Cyclotides Associate with Leaf Vasculature and Are the Products of a Novel Precursor in Petunia (Solanaceae)*1,2

Aaron G. Poth1,2, Joshua S. Mylne1, Julia Grassl1, Russell E. Lyons5, A. Harvey Millar*1,2, Michelle L. Colgrave5, and David J. Craik

From the1 Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia, 5 CSIRO Division of Livestock Industries, St. Lucia, Queensland 4067, Australia, and the 3 Australian Research Council Centre of Excellence in Plant Energy Biology and Centre for Comparative Analysis of Biomolecular Networks, University of Western Australia, Crawley, Western Australia 6009, Australia

Background: Cyclotides are defense-related cyclic plant peptides.

Results: Petunia cyclotides are encoded by novel cyclotide genes and occur in a discrete pattern in leaf architecture.

Conclusion: Novel cyclotides exist in the Solanaceae and are abundant in vascular tissues.

Significance: Cyclotide localization is consistent with an anti-herbivory role. Novel Solanaceae genes provide opportunities for expressing designer cyclic peptides in major crop species.

Cyclotides are a large family of plant peptides that are structurally defined by their cyclic backbone and a trifacta of disulfide bonds, collectively known as the cyclic cystine knot (CCK) motif. Structurally similar cyclotides have been isolated from plants within the Rubiaceae, Violaceae, and Fabaceae families and share the CCK motif with trypsin-inhibitory knottins from a plant in the Cucurbitaceae family. Cyclotides have previously been reported to be encoded by dedicated genes or as a domain within a knottin-encoding PA1-albumin-like gene. Here we report the discovery of cyclotides and related non-cyclic peptides we called “acyclotides” from petunia of the agriculturally important Solanaceae plant family. Transcripts for petunia cyclotides and acyclotides encode the shortest known cyclotide precursors. Despite having a different precursor structure, their sequences suggest that petunia cyclotides mature via the same biosynthetic route as other cyclotides. We assessed the spatial distribution of cyclotides within a petunia leaf section by MALDI imaging and observed that the major cyclotide component Phyb A was non-uniformly distributed. Dissected leaf mid-vein extracts contained significantly higher concentrations of this cyclotide compared with the lamina and outer margins of leaves. This is the third distinct type of cyclotide precursor, and Solanaceae is the fourth phylogenetically disparate plant family to produce these structurally conserved cyclopeptides, suggesting either convergent evolution upon the CCK structure or movement of cyclotide-encoding sequences within the plant kingdom.

Cyclotides are a family of backbone-cyclized plant peptides first discovered in Oldenlandia affinis from the Rubiaceae plant family but since found in a growing number of plants from the Violaceae, Cucurbitaceae, and Fabaceae families (1). Cyclotides are presumed to have a role in plant defense, given reports that ascribe insecticidal (2), molluscicidal (3), or anthelmintic (4) activities to isolated peptides. Since their initial discovery as the active constituents of a uterotonic traditional medicine (5), a host of other bioactivities have been attributed to cyclotides, including anti-HIV (6), cytotoxic (7), and neurotensin inhibitory activity (8).

The definitive structural feature common to cyclotides is the cyclic cystine knot (CCK) motif in which three disulfide bonds are entwined in a knotted conformation such that one disulfide bond is threaded through an opening bounded by two sections of the peptide backbone and the two disulfide bonds constraining them (9). The cystine knot has been demonstrated to be the feature that confers most of their stability at high temperatures, in extremes of pH, and against proteolytic enzymes (10, 11). The CCK motif is very tolerant to sequence variation of the cysteine residues, as exemplified by the observation that it occurs in two cyclic trypsin inhibitors, MCoTI-I and MCoTI-II (12), from a Cucurbitaceae plant that differ substantially in extremes of pH, and against proteolytic enzymes (10, 11). The CCK motif is very tolerant to sequence variation of the non-Cys residues, as exemplified by the observation that it occurs in two cyclic trypsin inhibitors, MCoTI-I and MCoTI-II (12), from a Cucurbitaceae plant that differ substantially in enzyme and host specificity. Novel Solanaceae genes provide opportunities for expressing designer cyclic peptides in major crop species.

*This work was supported by an Australian Research Council (ARC) grant for cyclotide analysis (Grant DP0984390), an ARC Grant for development of MALDI-MSI in plants (Grant DP0985873), and equipment purchased under ARC Linkage Grant LE0775603 housed in the ARC Centre of Excellence in Plant Energy Biology and Centre for Comparative Analysis of Biomolecular Networks, University of Western Australia.

1 An ARC Queen Elizabeth II Fellow (Grant DP0879133) and the John S. Mattick Fellowship.
2 An ARC Australian Professorial Fellow (Grant DP0771156).
3 A National Health and Medical Research Council Professorial Fellow. To whom correspondence should be addressed. Tel.: 61-7-3346-2019; E-mail: d.craik@imb.uq.edu.au.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EMBL Data Bank with accession number(s) JQ886398, JQ886399, and JQ886400.

The protein sequence data reported in this paper will appear in the UniProt Knowledgebase under the accession numbers B3EWH5, B3EWH6, B3EWH7, and B3EWH4.

The abbreviations used are: CCK, cyclic cystine knot; MSI, mass spectrometric imaging; EST, expressed sequence tag; ACN, acetonitrile; RACE, rapid amplification of cDNA ends; AEP, asparaginyl endopeptidase; SFTI, sunflower trypsin inhibitor.
Novel Genes and Vascular Localization of Solanaceae Cyclotides

Several recent studies have demonstrated the suitability of the CCK framework as a stable drug design scaffold, exemplified by the synthesis of modified cyclotides to incorporate bioactive peptide epitopes that would otherwise have short in vivo half-lives. Examples include cyclotide-based vascular endothelial growth factor-A (VEGF) agonists (15) or antagonists (16) and inhibitors of tryptase β from human mast cells (17). These studies highlight the potential value cyclotides have as peptide therapeutics and provide an impetus for investigating their biosynthesis in plants, potentially opening new opportunities for the expression of “designer” cyclotides with pharmaceutical traits in plants.

In Rubiaceae and Violaceae plants, cyclotides are products of dedicated genes that comprise an endoplasmic reticulum signal sequence and a pro-region, followed by up to three cyclotide domains, each flanked by an N-terminal pro-domain and a C-terminal tail (18, 19). Recently, we reported the occurrence of cyclotides in the Fabaceae plant Clitoria ternatea (20), and subsequently it was demonstrated that the Fabaceae cyclotides are encoded within a PA1b-like albumin where the cyclotide has “replaced” the first of its usual two domains (21, 22). Typical Fabaceae albumin-1 genes encode a PA1 pro-protein that is post-translationally cleaved to liberate PA1b (a member of the knottin family) and PA1a albumins (23), whereas in the C. ternatea albumin-1 gene, a cyclotide domain has replaced the PA1b knottin domain. Despite being encoded within its unusual gene architecture, Cter M, the best characterized cyclotide from C. ternatea, shares the key structural feature of a CCK motif with cyclotides derived from dedicated cyclotide genes. Interestingly, another of the isolated peptides from C. ternatea is identical in primary sequence to a previously reported cyclotide, Psyle F from Psychotria leptotheysra from Rubiaceae (24).

Although their gene expression does not appear to be dynamically regulated (25), cyclotides are known to be differentially expressed within a plant. In Viola hederacea, cyclotide vh1 is specific for only root tissue (26), whereas in O. affinis, Oak4 expression and its encoded peptide kalata B2 were absent from root tissue (25). Recent work has demonstrated that GFP-tagged cyclotide precursors accumulate in plant cell vacuoles (27). Several studies have reported insecticidal activity in cyclotides (2, 21, 28) and provided the basis for further structure-activity studies (29), but little is known about the distribution of cyclotides within individual plant tissues.

Matrix-assisted laser desorption/ionization-mass spectrometric imaging (MALDI-MSI) is an analytical technique in which mass spectra are collected in a raster pattern across a tissue section to generate an average mass spectrum, which, when overlaid upon an image of the sample, can reveal the spatial distribution and relative abundances of analytes (30). MALDI-MSI (31) has been applied in the study of animal and human tissues as a research tool as well as in a medical diagnostic capacity in the study of disease pathology (32–34) and to monitor drug pharmacokinetics (35, 36). Recent examples of plant MALDI-MSI providing insights through spatial information include the peptide analytes of developing soybean cotyledons (37), secreted peptide hormones involved in plant development in Arabidopsis roots (38), and small-molecule glucosinolate derivatives involved in plant defense from Arabidopsis leaves (39).

Our discovery of cyclotides in Petunia arose from interrogating expressed sequence tag (EST) databases by tBLASTN with the Cter A peptide sequence, which yielded many matches to potential cyclotide-encoding transcripts. Here we describe the characterization of cyclotides in Petunia x hybrida following their isolation and tandem MS sequencing and report a novel architecture of their genes based upon cloning of three full-length cDNA clones and a wealth of other EST-derived sequences.

The discovery of cyclotides in the Solanaceae is significant and exciting because this plant family includes many crop species, including potato and tomato, two of the largest food crops by global yield with a combined world annual production of more than 450 million tons. Given the demonstrated potent insecticidal activity of isolated cyclotides (2, 21, 28), knowledge of the Solanaceae cyclotide gene architecture might enable their expression in important food crops to potentially provide crops protection from predation by herbivores. The combination of MALDI-MSI on cyclotide expression and localization in petunia and MS analysis of leaf region extracts provided evidence of non-uniform distribution of the major cyclotide mass consistent with location in the vasculature of the leaf. A vascular location is common for small molecule (39), physical (40), and peptidic (41–43) herbivory defense systems and would allow an additive role for cyclotides in reducing predation by herbivores. Exploitation of Solanaceae cyclotide genes might thus allow production of novel, ultrastable therapeutics, lead to the enhancement of the staple crops as “functional foods” (44–46), and/or reduce crop losses to insect attack.

Previous studies investigating the expression of the cyclotide-encoding gene Oak1 from Rubiaceae in the model plant Nicotiana benthamiana reported the production of mainly misprocessed peptides (27, 47). The discovery of a cyclotide-encoding gene from the Solanaceae has great potential to improve the value of N. benthamiana as a research tool to study cyclotide processing and also to study the effects of cyclotides in plant defense.

**EXPERIMENTAL PROCEDURES**

**Materials**—P. x hybrida seedlings were sourced from Pohlman’s Nursery (Gatton, Queensland, Australia). Solid phase extraction cartridges and reverse phase HPLC columns were from Grace Vydac. All solvents and enzymes were supplied as described previously (20).

**Extraction**—P. x hybrida plants were rinsed extensively with distilled water to remove soil prior to separation of the various plant tissues. Fresh leaf (8.0 g) and root (5.3 g) samples were lyophilized prior to ball-milling using a Retsch MM300 homogenizer in three 30-s bursts at 25 Hz. Powdered plant samples were subsequently extracted in 60% acetonitrile (ACN), 1% formic acid with vortexing and probe sonication. Crude extracts were then centrifuged in a benchtop centrifuge at 4000 × g, and the supernatants were collected and diluted with 1% formic acid to give final solvent extract concentrations of 10% ACN,
using a 100-nm particle size) and misted onto the surface of the dried sample slides at a concentration of 7 mg/ml in 50% ACN, 0.2% trifluoroacetic acid. Reduced and endoproteinase Glu-C-digested samples were subject to a cleanup step using C18 ZipTips (Millipore) to remove salts and elicit a solvent exchange from aqueous solution to 80% ACN, 1% formic acid. Samples (3 μl) were transferred to nanospray tips (Proxeon, ES380) and nano-electrospray ionization was induced with a voltage differential of 900 V applied to the tip on a QSTAR Pulsar i QqTOF mass spectrometer (Applied Biosystems). TOF spectra were collected over the range m/z 400–2000. Product ion spectra were collected (m/z 100–2000) using collision energy voltages ranging from 10 to 60 V. Both TOF and product ion data were acquired using Analyst QS 1.5 software, and tandem MS spectra were manually assigned.

Nano-LC-MS/MS—Reduced, carbamidomethylated, and endoproteinase Glu-C-digested (linearized) cyclotide-containing crude extracts were analyzed via LC-MS/MS with conditions as described previously (21). Cryosectioning—Petunia leaf tissue was prepared for MALDI imaging with minor changes to a method described previously (37). Samples were cryosectioned using a Leica CM3050 cryotome with chamber temperature set at −19 °C and object temperature set at −17 °C. Frozen leaf tissue was floated on the surface of optimal cutting temperature (OCT) medium applied to the cryotome chuck and paradermal (axial longitudinal) cryosections sampled at 15-μm thickness.

LC/MS Analysis—The relative quantitation of Phyb A (m/z 3069) among leaf parts was performed using a method described previously (39) with some modifications. Leaves were removed from P. x hybrida plants, flash-frozen in liquid N₂, and lyophilized. These leaves were subsequently dissected to yield midvein, lamina, and peripheral leaf tissue samples, which were then weighed in separate tubes, and 500 μl of water was added per mg of dried plant tissue. Sealed sample tubes were placed into a heater block set at 95 °C for 75 min, cooled to room temperature, and centrifuged at 4000 × g for 10 min. Sample supernatants were introduced to a QSTAR Pulsar i QqTOF mass spectrometer (Applied Biosystems) equipped with a Turbospray ionization source, using an Agilent 1100 binary HPLC system (Agilent). Reversed phase separation of peptide analytes was achieved using a linear gradient comprising solvent A (0.1% formic acid) and solvent B (90% ACN, 0.1% formic acid (aqueous)) at a flow rate of 200 μl/min applied to a Jupiter C18 300-Å column (Phenomenex) of dimensions 150 mm × 2.0 mm with a particle size of 5 μm. TOF spectra were collected over the range m/z 400–2000 and analyzed using Analyst QS 2.0 software.

Cloning of PETUNITIDE Genes—We used tBLASTN to search the NCBI dbEST database with the amino acid sequence of Fabaceae cyclotide Cter A (GVIPEGESCVFIPCISTVIGC-
Novel Genes and Vascular Localization of Solanaceae Cyclotides

SCKNKVCYRN). Several Petunia ESTs matched with Expect (E) values of ≤1e7, including FN039431, FN022914, FN016979, FN020412, FN022913, FN039432, FN017679, and FN017678. Further analysis of these ESTs indicated that they were likely to encode cyclotides. These sequences were used to retrieve similar Petunia EST sequences. An alignment of the following 32 EST sequences was made: JI334153, JI334152, JI361658, JI335590, JI335591, DC242826, FN035504, FN039432, FN001371, JI333116, FN002680, FN020915, FN004316, FN003549, FN001318, FN005840, FN004550, FN005530, FN035505, FN039431, FN022914, FN016979, FN020412, FN017678, FN020916, and FN021459. Based on this alignment, there appeared to be several different cyclotide-encoding genes in Petunia. Against this alignment, primers that would amplify complete cyclotide domains were designed for 5′- and 3′-RACE: JM532 (5′-CAT ATA TAT GCC CCT CTC CT-3′) for 5′-RACE bound downstream from the stop codon in 19 of 32 sequences aligned; JM533 (5′-GAC GCA CGC GTA ATG GAT-3′) for 3′-RACE bound 30 bp upstream of the region encoding the cyclotide domain in 23 of 32 sequences aligned; JM534 (5′-TCA CGT GTG TTT CCT ACC A-3′) for 5′-RACE bound 4 bp upstream of the region encoding the cyclotide domain in 15 of 32 sequences aligned; and JM535 (5′-TGG GTG TTT CCT ACC A-3′) for 3′-RACE bound 4 bp upstream of the region encoding the cyclotide domain but was specific for a different group of 6 among the 32 sequences aligned.

RNA was extracted from the leaves, flowers, and roots of *P. x hybrida* using phenol/chloroform extraction, and selective precipitation of RNA was performed using lithium chloride as described previously (49). Between 500 ng and 1 μg of total RNA was used to create 5′- and 3′-RACE libraries using the SMARTer RACE cDNA amplification kit (634923, Clontech) as per the manufacturer’s instructions. The three 5′-RACE libraries were PCR-amplified using JM532, whereas the three 3′-RACE libraries were PCR-amplified using JM533, JM534, and JM535. The 5′- and 3′-RACE products were cloned into pGEM-T (Promega), sequenced, and aligned. These partial sequences suggested that up to five different transcripts had been amplified. Using the transcript sequences tentatively named PETUNITIDE1 to -5, we designed the following primers in the 5′- and 3′-UTRs that would be specific and amplify the full ORF: JM549 (5′-CAT ACT CAG TGA TTT CCC ACC A-3′) bound to the 3′-UTR of the transcript tentatively named PETUNITIDE1; JM552 (5′-CAT ACT CAG CTA CAC ATA GTG C-3′) bound to the 3′-UTR of PETUNITIDE3; JM550 (5′-CAT ACC CAG TGA TTT TCC ACC A-3′) bound to the 3′-UTR of PETUNITIDE4; JM551 (5′-CAT ACC CAG TAA TTT CCT ACC A-3′) bound to the 3′-UTR of PETUNITIDE5; JM554 (5′-GCC AGC TAC ACA TAG TGC T-3′) bound to the 3′-UTR of most PETUNITIDE transcripts and is upstream from the binding sites of JM550-JM552; and JM553 (5′-TGG CAA AGA TAA TAC TTT CA-3′) bound to the 5′-UTR of both PETUNITIDE2 and PETUNITIDE4.

PCR amplification of the aforementioned 5′- and 3′-RACE libraries with these primers yielded products of the expected sizes that were subsequently cloned into pGEM-T and sequenced. This sequencing revealed three different PETUNITIDE transcript products each encoding a full ORF. For each PETUNITIDE transcript, at least three independent clones were obtained.

RESULTS

Searching for Cyclotide Sequences in Petunia Transcript Databases—Structurally homologous cyclotides have previously been characterized from plants of the Rubiaceae, Violaceae, and Fabaceae families, with the investigated species typically having been selected on the basis of an identified bioactivity. To search for potential cyclotide-encoding genes within publicly available bioinformatic data, Fabaceae cyclotide Cter A was used as a BLASTN search string to interrogate the EST database at NCBI. Numerous putative cyclotide-encoding ESTs from the genus Petunia matched the submitted protein sequence (summarized in Fig. 1, with their accession numbers under “Experimental Procedures”). We named these putative petunia cyclotides Phyb A through Phyb L in accordance with a previously established convention (50). PETUNITIDE1 to -3 and the related ESTs appeared to encode precursor proteins possessing an endoplasmic reticulum target signal and, toward the end, a cyclotide domain containing six cysteines of typical

![FIGURE 1. BOXSHADE alignment of cyclotide precursors in genus Petunia.](link)
spacing, the highly conserved Glu in loop 1, a proto-N-terminal Gly, and the usual proto-C-terminal Asx (i.e. Asn or Asp). This arrangement differs from cyclotide precursors from the Violaceae and Rubiaceae, which have longer regions between the signal peptide region and the mature cyclotide domain(s), making the PETUNITIDE proteins essentially the same size as very recently described precursors from the Rubiaceae plant Chassalia chartacea, which are much shorter than previously described cyclotide precursors (supplemental Fig. S1).

Commonly trailing the cyclotide domains’ proto-C-terminal Asx is an Ile or Leu located two residues downstream (at P2'). This residue is consistently observed among the corresponding regions of Violaceae, Rubiaceae, and Fabaceae cyclotide genes and appears to be an important residue for processing (51). In previously reported cyclotide genes, the amino acid at P1’ is typically a Gly, but the Solanaceous precursors exhibit either a Gly or Glu. We also observed that some ESTs encode PETUNITIDE proteins that are punctuated with stop codons immediately following their cyclotide domains (Fig. 1; encoding putative Phyb J, Phyb K, and Phyb L). What effect this has on peptide maturation remained to be determined by analyzing the peptide profile of petunia. A list of ESTs that would encode the putative mature cyclotides and acyclotides is given in supplemental Table S1.

Detection and Sequencing of Cyclotides and Acyclotides from *P. x hybrida*—To confirm the synthesis and accumulation of the predicted cyclic and acyclic peptides, we prepared extracts of various *P. x hybrida* tissues and analyzed these directly using MALDI-TOF MS. As shown in Fig. 2, A and D, both leaf and root extracts exhibited signals within the mass range m/z 2800–3700 typical of cyclotides. A dominant signal observed at m/z 3069 (peak A) appeared in both leaf and root extracts and was consistent with the mass of the predicted cyclotide domain of PETUNITIDE1. After reduction and alkylation, a mass 348 Da larger than peak A was observed at m/z 3417 (Fig. 2B), consistent with the alkylation of six cysteine residues. Following digestion with endoproteinase Glu-C, an additional increase in mass of 18 Da was observed. This signal corresponded to peak A/366 Da and appeared at m/z 3435 (Fig. 2C), consistent with a single peptide backbone cleavage, as would be expected for a cyclotide. Tandem MS characterization confirmed the identification of the peak at m/z 3435 as the bracelet cyclotide Phyb A with sequence SCVWIPCVSAAGCSCSNKICYRNGIGCGE, in agreement with the sequence of PETUNITIDE1 (Fig. 3A).

The other dominant peak observed in MALDI-TOF analyses of root extracts appeared at m/z 3388 (peak B), but no corresponding peak at +366 Da was observed in the endoproteinase Glu-C-digested sample. Apart from the peak observed at m/z...
3435 (Phyb A), the dominant signal in the spectrum from endoproteinase Glu-C-digested root extract was a peak at \( m/z 2978 \) (Fig. 2F). Examination of the TOF-MS spectrum of the reduced and alkylated extract revealed a dominant peak at \( m/z 3736 \) (Fig. 2E) corresponding to the addition of 348 Da (alkylation of the six Cys residues) to the peak at \( m/z 3388 \) in the native root extract. Following endoproteinase Glu-C digestion, the peak at \( m/z 3736 \) was noticeably absent, suggesting that proteolysis had occurred. A peak at \( m/z 3754 \) would be expected following cleavage of a cyclic peptide backbone (at Glu in loop 1). However, no such peak was observed, raising two possibilities: 1) that the peptide contained more than the single Glu in loop 1 and/or 2) that the peptide was linear and was therefore cleaved into more than one fragment. Under MS/MS conditions, cyclic peptides typically do not fragment as readily as linear peptides. Extensive fragmentation of the precursor at \( m/z 3736 \) (Fig. 3B) in tandem MS indicated its peptide backbone to be non-cyclic. The assigned peptide sequence was homologous to the cyclotide-like gene product predicted by EST FN005530, with sequence pQSISCAESCVIPCATSLIGCSCVNSRCIYSK, which we named Phyb M, and was found to incorporate a pyroglutamyl modification at its N terminus. The dominant peak at \( m/z 2978 \) observed after endoproteinase Glu-C digestion was subjected to tandem MS analysis, revealing its identity as the C-terminal portion of Phyb M.

During LC-MS/MS analysis of reduced, alkylated, and endoproteinase Glu-C-digested root extracts, another acyclic peptide was observed (Fig. 3C) with a parent ion mass and fragmentation pattern matching the cyclotide sequence encoded by EST FN001318 (STDCGEPCVYIPCTITALLGCSCLNKVCVRP) and which we named Phyb K. The masses of the characterized as well as putative cyclotides are reported in Table 1. Fig. 4 illustrates an alignment of Phyb A with the cyclotide sharing the highest sequence homology, cycloviolacin O17 from Viola odorata, along with the petunia cyclotides Phyb K and Phyb M, demonstrating the conserved cysteine spacing.

Judging from LC-MS analyses, the abundance of petunia cyclotides in source plant material was within the range previously reported for cyclotides in Viola odorata and O. affinis (52), with Cte A in wet leaf material estimated to be 30.0 µg/g, whereas Cte K and Cte M were present in wet root material at 2.3 and 7.6 µg/g, respectively.

**Determination of Cyclotide Distribution in Petunia Leaf Tissue**—To examine the spatial distribution of petunia cyclotides within plant tissue, we used MALDI-MSI to analyze a paradermal leaf section, generating an ion intensity map.
shown in the average mass spectrum in Fig. 5A, numerous peaks were detected in the range m/z 3000–3600. Analysis of leaf extracts showed a single dominant peak at m/z 3069, which was sequenced and named Phyb A (Fig. 2A), and this peak was observed in the MALDI-MSI experiment along with peaks at m/z 3110, 3426, and 3463. Apart from m/z 3069, none of these other masses corresponded to isolated or predicted cyclotides (Figs. 1 and 2). LC-MS/MS analysis of leaf extracts was undertaken to determine if the MALDI-MSI peaks observed at m/z 3110, 3426, and 3463 were cyclotides; however, only a precursor at m/z 3424 (corresponding to m/z 3426 average mass in MALDI-MSI) was observed. Limited fragmentation of this precursor was observed, but following a reduction step, a mass increase of 2 Da was observed (supplemental Fig. S2, A and B), indicating the presence of a single intramolecular disulfide bond, and leading to extensive fragmentation in subsequent tandem MS (supplemental Fig. S3). The sequence resulting from manual de novo mass spectral interpretation (DEEP-KRGTPAEKKKYSVCTNPARTICRY) was used in a BLAST search and found to be consistent with a translated EST (FN008610) from petunia encoding a sequence homologous to known precursors. This is the fourth architecturally distinct cyclotide from the Solanaceae plant family. These peptides arise from the shortest known cyclotide precursors and are distinct from previously known precursors. This is the fourth architecturally distinct precursor from which cyclotides emerge, provoking interesting questions about the evolutionary origin of their structurally identical CCK framework peptides. The new precursors present opportunities for designing synthetic peptides capable of being cyclized efficiently in planta for a range of agricultural or pharmaceutical applications. Furthermore, we have confirmed enrichment of a cyclotide in the vasculature of leaves, a finding that is consistent with a proposed general role of cyclotides in herbivory defense.

**TABLE 1**

**Sequence alignment of cyclotides observed in *P. x hybrida* extracts**

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Amino acid sequence&lt;sup&gt;a,b&lt;/sup&gt;</th>
<th>Expt. m/z</th>
<th>Expt. Da</th>
<th>Theor. mass</th>
<th>Error&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Theor. mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phyb A</td>
<td>GIGGSGSCWVIPC-VSAAIAGCSKCNKC-YRN</td>
<td>3434.47</td>
<td>3434.47</td>
<td>3434.55</td>
<td>-0.22</td>
<td>3068.29</td>
</tr>
<tr>
<td>Phyb K</td>
<td>STDGEPGVYIPCTITALLGCSCNKVC-VRP</td>
<td>1200.89</td>
<td>3599.67</td>
<td>3599.71</td>
<td>-0.11</td>
<td>3251.46</td>
</tr>
<tr>
<td>Phyb M</td>
<td>pQSISCESCWVIPC-STLIGCSKNSRCY</td>
<td>1246.17</td>
<td>3735.51</td>
<td>3735.73</td>
<td>-0.58</td>
<td>3387.47</td>
</tr>
</tbody>
</table>

<sup>a</sup> Leu and Ile in Phyb M were assigned on the basis of homology to Phyb L.

<sup>b</sup> Phyb K and Phyb M are acyclic.

<sup>c</sup> Expt., experimental; Theor., theoretical; linearized peptide masses after reduction and carbamidomethylation of cysteines.

**DISCUSSION**

Here we report the discovery and characterization of cyclotides from *P. x hybrida* of the agronomically important Solanaceae plant family. These peptides arise from the shortest known cyclotide precursors and are distinct from previously known precursors. This is the fourth architecturally distinct precursor from which cyclotides emerge, provoking interesting questions about the evolutionary origin of their structurally identical CCK framework peptides. The new precursors present opportunities for designing synthetic peptides capable of being cyclized efficiently in planta for a range of agricultural or pharmaceutical applications. Furthermore, we have confirmed enrichment of a cyclotide in the vasculature of leaves, a finding that is consistent with a proposed general role of cyclotides in herbivory defense.

**Existence of Cyclotides in the Solanaceae**—The discovery of cyclotides within the Solanaceae plant family is an exciting and important development, given the significance of this plant family to human nutrition, and follows the recent landmark discovery of cyclotide genes in a member of the Fabaceae family (21, 22). The Solanaceae is host to more than 3000 species,
including staple crops, such as *Solanum tuberosum* (potato) and *Solanum lycopersicum* (tomato), which constitute two of the most important vegetable crops cultivated, with combined worldwide annual production exceeding 450 million tons.

Plant species previously investigated in the search for cyclotides have typically been selected on the basis of an identified bioactivity in their extracts, such as uterotonic activity in *O. affinis* (53) and *C. ternatea* (20), anti-HIV activity in *Pali-courea condensata* (54), hemolytic activity in *Viola* extracts, and trypsin inhibitory activity in *Momordica cochinchinensis* (12). In the current study, we examined *P. x hybrida* following the identification of ESTs from the genus *Petunia* via a database search. Further experiments confirmed petunia cyclotides to be the products of dedicated genes with a novel precursor structure.

**Structural and Evolutionary Implications of the Novel Precursors from Genus Petunia**—The sequences of three cyclotide-encoding genes, named *PETUNITIDE1* to -3, are shown in Fig. 1 alongside the translated amino acid sequences deriving from the BLAST-matched ESTs, where they encode precursor proteins of 79 residues comprising an endoplasmic reticulum signal sequence, a pro-region of 15 residues, a single cyclotide-encoding domain, and a six-residue C-terminal tail sequence. A distinguishing feature of Solanaceae cyclotide precursors is their relatively short (15 residue) N-terminal pro-regions compared with those from Rubiaceae (22–69 residues) and Violaceae (28–45 residues) cyclotide genes. In combination with their short C-terminal tails, the Solanaceae cyclotides are encoded by relatively short cyclotide-encoding precursors, similar in size to recently reported atypical cyclotide precursors from Rubiaceae (55).
Novel Genes and Vascular Localization of Solanaceae Cyclotides

Some of the BLAST-matched Petunia ESTs appeared to terminate with stop codons directly C-terminal to the cyclotide-encoding domains, indicating that acyclic cyclotides ("acyclotides") might be produced in planta. Accordingly, we characterized a peptide matching one of these predicted acyclic ESTs.

The first acyclotide characterized was violacin A from the Violaaceae plant Viola odorata, which we referred to at the time as a "linear cyclotide" (56). Later, in O. affinis, the transcript Oak9 was found to encode kalata B20-lin, an acyclotide seemingly arising from a single nucleotide change that introduces a stop codon (25). In two recent studies of Rubiaceae plants Hedysotis biflora and Chassalia chartacea, panels of novel "linear cyclotides" were characterized and referred to as "uncyclotides" (55, 57). We prefer the term "acyclotide" for the following two reasons. 1) This is in keeping with established practices in nomenclature of organic compounds as either cyclic or acyclic (58). 2) Selectional restrictions on English language pre-fixes mean that the "un-" prefix can be taken to confer two meanings (cf. "unlockable"), and when added to the word "cyclic," the resultant "uncyclic" can be construed to convey that the item being described is "not cyclic" or alternatively that it is "no longer cyclic." Thus, the "a-" prefix is unambiguous and conveys only one meaning to "acyclic": that the item is "not cyclic." Interestingly, in some cases, the acyclotides have biological activity comparable with that of their cyclic counterparts (53), but in most cases, the linear homologues are devoid of the activity of the cyclic forms (59, 60).

Fig. 7 illustrates a comparison of the gene structures of representative cyclotide and related knottin-or acyclotide-encoding sequences, in which PETUNITIDE genes, in terms of overall structure and size, can be seen to bear the most similarity to recently characterized CHASSATIDE genes identified within Rubiaceae plant Chassalia chartacea (55). Despite the lack of peptide evidence for cyclotide-like sequences in Poaceae, they are likely to be produced as acyclic peptides due to their truncation by a predicted stop codon, as illustrated for "Zea mays B," and in this way bear similarity to genes encoding putative petunia acyclotides, including Phyb J, K, and L. Peptide evidence was found for hedyotide B2, an acyclotide found in Rubiaceae plant H. biflora, and the gene encoding it was found to have been truncated at the C terminus of the cyclotide domain by a stop codon (57), with the remainder of the gene exhibiting homology to other Rubiaceae cyclotide genes as indicated in Fig. 7, A and B. Other acyclotides have also been characterized from Rubiaceae plants, including kalata B20-lin from O. affinis (25) and Psyle C from P. leptothyrsra (24).

A single example of a linear cyclotide, violacin A, has been described in the Violaaceae plant V. odorata (56). However, the gene encoding violacin A is unique compared with other acyclotide-encoding genes in that the premature stop codon does not appear after the entire peptide domain but rather appears to truncate an otherwise complete cyclotide gene. In this case, the nucleotide sequence immediately following the stop codon is replete with sequence that would encode typical proto-C-terminal and CTR amino acids, suggesting that violacin A might be the result of a single nucleotide polymorphism.

Solanaceae cyclotides are encoded by PETUNITIDE genes that incorporate sequence motifs considered integral for cyclotide biosynthesis and backbone cyclization in previously described cyclotide genes (51), including a proto-C-terminal Asx followed by a hydrophobic amino acid two residues C-terminal (e.g. –Asx-Xaa-Leu/Ile/Val–). It has been posited that an asparaginyl endopeptidase (AEP) would be the logical candidate enzyme driving cyclotide biosynthesis (47, 51), due to the demonstrated in vitro cleavage and transpeptidation (ligation) activity of jackbean AEP to produce mature concanavalin A (61) and its activity at a wide range of Asx-Xaa bonds (62). A large body of work has demonstrated that AEP can mature seed storage globulins and albumins (63–68). In sunflowers (Asteraceae), a gene encoding a napin-like preproalbumin storage PawS1 albumin gives rise to mature seed storage albumin as well as small backbone-cyclized trypsin inhibitor embedded upstream of the albumin (69). Following transformation of PawS1 into an Arabidopsis aep null mutant, it was determined that AEP was required for cleavage reactions at the proto-N terminus of SFTI, the proto-C terminus of SFTI, and the proto-N terminus of the PawS1 small albumin subunit (69), and based on this, AEP was proposed as a good candidate enzyme for mediating ligation of N and C termini of SFTI-1. This might occur through attack of the thioester acyl intermediate of AEP.
by the freed glycine of SFTI-1, held close to the thioester by the disulfide bond (61).

Recently, it was discovered that the butterfly pea (C. ternatea) contains pea albumin-1-like genes in which a cyclotide domain has “replaced” the first of the PA1 domains, and this cyclotide domain is trailed by residues that would enable bio-processing via the same AEP-mediated mechanism (21, 22).

One of the peptides characterized in the current study, Phyb M, with the sequence pQSISCAESCVWIPCATSLIGCSCNVN-SRCIYSK is supported by EST FN005530 and incorporates a post-translational N-terminal pyro-Glu modification. The first pyroglutamyl modification of a linear cyclotide was reported recently in hedyotide B4 from H. biflora (57). We did not detect peptide masses corresponding to non-pyro-Glu Phyb M, which suggests that it is not a degradation artifact from a longer mature peptide. N-terminal pyro-Glu peptides are known to be more resistant to degradation than their corresponding N-terminal Gln homologs (70), so the incorporation of a modified N-terminal residue in lieu of backbone cyclization may be an alternative strategy to provide enhanced stability toward exopeptidase activity. Thus, Phyb M bridges an evolutionary gap between Phyb K, an acyclotide with a free N terminus, and Phyb A with its “complete” CCK motif. The discovery of all three peptide forms from P. x hybrida may represent “evolution in progress.”

Despite the isolation of PETUNITIDE2 and PETUNITIDE3 transcripts, we found no mass spectrometric evidence for peptide masses for the cyclotides they would encode (Phyb B and Phyb C; Fig. 1). The sequences of these peptides as well as those encoded within the identified ESTs and the rest of the peptides characterized in this study are mostly homologous to many previously described cyclotides and incorporate permutations of previously observed amino acids within loop regions. An exception to this is the translated sequence of ESTs FN020915 and FN020916 (GIPCGGSCVWIPCISGVCSCSNKIC-
Novel Genes and Vascular Localization of Solanaceae Cyclotides

YRN), in which the absolutely conserved Glu in loop 1 (Fig. 4), present in all previously characterized cyclotides, is replaced with a Gly residue.

The potentially wider prevalence of cyclotides among Solanaceae plants remains to be elucidated; however, BLAST searches using full-length PETUNITIDE sequences to query all GenBank™ nucleotide sequences, including EST databases, revealed only matches to the Petunia ESTs reported in this study. This search confirms the uniqueness of the precursor sequence, especially considering the depth of EST coverage among members of the Solanaceae (e.g. 334384 in tobacco, 297142 in tomato, 249761 in potato, and 118054 in capsicum), and suggests that cyclotides evolved independently within the Solanaceae.

Vascular Localization and Functional Significance—Trabi et al. (26) investigated the tissue-specific distribution of a panel of cyclotides from Viola hederacea by comparing LC/MS profiles of separate tissue extracts and demonstrated that cyclotides are differentially expressed among plant tissues. This phenomenon was observed in a subsequent study of cyclotide localization in O. affinis plant tissues, which, in addition to examination of extracted peptides via LC/MS, observed no cDNA encoding kalata B2 in root tissue (25). Complementary to these studies, recent work examined the subcellular location of cyclotides during their biosynthesis, the results of which indicate that they are processed and accumulate within plant cell vacuoles (27). However, despite these advances, details on the intratissue distribution of cyclotides are lacking.

To examine the localization of cyclotides within petunia leaves, we analyzed a tissue section using MALDI-MSI and observed a number of peptide masses appearing in the mass range diagnostic of cyclotides (Fig. 5A). One of the signals observed was consistent with Phyb A and appeared to correlate with the vascular structures of the prepared leaf section (Fig. 5C). Through LC/MS analysis of dissected leaf extracts, relative quantitation of Phyb A was assessed in midvein, laminar, and peripheral leaf areas. Phyb A was found to exist at ~2-fold higher concentrations within midvein tissue versus laminar or peripheral leaf tissue extracts (Fig. 6B). This distribution was unique compared with the trends observed for 16 control m/z signals, which were present either in equivalent abundance across the three leaf tissue areas or in higher abundance within laminar and/or peripheral leaf tissue extracts compared with midvein extracts (supplemental Fig. S4). The size and direction of the -fold change in Phyb A abundance might be of functional significance in the context of plant defense, given a previous study of Arabidopsis thaliana in which it was demonstrated that non-peptidic plant defensive glucosinolates were enriched at the midvein and the outer lamina of leaves (39). In the Arabidopsis study, it was further observed that Helicoverpa feeding preference could be influenced by as little as 1.3-fold relative changes in the concentration of indol-3-ylmethylglucosinolate, the major glucosinolate present. The observed localization pattern for Phyb A (m/z 3069) primarily in the vasculature of the leaf section mirrors the glucosinolate study and places Solanaceous cyclotides in the right location in leaves to be potential modulators of insect herbivory.

One of the limitations of MALDI-MSI is its inherent limited dynamic range, which is instrument-, matrix-, and analyte-dependent. Few studies have quantified the dynamic range of this technique, but a recent investigation demonstrated linearity of signal intensity increasing with analyte concentration from the limit of quantitation (femtomolar) over less than 2 orders of magnitude (71). Given the relative abundance of Cter A compared with other putative cyclotide signals in leaf extract (Fig. 2A), it is therefore unsurprising that low abundance putative cyclotide signals in the extract were not detected during MALDI-MSI, where the sample had not been deconvoluted through extraction, and the analyte was rather presented to the instrument in a complete, complex sample matrix.

Additional signals were observed during MALDI-MSI analysis of the leaf section that did not correspond to any of the calculated peptide masses from PETUNITIDE genes or translated EST sequences. Signals at m/z 3426 and 3463 appeared to be abundant in areas of the leaf section distinct from m/z 3069 or 3110. Tandem MS of the major peak observed at m/z 3426 (average) in the MALDI-MSI experiment (m/z 3424 monoisotopic in ESI) following reduction of a single disulfide bond permitted de novo sequencing and its further identification as nuclear PSII-T 5-kDa protein. Although a fragment of a homologous PSII-T protein has been sequenced from spinach (72), our work demonstrates the first mass spectral evidence of any nuclear PSII-T protein and describes a previously unreported disulfide bond. Given the conserved nature of the cysteines in homologous nuclear PSII-T proteins (not shown) a disulfide bond could be expected in all such proteins. The even distribution of PSII-T among all leaf areas, as shown in supplemental Fig. S4O, is consistent with the ubiquitous nature of photosynthetic proteins in leaf tissue.

Fig. 5G illustrates a difference map of signal intensities for the m/z 3426 (PSII-T) and 3069 (Phyb A) peptides and reflects the differential spatial expression of the two masses. The differential localization of the various m/z signals from the MALDI-MSI experiment indicated that the intensity of any particular signal was not significantly influenced by cell size or density and that signals for each peptide were heterogeneous across the tissue sample. This validates the sample preparation methodology and the suitability of the technique as a whole and demonstrates its ability to provide information on the spatial relative abundance of peptide analytes.

CONCLUSIONS

Here we have demonstrated the first evidence that cyclotides and acyclotides exist within the Solanaceae plant family as the products of a novel precursor structure. This work complements previous characterizations of cyclotide-encoding genes from Violaceae, Rubiaceae, and Fabaceae plant families (73). Analysis of the Solanaceae cyclotide (PETUNITIDE) genes implicates AEP in their proto-C-terminal processing, consistent with purported biosynthetic pathways of cyclotides in the literature (47) and consistent with the demonstrated requirement for AEP in processing the cyclic sunflower trypsin inhibitor SFTI-1 (69). Petunia cyclotides and their encoding genes have residues trailing the proto-C terminus consistent with those shown previously to be important for their correct bio-
Novel Genes and Vascular Localization of Solanaceae Cyclotides

... synthesis. Similar to CHASSATIDE genes from Rubiaceae, PETUNITTIDE genes are more compact than previously known cyclotide precursors. Subtle differences between the sequence motifs flanking the mature cyclotide sequences in Solanaceae and phylogenetically distinct Rubiaceae or Violaceae precursors might explain the low yields of cyclic products following expression of both natural and designed cyclotides in Solanaceae plants (47, 51). Thus, the discovery of novel cyclotide-encoding genes within the Solanaceae family might enable their application as an alternative option for circular peptide production compared with known cyclotide genes. PETUNITTIDE genes might also be employed to enhance crop protection within Solanaceae species important to human nutrition, such as potato, capsicum, and tomato, through genetic incorporation of custom cyclotide and/or acyclotide-encoding domains.

Our data demonstrate that cyclotides associate with the vascular features of petunia leaf tissues, which aligns with previously characterized small molecule and peptic mediators of plant defense. Examples include glucosinolate (39) precursors of toxic cyanocompounds in Arabidopsis, terpenes involved in squirt-gun defenses in Bursera sp. (40), pumpkin fruit trypsin inhibitor (41), cysteine protease inhibitors in maize (42), and defensins in capsicum (43). Hence, the localization of increased concentrations of cyclotides in these areas could modulate herbivore feeding behavior and contribute to plant defense. This work adds to the known pool of cyclotide-producing plant families and provides an impetus for the further exploration of Solanaceae species for cyclotides. Judging from the variation in cyclotide gene structures now described, it seems likely that further significant variations will be discovered in yet to be described cyclotide-containing plant families. This combined knowledge will be crucial to understanding their evolutionary origins as either the products of convergent evolution or potentially the action of transposable elements.

Acknowledgments—We thank Alun Jones and the Molecular and Cellular Proteomics Facility (University of Queensland, Institute for Molecular Bioscience) for access to mass spectrometry instrumentation and Dr. Anne Rae (CSIRO Plant Industries) for helpful discussions.

REFERENCES