] Most of the new cyclotides comprised combinations of known loops from previously characterized cyclotides or novel loops with conservative amino acid substitutions. As a result, the majority of sequences displayed significant homology to known cyclotides. According to the sequence logo plot, the greatest variations in loop size and/or composition are in loops 3 and 6, consistent with data for all published cyclotide sequences as assessed using the "cyclotide loop view" tool within Cybase.[29]

Biochemical Properties of Novel Cyclotides. Since the initial discovery of the insecticidal activity of cyclotides,[15] several studies have demonstrated that this and other bioactivities are mediated through interactions with membranes.[17] An important physicochemical feature of cyclotides, with regard to membrane interaction, is a surface-exposed patch of hydrophobic residues. This surface-exposure presumably results from the exclusion of hydrophobic amino acids from the core of cyclotides owing to the presence of the CCK motif. In addition to the importance of defined hydrophobic moieties in potentiating cyclotide−membrane interactions, clusters of charged residues have been demonstrated as determinants of hemolytic, insecticidal, and anthelmintic activity. In particular, the hemolytic and anthelmintic properties of cyclotide variants correlate with these important structural features,[19] with the most active bracelet cyclotides displaying hydrophobic residues in loops 2 and 3 and positively charged residues in loops 5 and 6.

Among the novel Cter cyclotides identified here, Cter A has the largest net positive charge (2+) with basic residues clustered in loops 5 and 6, similar to the cycloviolacin peptides derived from *Viola odorata* that have been shown to possess potent anthelmintic activity.[19,20] The remaining peptides are clustered into groups with net positive 1+ (Cter G and Cter I), neutral (Cter B, Cter F, Cter H, Cter J, and Cter K), and those with net negative charge −1 (Cter C, Cter E, and Cter L). The bioactivities of cyclotides are further influenced by the manner in which they self-associate in membranes, which in turn is reliant upon the display of hydrophilic moieties on a "bioactive face" spatially distinct from the hydrophobic patches.[48,49] The proposed "bioactive face" is centered around a glutamic acid residue, an absolutely conserved feature among previously reported cyclotides. Consistent with previous findings, this glutamic acid is conserved among all novel cyclotides described in this study.

One Peptide or Two? Detection of N and D Peptide Isomers. Mass spectrometric analyses of a majority of isolated *C. ternatea* cyclotides generated peptide ions with ambiguous isotope patterns. The isotopic distributions of full-length linearized Cter B, as well as fragment peptides produced from dual enzyme digests of Cter B, are shown in Figure 6. As illustrated in panel A, the measured intensity of the monoisotopic peak at m/z 1092.4 relative to the rest of the isotopic envelope for full-length linearized Cter B is less than the theoretical intensity (indicated by dashed lines). Panel B demonstrates that the experimental and calculated isotopic distributions for the triply charged precursor at m/z 628.5 corresponding to the sequence SCVWIPCTV-TALLGCSCK match closely, whereas the experimental and calculated isotopic distributions for the doubly charged precursor at m/z 705.7 (panel C) corresponding to the sequence DKV-CYLNGVPCAE are clearly different. These mass spectral data indicate that multiple full-length cyclotide precursors are present in the sample and that the variable isotopic distributions observed among the precursor ions are associated with the cyclotide fragment corresponding to m/z 705.7.

Subsequent tandem MS analysis of the m/z 705.7 fragment was conducted to determine the point of variation in the peptide sequence. Panel D shows the tandem MS spectrum of the m/z 705.7 precursor, with diagnostic sequence ions indicated in bold. In panel E the b6 (DKVCYL, m/z 722.2) and y6 (GVPCAE, m/z 575.1) ions exhibit typical isotopic distributions for their size, with the monoisotopic peak appearing as the most intense and

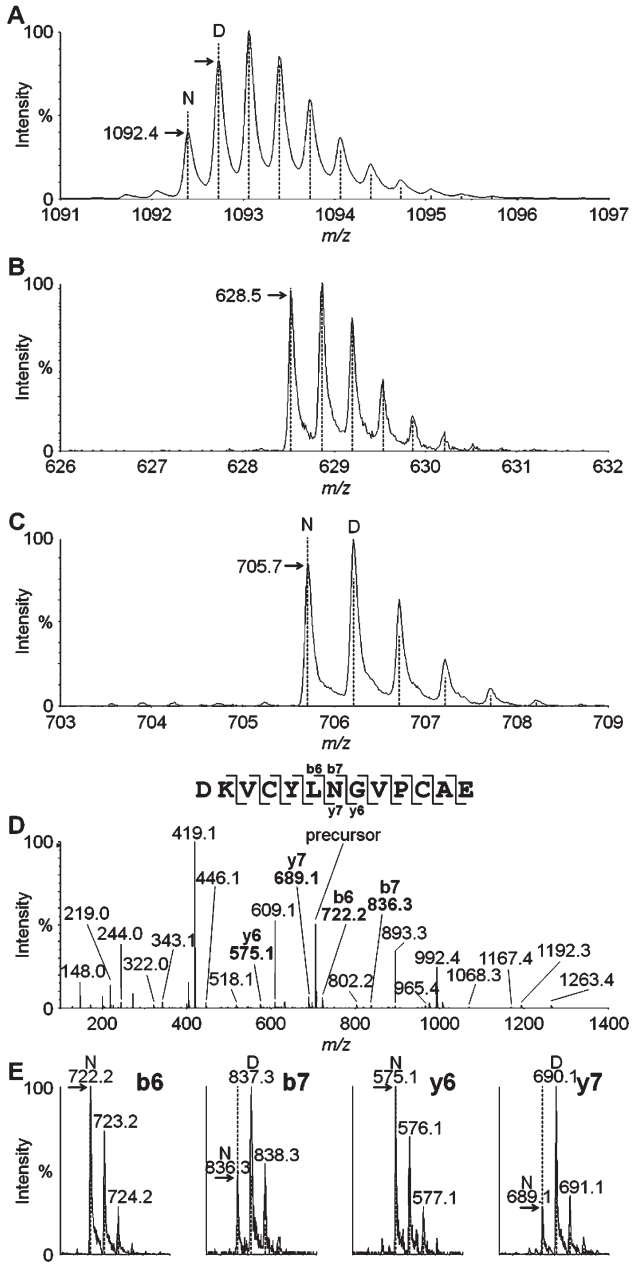


Figure 6. Isotopic distribution delineates isoform-specific sequence ions. Nanospray spectra for reduced and digested (trypsin and endoproteinase Glu-C) Cter B. (A) TOF-MS spectrum of full-length linearized Cter B precursor 3274.2 Da. (B) TOF-MS spectrum of Cter B digest product with precursor 628.5^{3+} (1882.6 Da). (C) TOF-MS spectrum of Cter B digest product with precursor 705.7^{2+} (1409.4 Da). (D) Full product ion spectrum of precursor m/z 705.7^{2+} (1409.4 Da). Sequence ions shown in bold represent cleavage of the amide bonds either side of the amino acid at position 7. (E) Isotopic distributions of diagnostic fragment ions b6, b7, y6, and y7 indicate the presence of both Asn and Asp at position 7 and thus the heterogeneous nature of the selected precursor ion within the transmission window. Dotted lines illustrate the theoretical isotopic distributions for precursor and fragment ions, assuming that the residue at position 7 is an asparagine. Arrows (→) indicate the observed intensities of labeled monoisotopic peaks.

with isotopic patterns matching closely with the theoretical patterns. The distributions for b7 (DKVCYLN, m/z 836.3) and y7 (NGVPCAE, m/z 689.1) ions, however, are skewed such that the most intense peak within their respective isotopic envelopes is that which normally corresponds to the monoisotopic peak of an analyte bearing a single ^{13}C atom. The fact that the peptide fragments in question are too small for this to be the case, along with the abrupt deviations in isotopic distribution from adjacent sequence ions, suggests the coexistence of peptides with an Asn or Asp at position 7 within the m/z 705.7 fragment, *i.e.*, DKVCYLNGVPCAE and DKVCYLDGVPCAE, corresponding to position 31 in the sequence of Cter B shown in Table 1. Of the *C. ternatea* cyclotides listed in Table 1, five pairs of sequences appear to be related through dual-isotope patterns of this nature, *i.e.*, Cter B and C; Cter D and E; Cter G and H; Cter I and J; and Cter K and L.

In the initial report detailing the discovery of cyclotides from *Viola odorata*,[1] the reported cyclotides, named cycloviolacins, all possessed an Asn in loop 6 corresponding to the C-terminus of linear precursor proteins. Subsequent examination of *V. odorata* using modified HPLC conditions[50] uncovered a range of novel cyclotides. The novel peptides included cycloviolacin O19 and cycloviolacin O20 whose sequences are highly homologous to those of previously reported cyclotides cycloviolacin O8 and cycloviolacin O3, respectively. Cycloviolacin O19 and cycloviolacin O20 possess a loop 6 Asp, in the place of Asn, and were not reported in the earlier study.[1] The study by Ireland *et al.* therefore provided the first evidence for the existence of Asn and Asp C-terminal cyclotide isoforms in *V. odorata*, indicating that highly homologous cyclotides differing by a C-terminal Asn or Asp or other single amino acid substitutions coelute during standard HPLC separations.[50] Given that most cyclotide separations reported in the literature have relied on these standard HPLC conditions, it is likely that in these studies, cyclotides with C-terminal Asn and Asp coeluted, thus eluding analysis.

The MS analysis undertaken in the current study demonstrates that Asn and Asp variants can be identified in a mixture through careful scrutiny of MS data. Furthermore, this study suggests that cyclotides with C-terminal Asp might be more common than previously reported, being missed in earlier MS analyses. The possibility also exists that cyclotides differing by 1 Da but whose sequences are homologous such as those that would result from the differential incorporation of Gln or Glu or those that differ at a range of positions may coelute.

N and D Peptide Isomers Exist Naturally *in Planta*. Of the more than 150 cyclotides characterized previously, only four pairs share sequences otherwise identical to each other apart from Asn and Asp variation in loop 6, *i.e.*, kalata B1 and B4, kalata B6 and B10, cycloviolacin O8 and O19, and cycloviolacin O3 and O20.[29] Therefore, the high incidence of Asn and Asp variants identified in this study warranted further examination to rule out deamidation as a possible cause of the synonymous sequences. Deamidation of Asn residues during sample workup is a commonly observed artefact in proteomic analyses, catalyzed by exposure of the sample to elevated temperatures and basic pH,[51] typically during enzymatic cleavage and occurring most frequently at Asn residues immediately N-terminal to Gly, as would be the case in these cyclic proteins. However, the isotopic distributions of "native" cyclotides extracted from fresh plant material at low pH and not heated before MS analysis suggest that Asn and Asp cyclotide variants characterized in this study, *i.e.*, Cter B and C; Cter D and E; Cter G and H; Cter I and J; and Cter K and L, coexist naturally. The existence of Cter A and Cter F, which do not display Asn or Asp variability and which were isolated from the same starting material and processed in parallel,

supports the natural coexistence of Asn and Asp C-terminal cyclotide isoforms.

A recent study of ESTs from the cyclotide-producing plant *O. affinis* reports high relative expression of a protein with close homology to asparaginase, whose biological function is the conversion of asparagine to aspartic acid.[52] With respect to pairs of cyclotides isolated from *O. affinis* differing only at the nascent C-terminus, the fact that only kalata B1 and kalata B6 (C-terminal Asn) genes have been found despite peptide evidence for kalata B1 and B4, and kalata B6 and B10 (each pair identical except for C-terminal Asn or Asp), led Qin *et al.* to suggest the alternative possibility that the "Asp" peptides are a product of post-translational processing occurring *in planta*.[52] A similar situation exists for related *V. odorata* peptides cycloviolacin O8 (C-terminal Asn) and cycloviolacin O19 (C-terminal Asp), with only the gene encoding the former cyclotide having been characterized.[53] However, it remains to be determined whether the observed "Asn or Asp" variable peptide pairs from *O. affinis* and *V. odorata* are a product of enzymatic post-translational processing and, further, whether a similar enzyme is involved in the biosynthesis of some metabolites with C-terminal Asp described from *C. ternatea* in this study.

Variable Residues in the Ligation Site Imply Catalytic Promiscuity. Since the discovery of the first cyclotide-encoding gene,[15] it has been evident that amino acids participating in cyclization are located in loop 6 of fully formed cyclotides. Recent studies exploring the structural characteristics of cyclotide precursor sequences involved in their cyclization[54,55] emphasize the importance of tripeptide motifs (typically Gly-Leu-Pro or Ser-Leu-Pro or Ala-Leu-Pro) demarcating the cyclotide domain and the positioning of an Asn or Asp residue immediately prior to the C-terminal tripeptide. In addition, these studies suggest that an as yet unidentified asparaginyl endopeptidase (AEP) is responsible for the ligation of cyclotide proto-termini as the final step of cyclotide biosynthesis.

Among the new cyclotides identified here, Cter G and Cter H, and Cter K and Cter L, contain novel amino acid sequences at their respective predicted sites of *in planta* cyclization. In the case of Cter G and Cter H, the loop 6 sequences "YNNGLP" and "YNDGLP" present the unique motifs Asn-Asn-Gly and Asn-Asp-Gly, which are noteworthy because they present two possible cyclization sites. The position of the peptide bond formed during cyclization of linear cyclotide precursors, as corroborated by gene sequencing efforts, is frequently observed at an Asn-Gly or the Asp-Gly junction. By itself, this information would suggest that the cyclization site in Cter H is Asp-Gly; however, the demonstrated cyclic nature of cycloviolacin O25,[50] which presents a loop 6 sequence "YFNDIF", tenders the alternative possibility that the cyclization reaction takes place between Asn-Asp. In the case of Cter K and Cter L, the loop 6 sequences are "YNHEP" and "YDHEP" with presumed novel cyclization sites Asn-His or Asp-His. Although there are other examples of cyclotides with a positively charged residue following Asp in the cyclization site (for example, "YHDKIP" in circulin D and circulin E),[56] this is the first example with an acidic residue in place of the typically small hydrophobic residue (Ala, Ile, Leu, or Val) at this position (second residue of mature cyclotide in presumed gene sequence). The existence of mature cyclic peptides with unusual residues within the N-terminal tripeptide motif (*e.g.*, His-Glu-Pro in Cter K and Cter L) suggests greater flexibility in cyclotide processing mechanisms within *C. ternatea* than observed in other cyclotide-producing species. A recent study in which a modified cyclotide gene was expressed in transgenic noncyclotide-containing plant species reported that mechanisms central to the processing of fully formed cyclotides are sensitive to changes in N-terminal sequence. In particular, Ala mutations at Gly_1 or Leu_2 in kalata B1 genes were found to disrupt the formation of cyclic products.[54]

Legumain, an AEP with transpeptidation (peptide ligation) activity, first described in jack beans,[57,58] is of potential significance to the processing of cyclotides. In particular, the demonstrated flexibility of a Fabaceae legumain that cleaves at almost all Asn-Xaa bonds[59] and to a lesser extent Asp-Xaa bonds[60] may prove to be relevant in the biosynthesis of cyclotides from *C. ternatea*. All Fabaceae cyclotides investigated in the current study, including those with nontypical sequence in loop 6, for example, "YNHEP" or "YDHEP" in Cter K and Cter L and "YNNGIP" or "YNDGIP" in Cter G and Cter H were observed only as cyclized peptides; no linear or misprocessed forms were detected.

Besides *C. ternatea* cyclotides possessing novel loop 6 sequences, there are a number of "orphan" cyclotides whose loop 6 sequences appear incompatible with the typical activity of AEPs previously implicated in cyclotide bioprocessing. Apart from cycloviolacin O25, whose loop 6 sequence indicates the lack of typical putative N-terminal amino acids Gly, Ser or, Ala in putative precursors, *Chassalia parvifolia* cyclotides circulin D and circulin E are distinct from other known cyclotides in that they have positively charged proto-N-termini, whereas circulin F does not have an Asn or Asp in loop 6. Cyclization by AEP is one of the proposed biosynthetic mechanisms proposed as being central to the cyclization of SFTI-1 [61] in *Helianthus annuus*; however, the gene sequence corresponding to amino acids surrounding the expressed protein sequence does not indicate the involvement of Gly-Leu-Pro tripeptide motifs regarded as essential in cyclotide precursor proteins.[54,55] In addition, sequence alignments of cyclic trypsin inhibitors MCoTI-I and MCoTI-II from *Momordica cochinchinensis* with related linear trypsin inhibitors[62] suggest that they exhibit C-terminal Gly as unprocessed precursor proteins. Therefore, it is tempting to speculate that cyclization strategies utilized by organisms in the production of cyclotides and other cyclic proteins vary between species, based in part upon the capabilities of available processing enzymes.

Distribution and Evolution: Cyclotides Are Expressed in the Agriculturally and Economically Important Fabaceae Family. The discovery of novel cyclotides in the Fabaceae is highly significant because it is the third-largest family of flowering plants, comprising more than 19,000 species, and now represents the largest family of plants from which cyclotides have been described. The leguminous plants *Arachis hypogea* (peanut), *Glycine max* (soybean), and *Cicer arietinum* (chickpea) all belong to the Fabaceae plant family, making it one of the most important families in terms of worldwide nutrition and agricultural practices. Aside from the direct importance of the grain legumes to human nutrition, fabaceous crops are important to primary production, where forage legumes represent an integral component of stock feed. Fallow legumes are important to the success of other crops, being rotated with them in seasonal cycles to take advantage of the replenished soil nitrogen levels brought about through nitrogen fixation in root nodules.

To date, the only other proteins with antibiotic potential described previously from seed extracts of *C. ternatea* include a cysteine-rich defensin[63] with antifungal properties and a protein named "finotin", which displays broad-spectrum antibacterial and insecticidal activity in addition to antifungal activity against

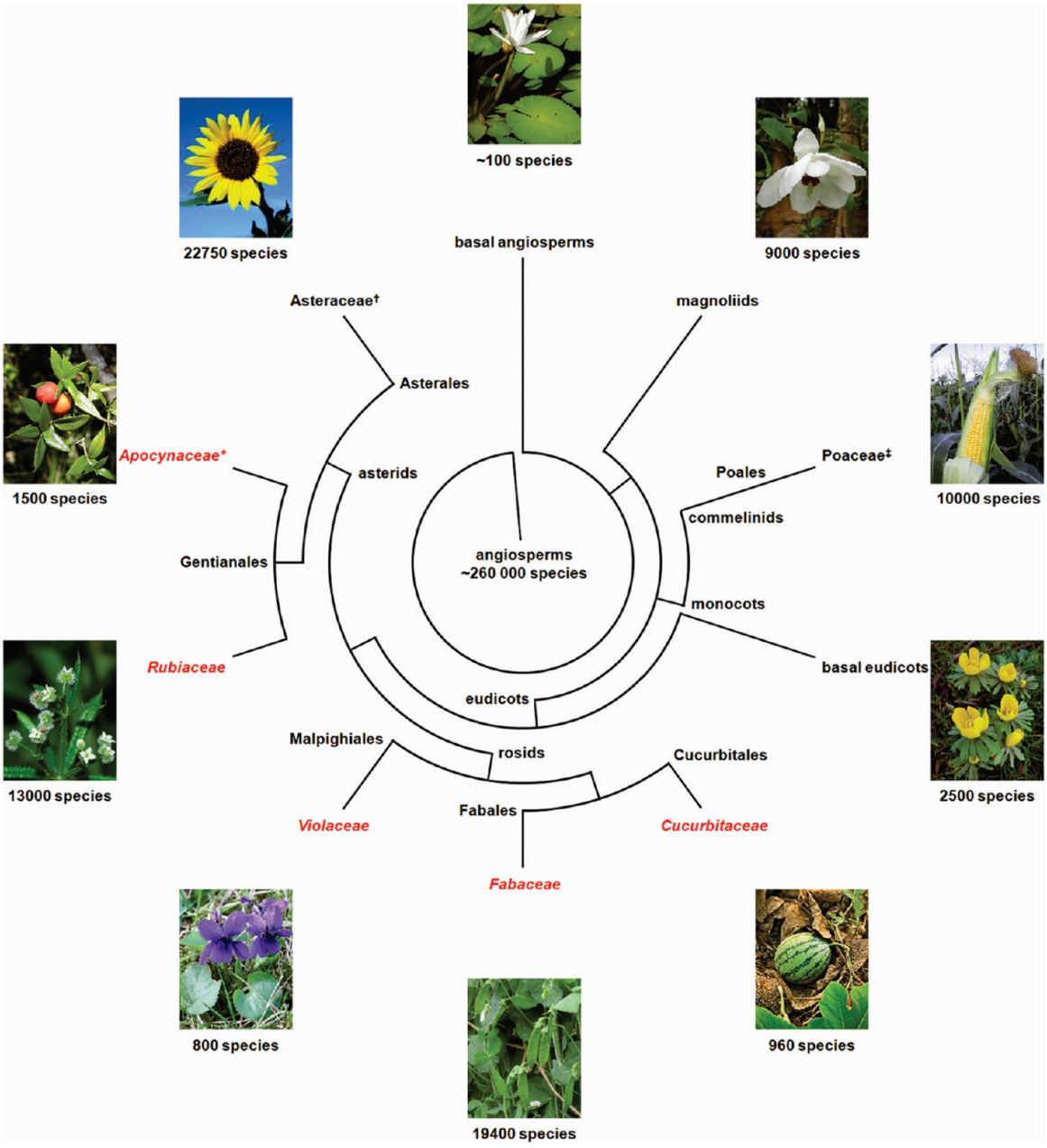


Figure 7. Distribution of ribosomally synthesized circular proteins within angiosperms. Cyclotide-containing plant families as reported in the literature appear in red italicized font. Notes: *A recent study reported evidence of cyclotides within the Apocynaceae family,[30] but no cyclotide peptide or nucleic acid sequences have been published yet. ‡Gene sequences encoding putative linear cyclotide-like proteins have been identified in several species within the Poaceae family. (These sequences lack the C-terminal Asn or Asp considered crucial for *in planta* cyclization). †Backbone-cyclized circular peptides distinct from cyclotides have been characterized from species within the Asteraceae family.

important plant pathogens.[35] Partial characterization of finotin revealed that it was approximately 20 kDa in size and that its activity could be attenuated following exposure to the proteolytic enzyme Pronase E (mycolysin). Cyclotides are typically smaller than 3700 Da and characteristically resilient to various proteolytic degradative processes. The demonstrated features of finotin are therefore incompatible with these criteria, suggesting that it is not a cyclotide.

Figure 7 summarizes current knowledge of the distribution of circular proteins in the plant kingdom, highlighting cyclotide-containing families among orders Gentianales, Malpighiales, Cucurbitales, and now Fabales, alongside Asterales, from which the small trypsin inhibitor SFTI-1 has been isolated. Genes encoding putative cyclotide-like cystine knot proteins have been identified in several Poales species (within the Poaceae family) and are significant in that they harbor a stop codon in place of a C-terminal Asp or Asn, which is thought to be a prerequisite for cyclization during the biosynthesis of cyclotides.[31]

Previous studies suggest that cyclotides most likely arose through convergent evolution from genetically distinct, linear cyclotide-like precursors.[30,64] The important feature common to cyclotide-producing plants appears to be the recruitment of the ubiquitous vacuole processing enzyme asparaginyl endopeptidase (AEP) to facilitate cyclization.[54,55] However, the insertion

of a cyclotide gene into *Arabidopsis thaliana* or *Nicotiana benthamiana* results primarily in the production of linear misprocessed proteins, highlighting differences in the capabilities of processing enzymes in different species. Combined with the current study, previous data outlined above suggest that it is likely that the evolution of cyclotides among the various plant species in which they are produced is a product of both the positioning of an appropriate amino acid at the C-terminal position of a knottin coding sequence and the parallel evolution of the biosynthetic infrastructure capable of bringing about efficient cyclization. It is worth noting that in cyclotide-containing species, only fully formed (cyclized) cyclotides have been characterized (with the exception of violacin A[65] and kalata B20-lin,[66] which lack a C-terminal Asn or Asp); no corresponding linear forms have been detected, further illustrating the efficiency of cyclization machinery involved in natural cyclotide biosynthesis.

In summary, in this study we have applied refined diagnostic criteria to the screening of *C. ternatea,* a member of plant family Fabaceae, enabling the unequivocal identification of cyclotides therein. Mass spectrometric experiments yielded the peptide sequences of a dozen novel bracelet cyclotides. Furthermore, the cyclotides described in this study, possessing novel putative N-termini, are suggestive of significantly different specificities of the enzymes involved in cyclotide biosynthesis from those reported earlier. Lastly, it was demonstrated that coeluting Asn and Asp variants can be identified in a mixture through careful analysis of MS data. This finding, combined with the high abundance of these variants observed within *C. ternatea*, suggests that cyclotides with a C-terminal Asp may be underreported in the literature. This study provides impetus for the discovery and characterization of cyclotides from a broad range of Fabaceae species to determine the proportion producing cyclotides, as well as the mechanistic capabilities of fabaceous processing enzymes. The discovery of novel cyclotide sequences alongside other circular peptide sequences from a wide range of plants is key to defining the cyclization strategies employed by plants and their biosynthetic pathways.

METHODS

Seed Extraction. Seed material (~20 g) from *C. ternatea* (Milgarra variety as supplied by Heritage Seeds) was ground in a mortar and pestle prior to solvent extraction with 100 mL of 50% (v/v) acetonitrile, 2% (v/v) formic acid. Crude extract was centrifuged for 4 min at 4,000*g*, and the supernatant passed through a 0.45 μm syringe filter prior to lyophilization, yielding 430 mg material.

Solid-Phase Extraction (SPE). Crude plant extracts were redissolved in 1% (v/v) formic acid and underwent an SPE cleanup step prior to further analysis. Waters C18 SPE cartridges of 100 mg to 10 g resin capacity were activated with 10 bed volumes of methanol and subsequently equilibrated with 10 bed volumes of 1% (v/v) formic acid. Following application of crude plant extracts, the cartridges were washed with a further 10 bed volumes of 1% (v/v) formic acid. Interfering substances were eluted from the cartridges in 10% (v/v) acetonitrile, and cyclotides collected in 20% to 80% (v/v) acetonitrile elution steps as separate fractions and lyophilized.

HPLC Purification. Separation of cyclotides from crude *C. ternatea* extracts or SPE fractions was carried out using preparative or semipreparative HPLC. For preparative HPLC, samples were reconstituted in 10% (v/v) acetonitrile, 1% (v/v) trifluoroacetic acid and introduced to a Phenomenex C18 RP-HPLC column (250 mm × 21.2 mm, 15 μm, 300 Å). Using a Waters 600E HPLC unit, a linear 1% min^{-1} acetonitrile gradient was delivered to the column at a flow rate of 8 mL min^{-1}, the eluent was monitored using a dual wavelength UV detector set to 214 and 280 nm, and fractions were collected. In semipreparative HPLC separations, a Phenomenex C18 RP-HPLC column (250 mm × 10 mm, 10 μm, 300 Å) was utilized with a flow rate of 3 mL min^{-1}. Selected cyclotides were purified to >95% purity through repetitive RP-HPLC and duplicate samples submitted for amino acid analysis.

MALDI-TOF MS. MALDI-TOF analyses were conducted using an Applied Biosystems 4700 TOF-TOF Proteomics Analyzer. Samples were prepared through 1:1 dilution with matrix consisting of 5 mg mL^{-1} CHCA in 50% (v/v) acetonitrile, 1% (v/v) formic acid prior to spotting on a stainless steel MALDI target. MALDI-TOF spectra were acquired in reflector positive operating mode with source voltage set at 20 kV and Grid1 voltage at 12 kV, mass range 1000−5000 Da, focus mass 1500 Da, collecting 1500 shots using a random laser pattern and with a laser intensity of 3500. External calibration was performed by spotting CHCA matrix 1:1 with Applied Biosystems Sequazyme Peptide Mass Standards Kit calibration mixture diluted 1:400 as described previously.[67]

Enzymatic Digestion. Prior to tandem MS analyses, cyclotides were cleaved to produce linearized fragments following reduction and alkylation to prevent reoxidation. Lyophilized samples were reconstituted in 100 mM NH_4HCO_3 (pH 8) and a 10 μL portion was reduced by addition of 10 μL of 10 mM dithiothreitol and incubation at 60 °C for 30 min in a nitrogenous atmosphere. Incubation with a further 10 μL of 100 mM iodoacetamide followed for 30 min at RT. Samples were split into three ~7 μL fractions for digestion by endoproteinase Glu-C (Sigma P2922), TPCK-treated bovine trypsin (Sigma T1426) or a combination of both enzymes. In the case of the single-enzyme digests, a sample of ~7 μL received 5 μL of 40 ng μL^{-1} enzyme and 5 μL of 100 mM NH_4HCO_3. For the double-enzyme digest, a sample of ~7 μL was mixed with 5 μL of 40 ng μL^{-1} of each enzyme. All three digests were incubated at 37 °C for 3 h and then quenched with formic acid. All samples were retained at 4 °C until further analysis.

Nanospray on QSTAR Pulsar. Reduced and enzymatically digested samples were processed using C18 ziptips (Millipore) to remove salts and elicit a solvent exchange from aqueous solution to 80% (v/v) acetonitrile, 1% (v/v) formic acid. Samples (3 μL) were introduced to nanospray tips (Proxeon ES380) and 900 V was applied to the tip to induce nanoelectrospray ionization on a QSTAR Pulsar I QqTOF mass spectrometer (Applied Biosystems). The collision energy (CE) was varied from 10 to 60 V. Both TOF and product ion mass spectra were acquired and manually assigned using Analyst QS 1.1 Software.

Database Searching. Nucleotide sequences appearing in the GenBank nonredundant database corresponding to Fabaceae species were interrogated using the Basic Local Alignment Search Tool (TBLASTN) with queries consisting of the deduced amino acid sequences of novel *C. ternatea* cyclotides.

ASSOCIATED CONTENT

Supporting Information. This material is available free of charge *via* the Internet at http://pubs.acs.org.

Accession Codes

Deduced *C. ternatea* cyclotide sequences appear in the UniProtKB database with accession numbers as follows: Cter A, P86841; Cter B, P86842; Cter C, P86843; Cter D, P86844; Cter E, P86845; Cter F, P86846; Cter G, P86847; Cter H, P86848; Cter I, P86849; Cter J, P86850; Cter K, P86851; Cter L, P86852.

AUTHOR INFORMATION

Corresponding Author

*Tel: +61 7 3346 2019. Email: d.craik@uq.edu.au.

■ ACKNOWLEDGMENT

This research was supported by a grant from the Australian Research Council (DP0984390). D.J.C. is a NHMRC Professorial Research Fellow. N.L.D. is a Queensland Smart State Fellow. We are grateful to A. Jones of the Molecular and Cellular Proteomics Mass Spectrometry Facility at the Institute for Molecular Bioscience (The University of Queensland) for providing technical support and access to the mass spectrometry facility. The authors thank A. Cooper for revisions to the manuscript.

■ ABBREVIATIONS USED

kB1, kalata B1; RP-HPLC, reversed-phase high performance liquid chromatography; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; CCK, cyclic cystine knot; SFTI-1, sunflower trypsin inhibitor-1; AEP, asparaginyl endopeptidase; SPE, solid phase extraction; CHCA, α-cyano-4-hydroxycinnamic acid; CE, collision energy

■ REFERENCES

(1) Craik, D. J., Daly, N. L., Bond, T., and 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.

(2) Colgrave, M. L., and Craik, D. J. (2004) Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: The importance of the cyclic cystine knot. *Biochemistry 43*, 5965–5975.

(3) Gran, L. (1973) On the effect of a polypeptide isolated from "Kalata-Kalata" (*Oldenlandia affinis* DC) on the oestrogen dominated uterus. *Acta Pharmacol. Toxicol. 33*, 400–408.

(4) Saether, O., Craik, D. J., Campbell, I. D., Sletten, K., Juul, J., and Norman, D. G. (1995) Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. *Biochemistry 34*, 4147–4158.

(5) Gran, L. (1973) Oxytocic principles of *Oldenlandia affinis*. *Lloydia 36*, 174–178.

(6) Gustafson, K. R., Sowder, R. C., Henderson, L. E., Parsons, I. C., Kashman, Y., Cardellina, J. H., McMahon, J. B., Buckheit, R. W., Pannell, L. K., and Boyd, M. R. (1994) Circulins A and B: Novel HIV-inhibitory macrocyclic peptides from the tropical tree *Chassalia parvifolia*. *J. Am. Chem. Soc. 116*, 9337–9338.

(7) Daly, N. L., Koltay, A., Gustafson, K. R., Boyd, M. R., Casas-Finet, J. R., and Craik, D. J. (1999) Solution structure by NMR of circulin A: a macrocyclic knotted peptide having anti-HIV activity. *J. Mol. Biol. 285*, 333–345.

(8) Hallock, Y. F., Sowder, R. C., Pannell, L. K., Hughes, C. B., Johnson, D. G., Gulakowski, R., Cardellina, J. H., and Boyd, M. R. (2000) Cycloviolins A-D, anti-HIV macrocyclic peptides from *Leonia cymosa* *J. Org. Chem. 65*, 124–128.

(9) Witherup, K. M., Bogusky, M. J., Anderson, P. S., Ramjit, H., Ransom, R. W., Wood, T., and Sardana, M. (1994) Cyclopsychotride-A, a biologically-active, 31-residue cyclic peptide isolated from *Psychotria longipes*. *J. Nat. Prod. 57*, 1619–1625.

(10) Tam, J. P., Lu, Y.-A., Yang, J.-L., and Chiu, K.-W. (1999) An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proc. Natl. Acad. Sci. U.S.A. 96*, 8913–8918.

(11) Pranting, M., Loov, C., Burman, R., Göransson, U., and Andersson, D. I. (2010) The cyclotide cycloviolacin O2 from *Viola odorata* has potent bactericidal activity against Gram-negative bacteria. *J. Antimicrob. Chemother. 65*, 1964–1971.

(12) Schöpke, T., Hasan, A. M. I., Kraft, R., Otto, A., and Hiller, K. (1993) Compounds with hemolytic activity from *Viola tricolor* L. and *Viola arvensis* Murray. *Sci. Pharm. 61*, 145–153.

(13) Chen, B., Colgrave, M. L., Wang, C. N., and Craik, D. J. (2006) Cycloviolacin H4, a hydrophobic cyclotide from *Viola hederaceae*. *J. Nat. Prod. 69*, 23–28.

(14) Göransson, U., Sjögren, M., Svangård, E., Claeson, P., and Bohlin, L. (2004) Reversible antifouling effect of the cyclotide cycloviolacin O2 against barnacles. *J. Nat. Prod. 67*, 1287–1290.

(15) Jennings, C., West, J., Waine, C., Craik, D., and Anderson, M. (2001) Biosynthesis and insecticidal properties of plant cyclotides: The cyclic knotted proteins from *Oldenlandia affinis*. *Proc. Natl. Acad. Sci. U.S.A. 98*, 10614–10619.

(16) Jennings, C. V., Rosengren, K. J., Daly, N. L., Plan, M., Stevens, J., Scanlon, M. J., Waine, C., Norman, D. G., Anderson, M. A., and Craik, D. J. (2005) Isolation, solution structure, and insecticidal activity of Kalata B2, a circular protein with a twist: Do Möbius strips exist in nature? *Biochemistry 44*, 851–860.

(17) Barbeta, B. L., Marshall, A. T., Gillon, A. D., Craik, D. J., and Anderson, M. A. (2008) Plant cyclotides disrupt epithelial cells in the midgut of lepidopteran larvae. *Proc. Natl. Acad. Sci. U.S.A. 105*, 1221–1225.

(18) Colgrave, M. L., Kotze, A. C., Huang, Y. H., O'Grady, J., Simonsen, S. M., and Craik, D. J. (2008) Cyclotides: Natural, circular plant peptides that possess significant activity against gastrointestinal nematode parasites of sheep. *Biochemistry 47*, 5581–5589.

(19) Colgrave, M. L., Kotze, A. C., Ireland, D. C., Wang, C. K., and Craik, D. J. (2008) The anthelmintic activity of the cyclotides: Natural variants with enhanced activity. *ChemBioChem 9*, 1939–1945.

(20) Colgrave, M. L., Kotze, A. C., Kopp, S., McCarthy, J. S., Coleman, G. T., and Craik, D. J. (2009) Anthelmintic activity of cyclotides: *In vitro* studies with canine and human hookworms. *Acta Trop. 109*, 163–166.

(21) Plan, M. R. R., Saska, I., Cagauan, A. G., and Craik, D. J. (2008) Backbone cyclised peptides from plants show molluscicidal activity against the rice pest *Pomacea canaliculata* (golden apple snail). *J. Agric. Food Chem. 56*, 5237–5241.

(22) Craik, D. J., Mylne, J. S., and Daly, N. L. (2010) Cyclotides: macrocyclic peptides with applications in drug design and agriculture. *Cell. Mol. Life Sci. 67*, 9–16.

(23) Kolmar, H. (2008) Alternative binding proteins: biological activity and therapeutic potential of cystine-knot miniproteins. *FEBS J. 275*, 2684–2690.

(24) Gunasekera, S., Foley, F. M., Clark, R. J., Sando, L., Fabri, L. J., Craik, D. J., and Daly, N. L. (2008) Engineering stabilized vascular endothelial growth factor-A antagonists: Synthesis, structural characterization, and bioactivity of grafted analogues of cyclotides. *J. Med. Chem. 51*, 7697–7704.

(25) Thongyoo, P., Bonomelli, C., Leatherbarrow, R. J., and Tate, E. W. (2009) Potent inhibitors of beta-tryptase and human leukocyte elastase based on the MCoTI-II scaffold. *J. Med. Chem. 52*, 6197–6200.

(26) Gao, Y., Cui, T., and Lam, Y. (2010) Synthesis and disulfide bond connectivity-activity studies of a kalata B1-inspired cyclopeptide against dengue NS2B-NS3 protease. *Bioorg. Med. Chem. 18*, 1331–1336.

(27) Camarero, J. A., Kimura, R. H., Woo, Y. H., Shekhtman, A., and Cantor, J. (2007) Biosynthesis of a fully functional cyclotide inside living bacterial cells. *ChemBioChem 8*, 1363–1366.

(28) Thongyoo, P., Jaulent, A. M., Tate, E. W., and Leatherbarrow, R. J. (2007) Immobilized protease-assisted synthesis of engineered cysteine-knot microproteins. *ChemBioChem 8*, 1107–1109.

(29) Kaas, Q., and Craik, D. J. (2010) Analysis and classification of circular proteins in CyBase. *Peptide Sci. 94*, 584–591.

(30) Gruber, C. W., Elliot, A., G., Ireland, D. C., Delprete, P. G., Dessein, S., Göransson, U., Trabi, M., Wang, C. K., Kinghorn, A. B., Robbrecht, E., and Craik, D. J. (2008) Distribution and evolution of circular miniproteins in flowering plants. *Plant Cell 20*, 2471–2483.

(31) Mulvenna, J. P., Mylne, J. S., Bharathi, R., Burton, R. A., Shirley, N. J., Fincher, G. B., Anderson, M. A., and Craik, D. J. (2006) Discovery of cyclotide-like protein sequences in graminaceous crop plants: ancestral precursors of circular proteins? *Plant Cell 18*, 2134–2144.

(32) Fantz, P. R. (1991) Ethnobotany of *Clitoria* (Leguminosae). *Econ. Bot. 45*, 511–520.

(33) Mukherjee, P. K., Kumar, V., Kumar, N. S., and Heinrich, M. (2008) The Ayurvedic medicine *Clitoria ternatea*—from traditional use to scientific assessment. *J. Ethnopharmacol. 120*, 291–301.

(34) Gran, L. (1970) An oxytocic principle found in *Oldenlandia affinis* DC. An indigenous, Congolese drug "Kalata-Kalata" used to accelerate the delivery. *Medd. Norsk Farm. Selskap 32*, 173–180.
(35) Kelemu, S., Cardona, C., and Segura, G. (2004) Antimicrobial and insecticidal protein isolated from seeds of *Clitoria ternatea*, a tropical forage legume. *Plant Physiol. Biochem. 42*, 867–873.
(36) Malabadi, R., Mulgund, G., and Nataraja, K. (2005) Screening of antibacterial activity in the extracts of *Clitoria ternatea*. *J. Med. Aromat. Plant Sci. 27*, 26–29.
(37) Khadatkar, S. N., Manwar, J. V., and Bhajipale, N. S. (2008) *In-vitro* anthelmintic activity of root of *Clitoria ternatea* Linn. *Pharmacogn. Mag. 4*, 148–150.
(38) Kamilla, L., Mansor, S. M., Ramanathan, S., and Sasidharan, S. (2009) Effects of *Clitoria ternatea* leaf extract on growth and morphogenesis of *Aspergillus niger*. *Microsc. Microanal. 15*, 366–372.
(39) Mathew, N., Anitha, M. G., Bala, T. S. L., Sivakumar, S. M., Narmadha, R., and Kalyanasundaram, M. (2009) Larvicidal activity of *Saraca indica*, *Nyctanthes arbor-tristis*, and *Clitoria ternatea* extracts against three mosquito vector species. *Parasitol. Res. 104*, 1017–1025.
(40) Goel, A. K., Kulshreshtha, D. K., Dubey, M. P., and Rajendran, S. M. (2002) Screening of Indian plants for biological activity: Part XVI. *Indian J. Exp. Biol. 40*, 812–827.
(41) Gorman, C. F. (1969) Hoabinhian: A pebble-tool complex with early plant associations in Southeast Asia. *Science 163*, 671–673.
(42) Vietmeyer, N. D. (1986) Lesser-known plants of potential use in agriculture and forestry. *Science 232*, 1379–1384.
(43) Morris, J. (1997) Special-purpose legume genetic resources conserved for agricultural, industrial, and pharmaceutical use. *Econ. Bot. 51*, 251–263.
(44) Graham, P. H., and Vance, C. P. (2003) Legumes: Importance and constraints to greater use. *Plant Physiol. 131*, 872–877.
(45) Craik, D. J. (2009) Circling the enemy: cyclic proteins in plant defence. *Trends Plant Sci. 14*, 328–335.
(46) Tang, Y.-Q., Yuan, J., Ösapay, G., Ösapay, K., Tran, D., Miller, C. J., Ouellette, A. J., and Selsted, M. E. (1999) A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated α-defensins. *Science 286*, 498–502.
(47) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res. 14*, 1188–1190.
(48) Huang, Y. H., Colgrave, M. L., Daly, N. L., Keleshian, A., Martinac, B., and Craik, D. J. (2009) The biological activity of the prototypic cyclotide kalata B1 is modulated by the formation of multimeric pores. *J. Biol. Chem. 284*, 20699–20707.
(49) Huang, Y.-H., Colgrave, M. L., Clark, R. J., Kotze, A. C., and Craik, D. J. (2010) Lysine-scanning mutagenesis reveals an amendable face of the cyclotide kalata B1 for the optimization of nematocidal activity. *J. Biol. Chem. 285*, 10797–10805.
(50) Ireland, D. C., Colgrave, M. L., and Craik, D. J. (2006) A novel suite of cyclotides from *Viola odorata*: sequence variation and the implications for structure, function and stability. *Biochem. J. 400*, 1–12.
(51) Wright, H. T. (1991) Nonenzymatic deamidation of asparaginyl and glutaminyl residues in proteins. *Crit. Rev. Biochem. Mol. Biol. 26*, 1–52.
(52) Qin, Q., McCallum, E., Kaas, Q., Suda, J., Saska, I., Craik, D., and Mylne, J. (2010) Identification of candidates for cyclotide biosynthesis and cyclisation by expressed sequence tag analysis of *Oldenlandia affinis*. *BMC Genomics 11*, 111.
(53) Dutton, J. L., Renda, R. F., Waine, C., Clark, R. J., Daly, N. L., Jennings, C. V., Anderson, M. A., and Craik, D. J. (2004) Conserved structural and sequence elements implicated in the processing of gene-encoded circular proteins. *J. Biol. Chem. 279*, 46858–46867.
(54) Gillon, A. D., Saska, I., Jennings, C. V., Guarino, R. F., Craik, D. J., and Anderson, M. A. (2008) Biosynthesis of circular proteins in plants. *Plant J. 53*, 505–515.
(55) Saska, I., Gillon, A. D., Hatsugai, N., Dietzgen, R. G., Hara-Nishimura, I., Anderson, M. A., and Craik, D. J. (2007) An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclization. *J. Biol. Chem. 282*, 29721–29728.
(56) Gustafson, K. R., Walton, L. K., Sowder, R. C., Johnson, D. G., Pannell, L. K., Cardellina, J. H., and Boyd, M. R. (2000) New circulin macrocyclic polypeptides from *Chassalia parvifolia*. *J. Nat. Prod. 63*, 176–178.
(57) Carrington, D. M., Auffret, A., and Hanke, D. E. (1985) Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature 313*, 64–67.
(58) Min, W., and Jones, D. H. (1994) *In vitro* splicing of concanavalin A is catalyzed by asparaginyl endopeptidase. *Nat. Struct. Biol. 1*, 502–504.
(59) Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I., and Ishii, S. (1993) Asparaginyl endopeptidase of jack bean seeds. Purification, characterization, and high utility in protein sequence analysis. *J. Biol. Chem. 268*, 3525–3529.
(60) Halfon, S., Patel, S., Vega, F., Zurawski, S., and Zurawski, G. (1998) Autocatalytic activation of human legumain at aspartic acid residues. *FEBS Lett. 438*, 114–118.
(61) Mulvenna, J. P., Foley, F. M., and Craik, D. J. (2005) Discovery, structural determination, and putative processing of the precursor protein that produces the cyclic trypsin inhibitor sunflower trypsin inhibitor 1. *J. Biol. Chem. 280*, 32245–32253.
(62) Hernandez, J.-F., Gagnon, J., Chiche, L., Nguyen, T. M., Andrieu, J.-P., Heitz, A., Trinh Hong, T., Pham, T. T. C., and Le Nguyen, D. (2000) Squash trypsin inhibitors from *Momordica cochinchinensis* exhibit an atypical macrocyclic structure. *Biochemistry 39*, 5722–5730.
(63) Osborn, R. W., De Samblanx, G. W., Thevissen, K., Goderis, I., Torrekens, S., Van Leuven, F., Attenborough, S., Rees, S. B., and Broekaert, W. F. (1995) Isolation and characterization of plant defensins from seeds of Asteraceae, Fabaceae, Hippocastanaceae, and Saxifragaceae. *FEBS Lett. 368*, 257–262.
(64) Gruber, C. W. (2010) Global cyclotide adventure: A journey dedicated to the discovery of circular peptides from flowering plants. *Peptide Sci. 94*, 565–572.
(65) Ireland, D. C., Colgrave, M. L., Nguyencong, P., Daly, N. L., and Craik, D. J. (2006) Discovery and characterization of a linear cyclotide from *Viola odorata*: Implications for the processing of circular proteins. *J. Mol. Biol. 357*, 1522–1535.
(66) Mylne, J. S., Wang, C. K., van der Weerden, N. L., and Craik, D. J. (2010) Cyclotides are a component of the innate defence of *Oldenlandia affinis*. *Biopolymers 94*, 635–646.
(67) Saska, I., Colgrave, M. L., Jones, A., Anderson, M. A., and Craik, D. J. (2008) Quantitative analysis of backbone-cyclised peptides in plants. *J. Chromatogr. B 872*, 107–114.