Distribution of circular proteins in plants: large-scale mapping of cyclotides in the Violaceae

Robert Burman[1], Mariamawit Y. Yeshak[1,2], Sonny Larsson[1], David J. Craik[3], K. Johan Rosengren[1,4] and Ulf Göransson[1]*

[1] Division of Pharmacognosy, Department of Medicinal Chemistry, Uppsala University, Uppsala, Sweden, [2] Department of Pharmacognosy, School of Pharmacy, Addis Ababa University, Addis Ababa, Ethiopia, [3] Craik Lab, Chemistry and Structural Biology Division, Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia, [4] Laboratory for Peptide Structural Biology, School of Biomedical Sciences, The University of Queensland, Brisbane, QLD, Australia

During the last decade there has been increasing interest in small circular proteins found in plants of the violet family (Violaceae). These so-called cyclotides consist of a circular chain of approximately 30 amino acids, including six cysteines forming three disulfide bonds, arranged in a cyclic cystine knot (CCK) motif. In this study we map the occurrence and distribution of cyclotides throughout the Violaceae. Plant material was obtained from herbarium sheets containing samples up to 200 years of age. Even the oldest specimens contained cyclotides in the preserved leaves, with no degradation products observable, confirming their place as one of the most stable proteins in nature. Over 200 samples covering 17 of the 23–31 genera in Violaceae were analyzed, and cyclotides were positively identified in 150 species. Each species contained a unique set of between one and 25 cyclotides, with many exclusive to individual plant species. We estimate the number of different cyclotides in the Violaceae to be 5000–25,000, and propose that cyclotides are ubiquitous among all Violaceae species. Twelve new cyclotides from six phylogenetically dispersed genera were sequenced. Furthermore, the first glycosylated derivatives of cyclotides were identified and characterized, further increasing the diversity and complexity of this unique protein family.

Keywords: cyclotide, Violaceae, plant protein, plant peptide, stable proteins

INTRODUCTION

During recent decades plants have increasingly been recognized as an important source of bioactive proteins. As well as being attractive systems for protein production (Yusibov et al., 2011; Sack et al., 2015), plants provide unique protein scaffolds that are suitable for applications in drug design and agriculture (Boothe et al., 2010; Craik et al., 2010; Padovan et al., 2010; Göransson et al., 2012). Of particular interest from this applied perspective are small cysteine-rich proteins, which are highly stable and adaptable to provide many biological activities. One such family of plant proteins that possesses unique properties is the cyclotides (cyclo-peptides) that are found among species of violets (Violaceae). Cyclotides comprise a circular chain of approximately 30 amino acids, including six conserved cysteines that form the three disulfide bonds creating the cyclic cystine knot (CCK) motif (Craik et al., 1999). This motif imparts cyclotides with exceptional thermal, enzymatic, and chemical stability (Colgrave and Craik, 2004). Apart from the conserved cysteines, the sequences

of cyclotides are highly diverse, suggesting that the rigid scaffold can be adopted to accommodate a wide range of biological activities, making them ideal candidates for protein engineering ventures (Chan et al., 2011; Aboye et al., 2012; Wong et al., 2012; Koehbach et al., 2013b; Wang et al., 2014).

A number of biological activities have already been described for naturally occurring cyclotides, including cytotoxic effects (Lindholm et al., 2002; Svangård et al., 2004; Herrmann et al., 2008), hemolytic (Daly and Craik, 2000), anti-HIV (Gustafson et al., 2004; Wang et al., 2008a), and anti-fouling properties (Göransson et al., 2004). They have also been shown to have insecticidal (Jennings et al., 2001, 2005; Barbeta et al., 2008), nematocidal (Colgrave et al., 2008), and antimicrobial (Tam et al., 1999; Pränting et al., 2010; Ovesen et al., 2011) properties, suggesting that their biological function is probably linked to host defense of plants. Recent studies have shown that cyclotides interact with cell membranes (Svangård et al., 2007; Huang et al., 2009; Burman et al., 2011; Henriques and Craik, 2012), a phenomenon consistent with their broad range of biological effects. To date, more than 300 cyclotides have been described and they are generally divided into two main subfamilies, namely the Möbius and the bracelet subfamilies, which are characterized by the presence or absence, respectively, of a conserved *cis* proline (Craik et al., 1999). The structure and amino acid sequence of representative members from these two subfamilies are shown in **Figure 1**.

Despite extensive screening, until recently cyclotides had only been found in two phylogenetically distant families: the Violaceae and the Rubiaceae, with the exception of two more distantly related trypsin inhibitor peptides also carrying a CCK motif, found in the Cucurbitaceae (Hernandez et al., 2000; Felizmenio-Quimio et al., 2001; Heitz et al., 2001; Gruber et al., 2008; Koehbach et al., 2013a). However, cyclotides have now also been identified in members of the Fabaceae (Poth et al., 2010; Nguyen et al., 2011) and Solanaceae (Poth et al., 2012), and are probably also present in other families yet to be characterized.

Although the Violaceae has been identified as one of the major sources of cyclotides and cyclotide diversity, only a limited number of species have been examined. The majority of studies have been of plants of the genera *Viola* and *Hybanthus* (Broussalis et al., 2001; Simonsen et al., 2005; Ireland et al., 2006; Herrmann et al., 2008; Hashempour et al., 2013) with only a few studies having been conducted on other Violaceae genera, despite there being many of them (Hallock et al., 2000; Trabi et al., 2009; Burman et al., 2010). Consequently, sampling has been too narrow and biased for definitive conclusions to be drawn about the distribution of cyclotides among the Violaceae.

The family Violaceae comprises 23–31 genera and approximately 1050 species worldwide (Hekking, 1984, 1988; Munzinger and Ballard, 2003; Wahlert and Ballard, 2012; Wahlert et al., 2014; Tokuoka, 2008). Most genera contain a small number of species or are monotypic and restricted to the New World or Old World tropics. The three largest classical genera, *Viola*, *Hybanthus* sensu lato (including *Cubelium, Pigea, Pombalia,* and other proposed segregate genera), and *Rinorea* sensu lato (including *Dioryktandra* and other proposed segregated genera) account for more than 90% of Violaceae

species (Tokuoka, 2008; Wahlert et al., 2014). The largest of these three genera is *Viola*, the "true violets," which are characteristically herbs with bilaterally symmetrical, spurred flowers. The other genera have radially symmetrical flowers and are lianas, shrubs, or either large or small trees. The classification of Hekking recognizing three subfamilies are not supported by phylogenetic analyses based on nucleotide sequence data, subsuming the subfamilies Leonioideae and Violoideae into a single entity under the latter name (Tokuoka, 2008; Wahlert et al., 2014).
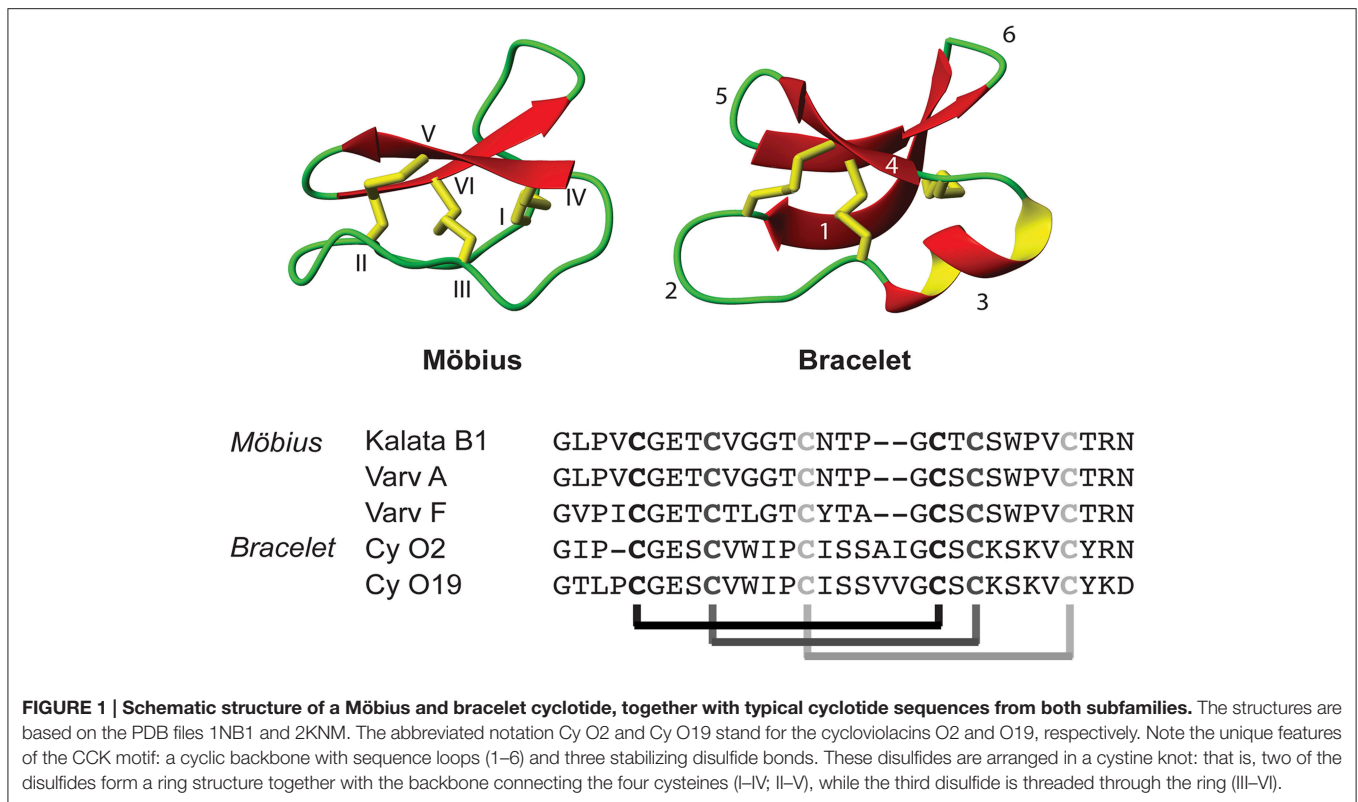
In the current study we undertook an extensive analysis of a broad collection of samples from within the family Violaceae. The main source of samples was from plant collections in three major Swedish herbaria. Due to the age and rarity of these samples it was necessary to make some adjustments to the current used methods to be able to identify cyclotides from the minute amounts of plant material that were available. Using this strategy we mapped the occurrence and distribution of cyclotides among the Violaceae on a large scale, analyzing more than 200 samples, representing approximately 1/6 of all species and the majority of genera. We show that certain cyclotides are more commonly expressed and occur in many species whereas other cyclotides are species specific. We conclude that cyclotides are ubiquitous in the Violaceae family of plants, that the chemical diversity of the cyclotide family is further expanded by the presence of glycosylated cyclotides, and demonstrate the exceptional stability of cyclotides by their presence as stable proteins in 200-year-old herbarium specimens.

RESULTS

A large-scale chemical screen was conducted on 143 species representing 17 of the possible 31 genera of Violaceae from all parts of the world. Sampling was focused in the three major classical genera, *Viola*, *Hybanthus* sensu lato, and *Rinorea* sensu lato, in order to include a significant sampling of all subgroups and to prevent problems that might have arisen from some genera being polyphyletic e.g., *Hybanthus* (Tokuoka, 2008; Wahlert et al., 2014). An overview of the number of samples screened from each genus, and the geographical origin of samples is shown in **Figure 2**. A complete list of species is found in the **Supplementary Table 1**.

Screening Procedure Development

Plant material was sampled at three major herbaria in Sweden, namely those located in Uppsala (UPS), Gothenburg (GB), and Stockholm (S), each of which possesses extensive plant collections. We strived to ensure that the sampling covered all parts of the phylogenetic tree of Violaceae. As a large number of samples were to be analyzed, but with only limited amounts of plant material for each species, we developed a simple, fast and sensitive method for cyclotide identification based on liquid chromatography-mass spectrometry (LC-MS). Briefly, plant material was first extracted with 60% acetonitrile, the extract was diluted, and cyclotides were captured on a C18-SPE column. The eluate was freeze-dried, re-dissolved to a concentration proportional to the original amount of plant

FIGURE 1 | Schematic structure of a Möbius and bracelet cyclotide, together with typical cyclotide sequences from both subfamilies. The structures are based on the PDB files 1NB1 and 2KNM. The abbreviated notation Cy O2 and Cy O19 stand for the cycloviolacins O2 and O19, respectively. Note the unique features of the CCK motif: a cyclic backbone with sequence loops (1–6) and three stabilizing disulfide bonds. These disulfides are arranged in a cystine knot: that is, two of the disulfides form a ring structure together with the backbone connecting the four cysteines (I–IV; II–V), while the third disulfide is threaded through the ring (III–VI).

material from which the extract had been taken, and then analyzed. With this procedure it was possible to get robust LC-MS results using extracts corresponding to as little as 1 mg of plant material. However, for practical reasons, and to allow for repeated analyses, most extractions were performed on approximately 10 mg of plant material, typically corresponding to less than 1 cm^2 of leaf material.

Identification of Cyclotides

Cyclotides were putatively detected by their late retention times and molecular masses between 2.8 and 3.8 kDa. **Figure 3** shows LC-MS base peak chromatograms of six species from five different genera. After visually inspecting all chromatograms, we compiled a database of the detected cyclotides. The data are given in the **Supplementary Table 2** as an Excel sheet, which can be organized according to plant species, retention time (i.e., relative hydrophobicity), molecular weight, and occurrence in more than one plant species. Peaks having a unique molecular weight or retention time (±1 min) were classified as individual cyclotides. Using these criteria, 744 different cyclotides were detected from the 143 species and their retention time relative to mass is plotted in **Figure 4**. The masses varied from 2806 to 3716 Da with an average of 3070 Da and a median of 3048 Da.
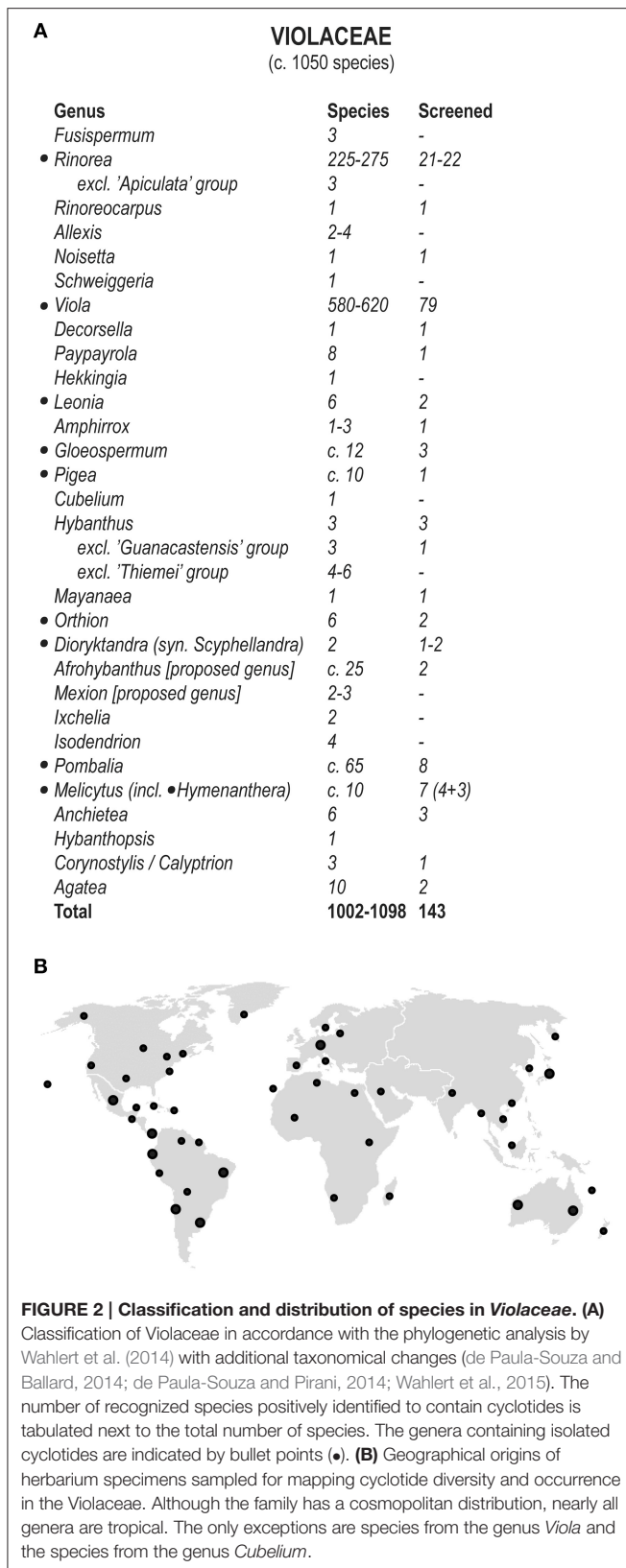
From the analyses it is clear that the number of cyclotides varies substantially among species. In some species only a single cyclotide was detected (e.g., *Orthion oblanceolatum* and *Leonia glycocarpa*), whereas other species contained 25 or more cyclotides (e.g., *Viola cazorlensis* and *Viola allchariensis*). The number of detected cyclotides in different species did not correlate with any specific genus or with the phylogenetic pattern

of the Violaceae. When the distribution of different cyclotides was compared among species, it was apparent that many cyclotides occurred in more than one species. For example, varv A was found in 70% of the analyzed species of the genus *Viola*, whereas varv E and kalata B1 were present in about 50%, and six other cyclotides were in 10–30% of the *Viola* species. The remainders of the cyclotides were only found in either one or a few (<10%) species.

Cyclotides were positively identified in all genera tested. The number of species that was shown to contain cyclotides corresponds to 15% of the total number of species in the Violaceae family. Although some samples gave an initial negative result (less than 10%), in that no ions corresponding to cyclotides were detected using LC-MS, no correlation was seen between the distribution pattern of negative samples and genera or species. Where it was possible to resample species using samples from other sources, cyclotides were usually found. The main reason for this is most likely due to different conservation techniques that might have been used to preserve the plant material: for example, a botanist might have preserved a plant specimen whole, or sprayed it with ethanol during the preparation of the specimen in order to remove living insects, fungi and bacteria. This latter procedure might have extracted hydrophobic proteins, including cyclotides, making it difficult to detect them with our analytical protocols.

An Evolutionary Perspective on the Cyclotide Fingerprints

The cyclotide fingerprints were then analyzed for their possible chemosystematic value, e.g., to either reinforce current

A   **VIOLACEAE**
(c. 1050 species)

Genus	Species	Screened
Fusispermum	3	-
• *Rinorea*	225-275	21-22
excl. 'Apiculata' group	3	-
Rinoreocarpus	1	1
Allexis	2-4	-
Noisetta	1	1
Schweiggeria	1	-
• *Viola*	580-620	79
Decorsella	1	1
Paypayrola	8	1
Hekkingia	1	-
• *Leonia*	6	2
Amphirrox	1-3	1
• *Gloeospermum*	c. 12	3
• *Pigea*	c. 10	1
Cubelium	1	-
Hybanthus	3	3
excl. 'Guanacastensis' group	3	1
excl. 'Thiemei' group	4-6	-
Mayanaea	1	1
• *Orthion*	6	2
• *Dioryktandra (syn. Scyphellandra)*	2	1-2
Afrohybanthus [proposed genus]	c. 25	2
Mexion [proposed genus]	2-3	-
Ixchelia	2	-
Isodendrion	4	-
• *Pombalia*	c. 65	8
• *Melicytus (incl. •Hymenanthera)*	c. 10	7 (4+3)
Anchietea	6	3
Hybanthopsis	1	
Corynostylis / Calyptrion	3	1
Agatea	10	2
Total	**1002-1098**	**143**

B

FIGURE 2 | Classification and distribution of species in *Violaceae*. (A) Classification of Violaceae in accordance with the phylogenetic analysis by Wahlert et al. (2014) with additional taxonomical changes (de Paula-Souza and Ballard, 2014; de Paula-Souza and Pirani, 2014; Wahlert et al., 2015). The number of recognized species positively identified to contain cyclotides is tabulated next to the total number of species. The genera containing isolated cyclotides are indicated by bullet points (•). **(B)** Geographical origins of herbarium specimens sampled for mapping cyclotide diversity and occurrence in the Violaceae. Although the family has a cosmopolitan distribution, nearly all genera are tropical. The only exceptions are species from the genus *Viola* and the species from the genus *Cubelium*.
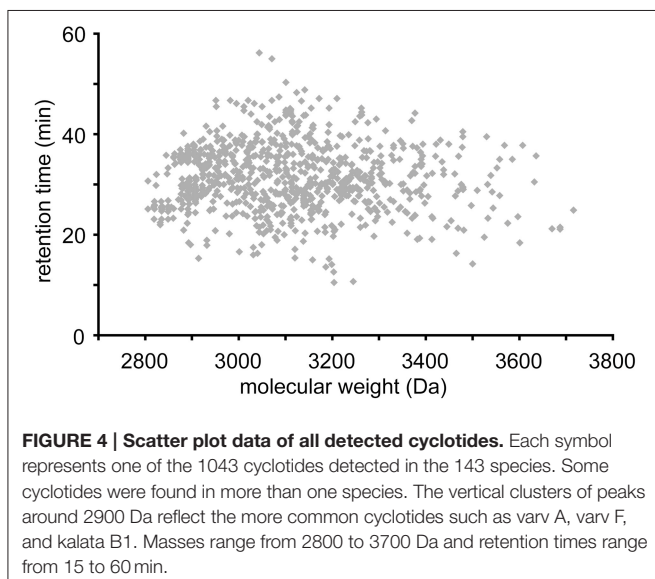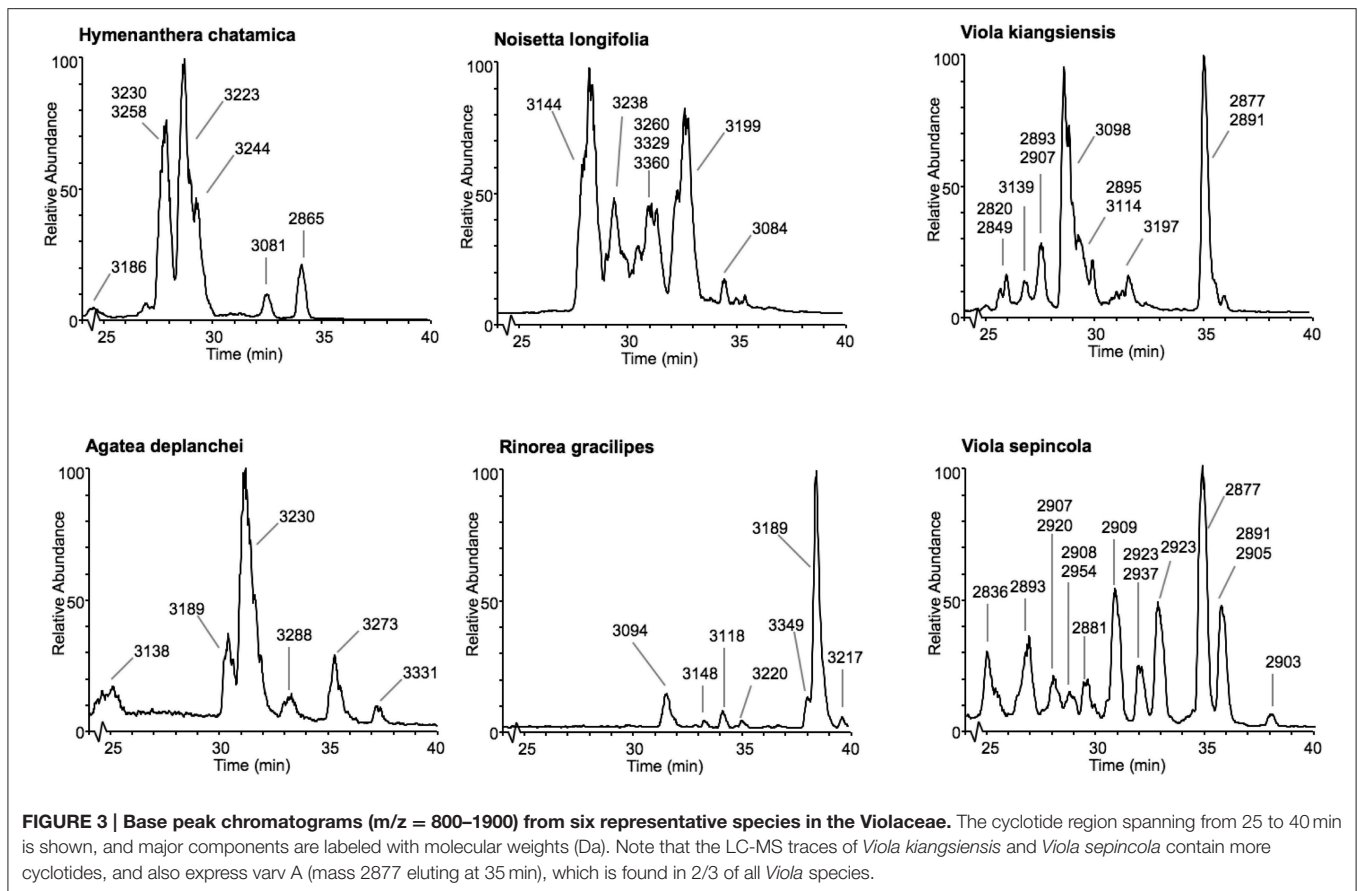
knowledge on the evolution of the Violaceae or predict the evolution of cyclotide proteins *per se*. The LC-MS profiles were used for clustering analysis by UPGMA and neighbor-joining [as implemented in PAUP*; (Swofford, 2003)], but the lack of the same cyclotide in two or more taxa did not provide enough resolution to draw conclusions about evolution of diversity of cyclotides. An example cluster analysis is shown in the **Supplementary Figure 1**. To increase the data points we added complete translated prepropeptides or the corresponding mRNAs available in NCBI nucleotide (GenBank) and protein databases. This approach indicates an origin of cyclotide diversity within each taxon, but also failed to produce results in agreement with botanical classification. Thus, more information of the genetic structure of cyclotides is needed to investigate the evolution of cyclotides.

Interestingly, approximately 1/3 of the cyclotides were detected in more than one species; thus, on average approximately 2/3 of the cyclotides present in a single species are unique to that species. Hence, by exploring any uninvestigated species in the Violaceae, it should be possible to discover an entirely new set of cyclotides with novel sequences. In the present study we detected on average 5 or 6 unique sequences per species. By extrapolation, there are potentially at least 5000 different cyclotides in the Violaceae alone, which is consistent with earlier estimates (Simonsen et al., 2005). Nevertheless, this is still likely to be an underestimate of the total number of cyclotides, since we are probably only able to detect the most abundant cyclotides in this screen. Based on the experience of species investigated in detail from large scale extractions, such as *Viola odorata* (Ireland et al., 2006), the real value is probably up to five times higher, i.e., approximately 25,000.

Sequencing of Novel Cyclotides

Abundant protein peaks from a diverse set of genera were purified by HPLC and reduced by dithiothreitol (DTE), followed by S-carbamidomethylation using iodoacetamide. Alkylation is an effective way to confirm the number of cysteines in a protein. Because cyclotides contain six cystines, i.e., three disulfide bonds, the reactions increase the mass by 348 Da (58×6 Da). The reactions were performed on almost 100 collected chromatographic peaks, all of which gave an increase of 348 Da, reinforcing their identity as cyclotides.

In order to be able to sequence small proteins efficiently with tandem MS, the proteins have to be in a linear conformation. Using microwave assisted enzymatic cleavage (together with endoproteinase GluC, which cleaves after the conserved Glu residue within cyclotides), we were able to get linear products amenable for sequencing within a matter of minutes. For full sequence coverage, the enzymatic digest was repeated with trypsin. Samples were analyzed using nanospray MS-MS using a Q-Tof MS. The procedure was performed both on known cyclotides to confirm the identity of protein peaks in the LC-MS analysis, as well as on novel cyclotides in order to identify new sequences.

FIGURE 3 | Base peak chromatograms (m/z = 800–1900) from six representative species in the Violaceae. The cyclotide region spanning from 25 to 40 min is shown, and major components are labeled with molecular weights (Da). Note that the LC-MS traces of *Viola kiangsiensis* and *Viola sepincola* contain more cyclotides, and also express varv A (mass 2877 eluting at 35 min), which is found in 2/3 of all *Viola* species.

FIGURE 4 | Scatter plot data of all detected cyclotides. Each symbol represents one of the 1043 cyclotides detected in the 143 species. Some cyclotides were found in more than one species. The vertical clusters of peaks around 2900 Da reflect the more common cyclotides such as varv A, varv F, and kalata B1. Masses range from 2800 to 3700 Da and retention times range from 15 to 60 min.

In total, 12 novel cyclotides were sequenced (**Table 1**) from nine different species in six genera. The total number of known cyclotides is now becoming so large that a uniform naming system is necessary. The problem is likely to increase exponentially as the number of new sequences increases in the near future, not the least with the current development of total RNA sequencing. In the present study we follow the proposed naming system in which a trivial name for a cyclotide is constructed as an indicative and pronounceable acronym of the Latin binomial of the plant from which it was first isolated, followed by a letter indicating its order of discovery (Broussalis et al., 2001; Arnison et al., 2013).

The novel cyclotides found in the present study resemble known sequences, but some differences are worth mentioning. For example, hyde A has an atypical sequence not characteristic of any of the two major subfamilies or hybrid variants thereof. The hybrids that have been found so far, e.g., kalata B8 (Daly et al., 2006), have a typical bracelet sequence but contain a shorter loop 3, which gives rise to an abnormal helix. Hyde A has, like kalata B8, a short loop 3 but instead of being similar to either bracelet or Möbius cyclotides in loop 5, it has a unique six amino acid sequence: GSTVPL. Most cyclotides have four amino acids in that loop, or exceptionally five amino acids, and the bracelet types often comprise 1–3 charged Lys or Arg residues, while Möbius forms have a aromatic amino acid in conjunction with a *cis*-Pro to create a conformational twist in the backbone. Hyde A also possesses a positively charged residue in loop 2, which is unusual and is seen only in a few other cyclotides.

TABLE 1 | Alignment of the isolated cyclotides of nine species in six different genera in the Violaceae, but the species are given as reported on herbarium sheets.

Species	Cyclotide	Sequence*	Mol weight(mono isot.)
		Loop 2 3 4 5 6 1	
Gloeospermum pauciflorum	Glopa F	SCVFL–PC–LSVSLGCSC**K**N**K**V––CY**R**NG**R**LPCG**E**	3310.5
Gloeospermum pauciflorum	Glopa G	SCVFL–PC–LSAVLGCSC**K**N**K**V––CY**R**NG**R**LPCG**E**	3294.5
*Hybanthus denticulatus***	Hyde A	SCVFD**R**TC–HLA––GCGCGSTVPLC**V**RNGVLPCG**E**	3228.4
*Hymenanthera oborata***	Hobo A	TCTLG–TC–NTP––GCTCSW––PLCT**K**NGLPTCG**E**	2936.2
Melicytus macrophyllus	Mema A	SCVWL–PCTVTALLGCSC**K**D**K**V––CY**R**NGL–PCA**AE**	3307.4
Melicytus macrophyllus	Mema B	SCVWL–PC–LTGLVGCSC**K**NNV––CYTNGTVPCG**E**	3395.3
Orthion oblanceolatum	Orto A	SCVYL–PCLLTAPLGCSC**K**N**K**V––CY**R**NGL–PCG**E**	3280.5
Rinorea gracilipes	Rigra A	SCVWL–PCTVTALLGC**K**CE**TR**G––CTLNGV–PCG**E**	3188.4
Rinorea lindeniana	Rili A	SCVWL–PCTVTALLGCTCVD**R**V––CFLDGL–PCA**E**	3262.4
Rinorea lindeniana	Rili B	TCAGG–TC–NTP––GCSCTW––PLCT**R**NGLPVCG**E**	2876.1
Viola decumbens	Vide A	SCVFL–PC–LTSALGCSC**K**S**K**V––CY**R**NGL–PCG**E**	3113.4
Viola nivalis	Vini A	SCVWL–PC–LSGLAGCSC**K**N**K**V––CYYDGSVPCG**E**	3216.3

Amino acids in bold indicate where trypsin or endoproteinase GluC cleaved the proteins prior to tandem MS sequencing.

*As MS sequencing does not discriminate between leucine and isoleucine both are represented as L.

**The species Hybanthus denticulatus *should otherwise be given as an as yet unnamed genus segregated from* Hybanthus, *and* Hymenanthera obovata *should be given as* Melicytus obovatus.

Glycosylation—A Further Increase in the Diversity of Cyclotides

During the MS analysis it was apparent that a total of 115 cyclotides from 53 different species were observed in their expected form but also as variants carrying post translational modifications corresponding to mass additions of 162 Da or multiples of 162, consistent with one or more hexose sugars (**Figure 5**; cyclotides are marked in **Supplementary Table 2**. One of these glycosylated cyclotides was related to the well-characterized cycloviolacin O2 (cyO2) and was isolated in sufficient quantities to allow for detailed characterization. Commercial monosaccharide analysis of hydrolyzed glycosylated cyO2 [cyO2(gly)] revealed hexose to be primarily glucose, but with traces of arabinose, fucose, and galactose also being detected (data shown in **Supplementary Figure 2**). Extensive analysis by tandem MS spectrometry showed that the hexose moiety in glycosylated variants of cyO2, mema A, and mram 1, was associated with loop 5, i.e., KSKV or KDKV. However, definitive evidence for which residue it was attached to be difficult to obtain solely based on MS data, due to poor fragmentation in this region. Instead we used both chemical and NMR spectroscopy strategies to delineate the glycosylation site.

To investigate whether the free amine groups of Lys23 and Lys25 remained unmodified, cyO2(gly) was acetylated using acetic anhydride as previously described (Herrmann et al., 2006). After acetylation the mass of cyO2 increased with 84 Da, corresponding to the addition of two acetyl groups, suggesting that hexose was not bound to either of the lysines.
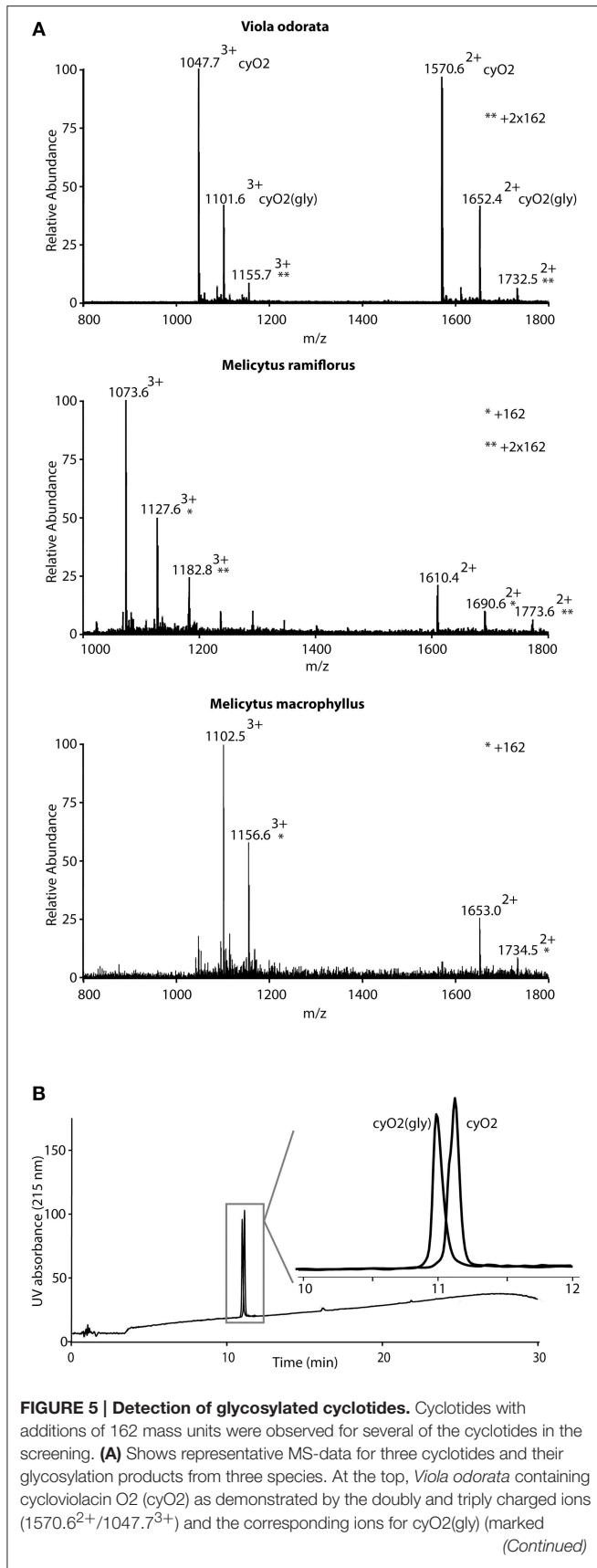
To investigate whether the addition of the hexose moiety resulted in structural changes, we structurally characterized cyO2(gly) using solution NMR. Homonuclear 2D NMR spectra were recorded, analyzed, and compared to spectra for cyO2 (Göransson et al., 2009). The NMR spectra for cyO2(gly) were of good quality in terms of signal dispersion, as generally observed for cyclotides. However, cyO2(gly) showed broader lines than native cyO2 and the spectra contained evidence of minor conformations or other inhomogeneities being present. The peaks corresponding to the main isomer were assigned using sequential assignment methods. From the secondary shifts, which are good indicators of secondary structure, it is clear that the cyclotide fold is retained after glycosylation. In particular, the Hα chemical shifts were very similar between the glycosylated and non-glycosylated forms, with the exception of the site around the modification. This was further supported by trends in the shifts of the amide protons as well as the side chain protons. Interestingly, the side chain of Lys25, which is rather flexible in solution in native cyO2, becomes more ordered in cyO2(gly), as evident from separation of the chemical shifts of the methylene pairs HD2/HD3 and HE2/HE3. However, the sugar moiety itself does not appear to be structured as no well resolved sharp peaks that could be assigned to specific protons of the sugar, or connections between the sugar and peptide, could be identified.

The combined analyses suggest that Ser24 is the site of glycosylation in loop 5 (KSKV) in cyO2(gly). Mema A do not possess an amino acid with a hydroxyl group in loop 5 (KDKV), but it may be that the hexose moiety in this case are attached to Asp25. Glycosylation, in particular O-glycosylation of Ser residues (Gomord et al., 2010), is relatively common in plants but has never been described for cyclotides. The functional role of cyO2(gly) was tested by comparing its cytotoxic activity to native cyO2 against U937-GTB lymphoma cells. Interestingly, the observed activity was identical to that for unmodified cyO2, with both having an IC_{50} of approximately $0.9\,\mu M$, and in terms of cytotoxic activity glycosylation is functionally irrelevant. Cycloviolacin O2 had an IC_{50} of $0.87 \pm 0.14\,\mu M$, and the glycosylated variant $0.91 \pm 0.08\,\mu M$. The dose response curve can be found in **Supplementary Figure 3**.

Stability of Cyclotides

Access to herbaria collections made it possible to examine cyclotide stability over time periods not attainable by other means. It was also necessary to verify that cyclotides could be
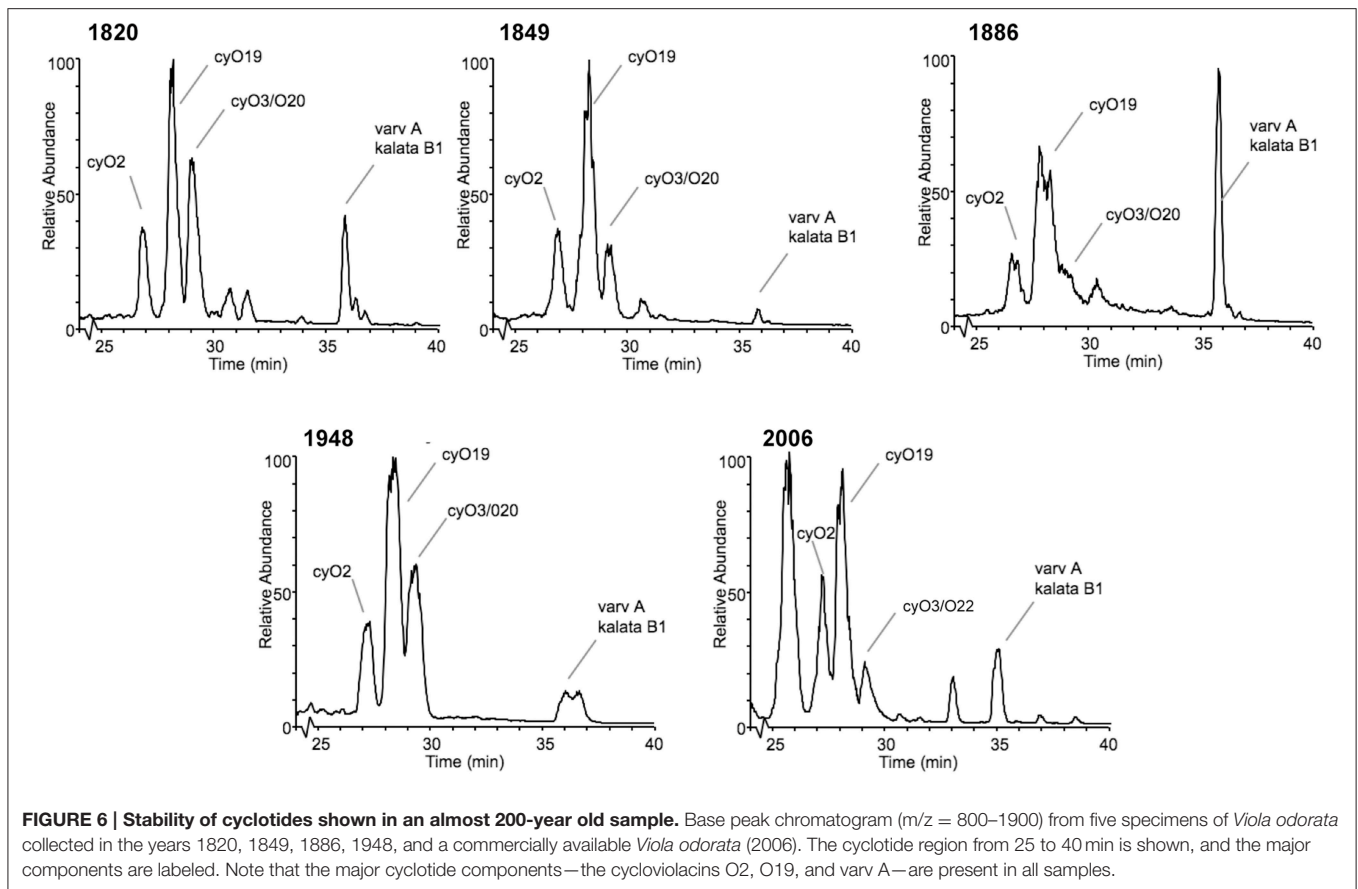
FIGURE 5 | Detection of glycosylated cyclotides. Cyclotides with additions of 162 mass units were observed for several of the cyclotides in the screening. **(A)** Shows representative MS-data for three cyclotides and their glycosylation products from three species. At the top, *Viola odorata* containing cycloviolacin O2 (cyO2) as demonstrated by the doubly and triply charged ions (1570.6^{2+}/1047.7^{3+}) and the corresponding ions for cyO2(gly) (marked

(Continued)

FIGURE 5 | Continued

with *). Cyclotides mram 1 from *Melicytus ramiflorus* (1610.4^{2+}/1073.6^{3+}) and mema A from *Melicytus macrophyllus* (1653.0^{2+}/1102.5^{3+}) and their respective glycosylated derivates show the same patterns of ions. Note that the cyO2 and mram 1 show signs of double glycosylations (**). The cyclotide derivates with additional 162 Da (corresponding to one sugar residue) shown in the figure have also been sequenced by MS-MS, demonstrating that the glycosylation is localized in loop 5. Cyclotides and their glycosylated derivates elute closely (or co-elute) on RP-HPLC, but may be separated. **(B)** Shows the separation of cyO2 and cyO2(gly), which unambiguously define them as two separate molecular species. The separation was done using a Phenomenex Kinetex C18 column (150 × 4.6 mm, 2.6 μm) operated at a flow rate of 1 ml/min and a 2%/min gradient of acetonitrile in water, containing 0.1% formic acid.

extracted from materials held in old plant collections. Using sweet violet, *Viola odorata*, which has a high level of cyclotide expression (Craik et al., 1999; Svangård et al., 2003; Ireland et al., 2006) and is represented in a number of different herbarium collections, we were able to sample specimens collected between 1820 and the present day. The LC-MS chromatograms of these samples and the standard reference cyclotides isolated and sequenced from recently collected and commercially available sweet violet are compared in **Figure 6**. The three dominant cyclotides (cycloviolacin O2, O19 and varv A) in all of the sweet violet LC-MS chromatograms had the same intensity of the peaks (equal order of magnitude) in both old and recent materials, indicating their concentration to be similar in all samples. The small differences seen are more similar to the seasonal variations reported for other cyclotides (Trabi et al., 2004). No signs of previously reported degradation products, e.g., oxidized tryptophan residues (Plan et al., 2007; Burman et al., 2011), were visible in any of the spectra, confirming the remarkable stability of cyclotides.

DISCUSSION

In this study we performed the first large-scale chemical screen for cyclotides in Violaceae. A key aspect of the study was the access to material from herbarium collections, which made it possible to include species from all continents and to cover a large part of the biological diversity of the Violaceae family. First, we adapted a small-scale extraction procedure applicable to the limited amount of dried plant material that we were permitted to sample from herbaria sheets. We then used the sensitivity and specificity of LC-MS to analyse the cyclotide content of the extracts. Using this strategy, 744 cyclotides were detected and compiled into a database for further analysis. Cyclotides were positively identified in 17 of the 23–31 genera, and in 15% of all the species in the Violaceae. The samples represented a wide spread of species across the Violaceae. More samples were taken from the larger genera, *Hybanthus* sensu lato, *Rinorea* sensu lato, and *Viola*, which together comprise approximately 90% of the total number of species of the family. The sampling also covered the majority of the smaller genera. The screen generated high-resolution data by mapping the cyclotide content found in species from different genera in the Violaceae, as well as in the subgroups of each genus.

FIGURE 6 | Stability of cyclotides shown in an almost 200-year old sample. Base peak chromatogram (m/z = 800–1900) from five specimens of *Viola odorata* collected in the years 1820, 1849, 1886, 1948, and a commercially available *Viola odorata* (2006). The cyclotide region from 25 to 40 min is shown, and the major components are labeled. Note that the major cyclotide components—the cycloviolacins O2, O19, and varv A—are present in all samples.

Two approaches were used to confirm that MS peaks were correctly identified as cyclotides. First we confirmed by reduction and alkylation of cystines that the major peaks contained the three disulfide bonds characteristic of cyclotides. This is not full evidence that the peak corresponds to a cyclotide, but it strongly suggests so. Some peaks also correspond to already known cyclotides. Subsequently, cyclotides that were isolated in sufficient amounts were subjected to enzymatic cleavage and MS-MS sequencing. We managed to sequence 12 novel cyclotides from six of the genera (*Gloeospermum,* a *Hybanthus* segregate, *Melicytus*, *Orthion*, *Rinorea*, and *Viola*). In total, cyclotides have now been fully sequenced from seven genera in the family Violaceae [cyclotides have previously been isolated from the genus *Leonia* (Hallock et al., 2000)]. These seven genera are phylogenetically well spread out across the family (**Figure 2A**). From these results, we can conclude that Violaceae is an extremely rich source of cyclotides, and that cyclotides are ubiquitous throughout the family.

Although the exceptional chemical, thermal, and biological stability of cyclotides have been demonstrated in earlier studies, our approach of sampling extensively from plant material held in various herbaria gave us a unique opportunity to assess cyclotide stability over time. To this end, we sampled specimens of *Viola odorata* (sweet violet) that were collected as long ago as 1820. The base peak chromatograms from those samples were essentially identical to those from recently collected specimens,

and no signs of any degradation products were visible. This result confirms the stability of cyclotides. Their compact, cyclic and knotted structure is probably the main factor accounting for this exceptional stability, as these characteristics limit opportunities for proteolysis, or for thermal or chemical degradation to occur (Colgrave and Craik, 2004).

Proteins are usually considered to be rather fragile biomolecules that are easily degraded by chemical and biotic factors. In the case of cyclotides, we show that their 3D-structures are retained intact in dry leaves for almost 200 years, and we have no doubt that they would be found intact in even older material if available. Indeed, our results establish cyclotides as one of the most stable protein families in nature. It is clear that this extended "shelf-life" of cyclotides is unparalleled in other protein families, making them ideal for applications in protein drug design.

When analysing the LC-MS data, we detected between one and 25 (average 7.3) different cyclotides per species. Detailed studies of a single species, e.g., *Viola odorata* that has previously been used as a model plant from the Violaceae, have shown it to express 30 cyclotides (Ireland et al., 2006). In comparison, the resulting numbers of the current study are relatively low; however, it should be borne in mind that we used only a very limited amount of plant material and it is likely that low abundant proteins may escape detection. The number of cyclotides from a single species in Rubiaceae is similar: *Oldenlandia affinis* has

been examined exhaustively for proteins and RNA and found to contain at least 18 unique sequences (Plan et al., 2007).
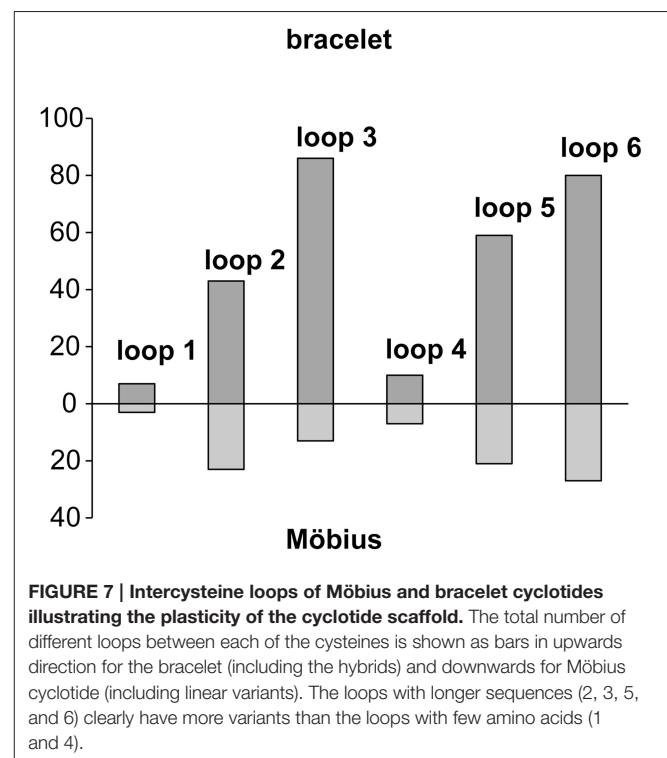
Previous studies have identified deamidation and oxidation products as cyclotide derivatives that further increase the maximum number of peaks that can be detected during LC-MS (Plan et al., 2007; Burman et al., 2011). Here, we describe for the first time the presence of glycosylation of cyclotides. A glycosylated variant of cyO2 carrying a hexose moiety was purified and characterized. The hexose was located to loop 5 and chemical and structural analyses indicate it to be o-linked to Ser24. NMR studies revealed that while the overall characteristic fold of the cyO2 was retained, the glycosylation had impact on the local structure of loop 5. Although this was not found to have any significant impact on the cytotoxic activity, glycosylation altering both chemical and structural properties represents a further major increase in the diversity of naturally occurring cyclotides and is possibly functionally relevant in other biological assays. For example, outside the plant kingdom glycosylation of host defense peptides has been shown to be important for antimicrobial activity (Bulet et al., 1999).

The current study demonstrates that each species is characterized by its own set of cyclotides, many of which are unique to that species. The systematic value of the data proved to be limited. The cluster analysis is based on the qualitative profiles of cyclotides in Violaceae, but including sequence data did not improve the analysis (data not shown). Comparison of cyclotide sequences is complicated by the presence of identical amino acid sequences in distantly related species (e.g., kalata B1 in *Viola odorata* and *Oldenlandia affinis*), as well as by the presence of multiple cyclotides sequentially encoded by the same mRNA (e.g., kalata S–kalata B1–kalata S in *Viola odorata*). This leads to problems in deducing homology among the peptides, and thus phylogenetic analysis based on amino acid sequences can be distorted. Other means of increasing the number of data points are analysing the complete translated prepropeptides or the corresponding mRNAs, which we tested on sequences available in NCBI nucleotide (GenBank) and protein databases. These approaches indicate an origin of cyclotide diversity within each taxon, but also failed to produce results in agreement with botanical classification. Thus, more information of the genetic structure of cyclotides is needed to investigate the evolution of this group of proteins. Although not supported by the cluster analysis, the genus *Viola* seems to hold a common set of cyclotides that are expressed in abundance and found in the majority of the species.

It has been suggested that cyclotides are expressed when a plant is stressed by external factors, such as grazing by insects or other herbivores. For instance, the cyclotide content in plants grown under glass is lower, than in plants grown in the field (Mylne et al., 2010), and another study has demonstrated the up-regulation of cyclotide-like genes after smut infection on maize plants (Basse, 2005). Thus, the fact that each species expresses its own characteristic cocktail of cyclotides may result from its combined response to a number of different stress factors during its growth. Alternatively, the fact that cyclotides are a fast evolving group of proteins,

may have enabled each species to produce its unique set of cyclotides.

To date, more than 250 cyclotide sequences have been described (see Cybase, http://www.cybase.org.au; Mulvenna et al., 2006; Wang et al., 2008b), 2/3 of which belong to the bracelet subfamily, the rest to the Möbius subfamily. In the current study we estimate the number of cyclotides in the Violaceae to be in the range of 5000–25,000, while in the Rubiaceae the number has been previously estimated to be between 10,000 and 50,000 (Gruber et al., 2008). Adding these values together gives a total of between 15,000 and 75,000, which clearly illustrates the plasticity of the cyclotide scaffold and the important role that the cyclotides must have in plant defense. It also makes the cyclotides one of the largest groups of proteins known, and taking into account the fact that they are also found in other plant families, the total number of naturally occurring cyclotides may be even higher. **Figure 7** shows the total number of loops between each of the cysteines of the two main subfamilies. By simply multiplying the number of variants of each loop, a theoretical library of 600 million different cyclotides is obtained. As all of these proteins will be composed of naturally occurring intercysteine sequences, it is likely that they, if expressed or synthesized, can be folded into native like cyclotide structures. Above all, this example demonstrates the plasticity of the structure, which in combination with its ability to harness diverse bioactivities (Leta Aboye et al., 2008; Chan et al., 2011; Aboye et al., 2012; Göransson et al., 2012; Wong et al., 2012), may be used by the plant and by us for applications in biotechnology as a privileged protein scaffold.

FIGURE 7 | Intercysteine loops of Möbius and bracelet cyclotides illustrating the plasticity of the cyclotide scaffold. The total number of different loops between each of the cysteines is shown as bars in upwards direction for the bracelet (including the hybrids) and downwards for Möbius cyclotide (including linear variants). The loops with longer sequences (2, 3, 5, and 6) clearly have more variants than the loops with few amino acids (1 and 4).

In summary, we have mapped the occurrence and distribution of cyclotides in the Violaceae by combining sampling of materials from archived herbarium collections and modern analytical techniques. Cyclotides demonstrate remarkable stability with no detectable degradation products in samples 200 years old. Cyclotides are expressed in plants as cocktails of up to 25 different cyclotides per species. Most cyclotides found in a single species seem to be unique to that species, although some are more generally expressed among different species. In particular, the Möbius cyclotides, varv A, varv E, and kalata B1 are found in the majority of the samples in the genus *Viola*. The total number of different cyclotides in the Violaceae is at least 5000, but may be as high as 25,000. Our mapping of cyclotide containing species covered approximately 1/6 of all species in the Violaceae, and included representatives from most genera, such that we can now definitely conclude that cyclotides are ubiquitous in the Violaceae.

METHODS

Plant Material

Samples of 5–50 mg of each plant species, comprising mostly leaf material but also fruit, were obtained from the herbaria held at the Swedish Museum of Natural History, Stockholm (S), and the Universities of Gothenburg (GB), and Uppsala (UPS). A few other samples were sourced from personal collections that had voucher specimens deposited at the Uppsala University Herbarium. We included two commercially available samples, *Viola odorata* and *Viola tricolor* (Alfred Galke GmbH, Gittelde/Harz, Germany). A complete list of the analyzed samples is provided in the Supplementary Material.

Extraction

The plant material was frozen in liquid nitrogen, ground to a powder, and extracted three times with 2.0 mL of 60% acetonitrile in 0.1% formic acid (eluent B). The extract was diluted with acidic water to 10% acetonitrile in 0.1% formic acid (eluent A), captured on a 500 mg C18 column, washed with eluent A, and eluted with eluent B. The eluate was freeze-dried and dissolved in eluent A to a concentration proportional to the original amount of extracted material (10 μL/mg).

Liquid Chromatography—Mass Spectrometry (LC-MS)

The extracts were analyzed by LC-MS using a Shimadzu LC10 HPLC (Shimadzu, Kyoto, Japan) connected to an LCQ electrospray ion trap MS (Thermo Finnigan, San Jose, CA), with a Grace Vydac Everest Narrowbore column (100 × 2.1 mm i.d., C18, 5 μm, 300 Å) and a linear gradient from 10 to 60% acetonitrile in 0.05% formic acid at a flow rate of 0.3 mL/min over 60 min. The capillary temperature was set at 220°C and the spray voltage at 4 kV.

Preparative High Performance Liquid Chromatography (HPLC)

Preparative RP-HPLC, using a Shimadzu LC10 system equipped with an SPD-M10Avp photodiode array detector, or an Äkta

system (Amersham Biosciences, Uppsala, Sweden) collecting data at 215, 254, and 280 nm, was carried out using a Grace Vydac Everest (250 × 4.6 mm i.d., C18, 5 μm, 300 Å) or a Phenomenex column (250 × 4.6 mm i.d., C18, 5 μm, 300 Å), run with linear gradients from 10 to 60% acetonitrile, with 0.1% trifluoroacetic acid, in water. Fractions were analyzed using MS.

Reduction, Alkylation, and Enzymatic Digestion

Isolated proteins were reduced by dithiothreitol (DTE) in 0.25 M Tris-HCl containing 1.0 mM EDTA and 6 M guanidine-HCl (pH 8.5) in the dark under nitrogen for 3 h. For S-carbamidomethylation, 50 mg iodacetamide were dissolved in 0.5 M Tris-HCl, and 2 mM EDTA was added to the reduced samples. After 10 min the reactions were quenched by adding 250 μl 0.5 M citric acid. The S-carbamidomethylated proteins were purified directly and desalted directly by HPLC. Digestion was done using trypsin and/or endoproteinase GluC dissolved in 25 mM NH$_4$HCO$_3$ using a Discover® microwave system (CEM, Matthews, NC) at 55°C, ΔT 5°C with the power 45 W, and a run time of 15 min.

MS Sequencing

The cleaved samples were dissolved in 50% methanol with 1% formic acid and analyzed with MS-MS. The samples were injected by syringe through a PicoTip® emittor at 0.3 μL/min connected to a Q-Tof Micro™ (Waters, Milford, MA) with the voltage set at 1.4 kV. The fragmentation spectra were analyzed with the help of MassLynx/BioLynx software (MaxEnt3 and Protein Sequencing).

Monosaccharide Analysis

Glycosylated cycloviolacin O2 was sent to Glycosolutions (Worchester, MA) for monosaccharide analysis (HPAEC Neutral Monosaccharide Analysis). The monosaccharide position in three cyclotides (cycloviolacin O2, mema A, and mram 1) was analyzed using MS-MS.

NMR Spectroscopy

For NMR spectroscopy studies of cyO2(gly), 2 mg of peptide was dissolved in 0.5 ml of 90% H$_2$O/10% D$_2$O, pH approximately 4 and 2D homonuclear datasets, including DQF-COSY, TOCSY, and NOESY were recorded at 298 K and 600 MHz on a Bruker Avance II spectrometer. All data were recorded and processed using Topspin 3.0 (Bruker) and analyzed using CARA (Keller, 2004). Chemical shifts were compared to random coil chemical shifts described by Wishart et al. (1995).

Cytotoxicity Assay

The cytotoxicity of individual cyclotides and mixtures was determined using the fluorometric microculture cytotoxicity assay (FMCA) (Larsson and Nygren, 1989; Lindhagen et al., 2008) with the human lymphoma cell line U-937 GTB. The cytotoxic properties of dilution series of cycloviolacin O2 with and without the monosaccharide were tested in duplicates at three occasions

as described before (Lindholm et al., 2002; Svangård et al., 2004; Herrmann et al., 2008).

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2015.00855

Supplementary Table 1 | Species that contain cyclotides, together with collection data.

Supplementary Table 2 | Database of detected cyclotides.

Supplementary Figure 1 | Tree from UPGMA clustering of presence-absence of cyclotides from the LC-MS profiles.

Supplementary Figure 2 | Sugar analyses using HPLC and electrochemical detection demonstrating that glucose is released after chemical hydrolysis.

Supplementary Figure 3 | Dose response curve of cyO2 and glycosylated cyO2 in the cytotoxicity assay.

REFERENCES

Aboye, T. L., Ha, H., Majumder, S., Christ, F., Debyser, Z., Shekhtman, A., et al. (2012). Design of a novel cyclotide-based CXCR4 antagonist with anti-human immunodeficiency virus (HIV)-1 activity. *J. Med. Chem.* 55, 10729–10734. doi: 10.1021/jm301468k

Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., et al. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* 30, 108–160. doi: 10.1039/C2NP20085F

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. doi: 10.1073/pnas.0710338104

Basse, C. W. (2005). Dissecting defense-related and developmental transcriptional responses of maize during *Ustilago maydis* infection and subsequent tumor formation. *Plant Physiol.* 138, 1774–1784. doi: 10.1104/pp.105.061200

Boothe, J., Nykiforuk, C., Shen, Y., Zaplachinski, S., Szarka, S., Kuhlman, P., et al. (2010). Seed-based expression systems for plant molecular farming. *Plant Biotechnol. J.* 8, 588–606. doi: 10.1111/j.1467-7652.2010.00511.x

Broussalis, A. M., Göransson, U., Coussio, J. D., Ferraro, G., Martino, V., and Claeson, P. (2001). First cyclotide from Hybanthus (Violaceae). *Phytochemistry* 58, 47–51. doi: 10.1016/S0031-9422(01)00173-X

Bulet, P., Hetru, C., Dimarcq, J. L., and Hoffmann, D. (1999). Antimicrobial peptides in insects; structure and function. *Dev. Comp. Immunol.* 23, 329–344. doi: 10.1016/S0145-305X(99)00015-4

Burman, R., Gruber, C. W., Rizzardi, K., Herrmann, A., Craik, D. J., Gupta, M. P., et al. (2010). Cyclotide proteins and precursors from the genus Gloeospermum: filling a blank spot in the cyclotide map of Violaceae. *Phytochemistry* 71, 13–20. doi: 10.1016/j.phytochem.2009.09.023

Burman, R., Herrmann, A., Tran, R., Kivelä, J. E., Lomize, A., Gullbo, J., et al. (2011). Cytotoxic potency of small macrocyclic knot proteins: structure-activity and mechanistic studies of native and chemically modified cyclotides. *Org. Biomol. Chem.* 9, 4306–4314. doi: 10.1039/c0ob00966k

Chan, L. Y., Gunasekera, S., Henriques, S. T., Worth, N. F., Le, S. J., Clark, R. J., et al. (2011). Engineering pro-angiogenic peptides using stable, disulfide-rich cyclic scaffolds. *Blood* 118, 6709–6717. doi: 10.1182/blood-2011-06-359141

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. doi: 10.1021/bi049711q

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. doi: 10.1002/cbic.200800174

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. doi: 10.1006/jmbi.1999.3383

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. doi: 10.1007/s00018-009-0159-3

Daly, N. L., and Craik, D. J. (2000). Acyclic permutants of naturally occurring cyclic proteins. Characterization of cystine knot and beta-sheet formation in the macrocyclic polypeptide kalata B1. *J. Biol. Chem.* 275, 19068–19075. doi: 10.1074/jbc.M000450200

Daly, N. L., Clark, R. J., Plan, M. R., and Craik, D. J. (2006). Kalata B8, a novel antiviral circular protein, exhibits conformational flexibility in the cystine knot motif. *Biochem. J.* 393, 619–626. doi: 10.1042/BJ20051371

de Paula-Souza, J., and Ballard, H. E. (2014). Re-establishment of the name Pombalia, and new combinations from the polyphyletic Hybanthus (Violaceae). *Phytotaxa* 183, 1–15. doi: 10.11646/phytotaxa.183.1.1

de Paula-Souza, J., and Pirani, J. R. (2014). Reestablishment of Calyptrion (Violaceae). *Taxon* 63, 1335–1339. doi: 10.12705/636.7

Felizmenio-Quimio, M. E., Daly, N. L., and Craik, D. J. (2001). Circular proteins in plants: solution structure of a novel macrocyclic trypsin inhibitor from *Momordica cochinchinensis*. *J. Biol. Chem.* 276, 22875–22882. doi: 10.1074/jbc.M101666200

Gomord, V., Fitchette, A. C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., et al. (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol. J.* 8, 564–587. doi: 10.1111/j.1467-7652.2009.00497.x

Göransson, U., Burman, R., Gunasekera, S., Strömstedt, A. A., and Rosengren, K. J. (2012). Circular proteins from plants and fungi. *J. Biol. Chem.* 287, 27001–27006. doi: 10.1074/jbc.R111.300129

Göransson, U., Herrmann, A., Burman, R., Haugaard-Jönsson, L. M., and Rosengren, K. J. (2009). The conserved glu in the cyclotide cycloviolacin O2 has a key structural role. *Chembiochem* 10, 2354–2360. doi: 10.1002/cbic.200900342

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. doi: 10.1021/np0499719

Gruber, C. W., Elliott, A. G., Ireland, D. C., Delprete, P. G., Dessein, S., Göransson, U., et al. (2008). Distribution and evolution of circular miniproteins in flowering plants. *Plant Cell* 20, 2471–2483. doi: 10.1105/tpc.108.062331

Gustafson, K. R., Mckee, T. C., and Bokesch, H. R. (2004). Anti-HIV cyclotides. *Curr. Protein Pept. Sci.* 5, 331–340. doi: 10.2174/1389203043379468

Hallock, Y. F., Sowder, R. C. II., Pannell, L. K., Hughes, C. B., Johnson, D. G., Gulakowski, R., et al. (2000). Cycloviolins A-D, anti-HIV macrocyclic peptides from *Leonia cymosa*. *J. Org. Chem.* 65, 124–128. doi: 10.1021/jo990952r

Hashempour, H., Koehbach, J., Daly, N. L., Ghassempour, A., and Gruber, C. W. (2013). Characterizing circular peptides in mixtures: sequence fragment assembly of cyclotides from a violet plant by MALDI-TOF/TOF mass spectrometry. *Amino Acids* 44, 581–595. doi: 10.1007/s00726-012-1376-x

Heitz, A., Hernandez, J. F., Gagnon, J., Hong, T. T., Pham, T. T., Nguyen, T. M., et al. (2001). Solution structure of the squash trypsin inhibitor MCoTI-II. A new family for cyclic knottins. *Biochemistry* 40, 7973–7983. doi: 10.1021/bi0106639

Hekking, W. H. A. (1984). Studies on neotropical Violaceae—the genus Fusispermum. *Proc. Kon. Ned. Akad. Wetensch. C Biol. Med. Sci.* 87, 121–130.

Hekking, W. H. A. (1988). Studies on neotropical Violaceae 2. Arrangement of leaves, inflorescences and branchlets in neotropical *Rinorea*. *Flora* 180, 345–376.

Henriques, S. T., and Craik, D. J. (2012). Importance of the cell membrane on the mechanism of action of cyclotides. *ACS Chem. Biol.* 7, 626–636. doi: 10.1021/cb200395f

Hernandez, J. F., Gagnon, J., Chiche, L., Nguyen, T. M., Andrieu, J. P., Heitz, A., et al. (2000). Squash trypsin inhibitors from *Momordica cochinchinensis* exhibit an atypical macrocyclic structure. *Biochemistry* 39, 5722–5730. doi: 10.1021/bi9929756

Herrmann, A., Burman, R., Mylne, J. S., Karlsson, G., Gullbo, J., Craik, D. J., et al. (2008). The alpine violet, *Viola biflora*, is a rich source of cyclotides with potent cytotoxicity. *Phytochemistry* 69, 939–952. doi: 10.1016/j.phytochem.2007.10.023

Herrmann, A., Svangård, E., Claeson, P., Gullbo, J., Bohlin, L., and Göransson, U. (2006). Key role of glutamic acid for the cytotoxic activity of the cyclotide cycloviolacin O2. *Cell. Mol. Life Sci.* 63, 235–245. doi: 10.1007/s00018-005-5486-4

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. doi: 10.1074/jbc.M109.003384

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. doi: 10.1042/BJ20060627

Jennings, C. V., Rosengren, K. J., Daly, N. L., Plan, M., Stevens, J., Scanlon, M. J., et al. (2005). Isolation, solution structure, and insecticidal activity of kalata B2, a circular protein with a twist: do Mobius strips exist in nature? *Biochemistry* 44, 851–860. doi: 10.1021/bi047837h

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. doi: 10.1073/pnas.191366898

Keller, R. L. J. (2004). *The Computer Aided Resonance Assignment Tutorial*. Goldau: Cantina Verlag.

Koehbach, J., Attah, A. F., Berger, A., Hellinger, R., Kutchan, T. M., Carpenter, E. J., et al. (2013a). Cyclotide discovery in Gentianales revisited-identification and characterization of cyclic cystine-knot peptides and their phylogenetic distribution in Rubiaceae plants. *Biopolymers* 100, 438–452. doi: 10.1002/bip.22328

Koehbach, J., O'Brien, M., Muttenthaler, M., Miazzo, M., Akcan, M., Elliott, A. G., et al. (2013b). Oxytocic plant cyclotides as templates for peptide G protein-coupled receptor ligand design. *Proc. Natl. Acad. Sci. U.S.A.* 110, 21183–21188. doi: 10.1073/pnas.1311183110

Larsson, R., and Nygren, P. (1989). A rapid fluorometric method for semiautomated determination of cytotoxicity and cellular proliferation of human tumor cell lines in microculture. *Anticancer Res.* 9, 1111–1119.

Leta Aboye, T., Clark, R. J., Craik, D. J., and Göransson, U. (2008). Ultra-stable peptide scaffolds for protein engineering-synthesis and folding of the circular cystine knotted cyclotide cycloviolacin O2. *Chembiochem* 9, 103–113. doi: 10.1002/cbic.200700357

Lindhagen, E., Nygren, P., and Larsson, R. (2008). The fluorometric microculture cytotoxicity assay. *Nat. Protoc.* 3, 1364–1369. doi: 10.1038/nprot.2008.114

Lindholm, P., Göransson, U., Johansson, S., Claeson, P., Gullbo, J., Larsson, R., et al. (2002). Cyclotides: a novel type of cytotoxic agents. *Mol. Cancer Ther.* 1, 365–369. Available online at: http://www.jstor.org/stable/3094003

Mulvenna, J. P., Wang, C., and Craik, D. J. (2006). CyBase: a database of cyclic protein sequence and structure. *Nucleic Acids Res.* 34, D192–D194. doi: 10.1093/nar/gkj005

Munzinger, J. K., and Ballard, H. E. (2003). *Hekkingia* (Violaceae), a new arborescent violet genus from French Guiana, with a key to genera in the family. *Syst. Bot.* 28, 345–351. Available online at: http://www.jstor.org/stable/3094003

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. doi: 10.1002/bip.21419

Nguyen, G. K., Zhang, S., Nguyen, N. T., Nguyen, P. Q., Chiu, M. S., Hardjojo, A., et al. (2011). Discovery and characterization of novel cyclotides originated from chimeric precursors consisting of albumin-1 chain a and cyclotide domains in the Fabaceae family. *J. Biol. Chem.* 286, 24275–24287. doi: 10.1074/jbc.M111.229922

Ovesen, R. G., Brandt, K. K., Göransson, U., Nielsen, J., Hansen, H. C., and Cedergreen, N. (2011). Biomedicine in the environment: cyclotides constitute potent natural toxins in plants and soil bacteria. *Environ. Toxicol. Chem.* 30, 1190–1196. doi: 10.1002/etc.496

Padovan, L., Crovella, S., Tossi, A., and Segat, L. (2010). Techniques for plant defensin production. *Curr. Protein Pept. Sci.* 11, 231–235. doi: 10.2174/138920310791112101

Plan, M. R., Göransson, U., Clark, R. J., Daly, N. L., Colgrave, M. L., and Craik, D. J. (2007). The cyclotide fingerprint in *Oldenlandia affinis*: elucidation of chemically modified, linear and novel macrocyclic peptides. *Chembiochem* 8, 1001–1011. doi: 10.1002/cbic.200700097

Poth, A. G., Colgrave, M. L., Philip, R., Kerenga, B., Daly, N. L., Anderson, M., et al. (2010). The discovery of cyclotides in the Fabaceae plant family provides new insights into the cyclization, evolution and distribution of circular proteins. *ACS Chem. Biol.* 6, 345–355. doi: 10.1021/cb100388j

Poth, A. G., Mylne, J. S., Grassl, J., Lyons, R. E., Millar, A. H., Colgrave, M. L., et al. (2012). Cyclotides associate with leaf vasculature and are the products of a novel precursor in petunia (*Solanaceae*). *J. Biol. Chem.* 287, 27033–27046. doi: 10.1074/jbc.M112.370841

Pränting, M., Lööv, 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. doi: 10.1093/jac/dkq220

Sack, M., Hofbauer, A., Fischer, R., and Stoger, E. (2015). The increasing value of plant-made proteins. *Curr. Opin. Biotechnol.* 32, 163–170. doi: 10.1016/j.copbio.2014.12.008

Simonsen, S. M., Sando, L., Ireland, D. C., Colgrave, M. L., Bharathi, R., Göransson, U., et al. (2005). A continent of plant defense peptide diversity: cyclotides in Australian Hybanthus (Violaceae). *Plant Cell* 17, 3176–3189. doi: 10.1105/tpc.105.034678

Svangård, E., Burman, R., Gunasekera, S., Lövborg, H., Gullbo, J., and Göransson, U. (2007). Mechanism of action of cytotoxic cyclotides: cycloviolacin O2 disrupts lipid membranes. *J. Nat. Prod.* 70, 643–647. doi: 10.1021/np070007v

Svangård, E., Göransson, U., Hocaoglu, Z., Gullbo, J., Larsson, R., Claeson, P., et al. (2004). Cytotoxic cyclotides from *Viola tricolor*. *J. Nat. Prod.* 67, 144–147. doi: 10.1021/np030101l

Svangård, E., Göransson, U., Smith, D., Verma, C., Backlund, A., Bohlin, L., et al. (2003). Primary and 3-D modelled structures of two cyclotides from *Viola odorata*. *Phytochemistry* 64, 135–142. doi: 10.1016/S0031-9422(03)00218-8

Swofford, D. L. (2003). *PAUP*. Phylogenetic Analysis using Parsimony (*and Other Methods)*. 4.0b10. Available online at: http://paup.csit.fsu.edu

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. doi: 10.1073/pnas.96.16.8913

Tokuoka, T. (2008). Molecular phylogenetic analysis of Violaceae (Malpighiales) based on plastid and nuclear DNA sequences. *J. Plant Res.* 121, 253–260. doi: 10.1007/s10265-008-0153-0

Trabi, M., Mylne, J. S., Sando, L., and Craik, D. J. (2009). Circular proteins from Melicytus (Violaceae) refine the conserved protein and gene architecture of cyclotides. *Org. Biomol. Chem.* 7, 2378–2388. doi: 10.1039/b823020j

Trabi, M., Svangård, E., Herrmann, A., Göransson, U., Claeson, P., Craik, D. J., et al. (2004). Variations in cyclotide expression in *Viola* species. *J. Nat. Prod.* 67, 806–810. doi: 10.1021/np034068e

Wahlert, G., Ballard, H. Jr., and de Paula-Souza, J. (2015). Ixchelia, a new genus of Violaceae from Mexico and Mesoamerica. *Brittonia* 1–11. doi: 10.1007/s12228-015-9375-3

Wahlert, G. A., and Ballard, H. E. (2012). A Phylogeny of *Rinorea* (Violaceae) Inferred from Plastid DNA Sequences with an Emphasis on the African and Malagasy species. *Syst. Bot.* 37, 964–973. doi: 10.1600/036364412X656392

Wahlert, G. A., Marcussen, T., de Paula-Souza, J., Feng, M., and Ballard, H. E. (2014). A phylogeny of the Violaceae (Malpighiales) inferred from plastid DNA sequences: implications for generic diversity and intrafamilial classification. *Syst. Bot.* 39, 239–252. doi: 10.1600/036364414X678008

Wang, C. K., Colgrave, M. L., Gustafson, K. R., Ireland, D. C., Goransson, U., and Craik, D. J. (2008a). Anti-HIV cyclotides from the Chinese medicinal herb *Viola yedoensis*. *J. Nat. Prod.* 71, 47–52. doi: 10.1021/np070393g

Wang, C. K., Gruber, C. W., Cemazar, M., Siatskas, C., Tagore, P., Payne, N., et al. (2014). Molecular grafting onto a stable framework yields novel cyclic peptides for the treatment of multiple sclerosis. *ACS Chem. Biol.* 9, 156–163. doi: 10.1021/cb400548s

Wang, C. K., Kaas, Q., Chiche, L., and Craik, D. J. (2008b). CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. *Nucleic Acids Res.* 36, D206–D210. doi: 10.1093/nar/gkm953

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., et al. (1995). 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* 6, 135–140. doi: 10.1007/BF00211777

Wong, C. T., Rowlands, D. K., Wong, C. H., Lo, T. W., Nguyen, G. K., Li, H. Y., et al. (2012). Orally active peptidic bradykinin B1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment. *Angew. Chem. Int. Ed Engl.* 51, 5620–5624. doi: 10.1002/anie.201200984

Yusibov, V., Streatfield, S. J., and Kushnir, N. (2011). Clinical development of plant-produced recombinant pharmaceuticals: vaccines, antibodies and beyond. *Hum. Vaccin.* 7, 313–321. doi: 10.4161/hv.7.3.14207