Circular proteins from Melicytus (Violaceae) refine the conserved protein and gene architecture of cyclotides

Manuela Trabi,† Joshua S. Mylne, Lillian Sando‡ and David J. Craik*

Received 5th January 2009, Accepted 12th March 2009
First published as an Advance Article on the web 6th April 2009
DOI: 10.1039/b823020j

Cyclotides are cyclic disulfide rich mini-proteins found in various Rubiaceae (coffee family), Violaceae (violet family) and Cucurbitaceae (squash family) plant species. Within the Violaceae, cyclotides have been found in numerous species of the genus Viola as well as species from two other genera, namely Hybanthus and Leonia. This is the first in-depth report of cyclotides in the genus Melicytus (Violaceae). We present the chromatographic profiles of extracts of eight Melicytus species and one Melicytus hybrid that were found to contain these circular peptides. We isolated and characterised five novel cyclotides (mra1 to mra5) from the aerial parts of a common New Zealand tree, *Melicytus ramiflorus.* All five peptides show the characteristics of the bracelet subfamily of cyclotides. Furthermore, we isolated 17 non-redundant cDNA clones from the leaves of *Melicytus ramiflorus* encoding cyclotide prepropeptides. This detailed report on the presence of cyclotides in several species of the genus Melicytus further strengthens our hypothesis that cyclotides are ubiquitous in Violaceae family plants and provides additional insight into the biochemical processing mechanisms that produce the cyclic protein backbone of this unique family of ultra-stable plant proteins.

Introduction

In recent years there has been an ever growing number of naturally occurring circular proteins discovered in bacteria, plants and animals.[1,2] The largest family amongst these circular proteins are the cyclotides,[3] isolated from plants of the Rubiaceae, Violaceae and Cucurbitaceae families (Table 1). Cyclotides generally contain 28–37 residues, including six absolutely conserved cysteine residues, which are arranged in a cystine knot motif, in which two disulfide bonds and their connecting backbone segments form a ring that is penetrated by the third disulfide bond. Knotted disulfide arrangements are also known from animal toxins and growth factors,[4] but only in the cyclotides is the cystine knot combined with a circular peptide backbone, resulting in the so called cyclic cystine knot (CCK) motif.[5] The CCK motif renders the cyclotides exceptionally stable against enzymatic digestion and thermal degradation.[6] This remarkable stability, combined with the multitude of natural bioactivities, including anti-HIV,[7–10] antimicrobial[11] and insecticidal activities[12–14] make cyclotides a prime target for pharmaceutical and agrochemical applications.

Although neither its circular backbone nor its knotted structure was recognised at the time, kalata B1 was the first cyclotide reported in the literature. It was discovered as one of the active uterotonic ingredients of a decoction made from the dried aerial parts of the herbaceous plant *Oldenlandia affinis* DC (Rubiaceae), called kalata-kalata in the native Lulua language, used in the Congo to accelerate childbirth.[15–17] From the mid 1990s onwards, reports of related macrocyclic peptides in species from other Rubiaceae genera appeared. They included cyclopsychotride A from *Psychotria longipes*,[18] circulins A–F from *Chassalia parvifolia*[8,9] and palicourein from *Palicourea condensata.*[19] Earlier, Schöpke *et al.*[20] discovered the presence of macrocyclic peptides in *Viola arvensis* Murray and *V. tricolor* L., both plants of the family Violaceae. Over time, this plant family proved to be an excellent source of macrocyclic peptides, with examples found in species of the genera Viola,[3,21–24] Hybanthus[25] and Leonia.[26] The term cyclotides was introduced in 1999,[3] when it became clear that these macrocyclic peptides belonged to a common structural family comprising a CCK motif surrounded by six backbone loops between the conserved Cys residues. Fig. 1 shows the CCK motif and location of loops 1–6 in the prototypic cyclotide kalata B1.

To date, more than 100 cyclotides, summarised for example in references 27 and 28, have been characterised. They fall into three subfamilies distinguished by amino acid sequence and structural features. The two main subfamilies, termed Möbius and bracelet cyclotides, comprise sequences that are relatively conserved within, but not between the subfamilies. Both contain only three residues in loop 1, including an absolutely conserved Glu residue, as opposed to a six-residue loop 1 in the third subfamily, namely the trypsin inhibitor subfamily from Cucurbitaceae (squash family). In the latter subfamily, the amino acid sequences are very similar to those of linear squash trypsin inhibitors, also referred to as knottins.[29] Members of the Möbius subfamily are characterised by a conserved *cis*-proline residue in loop 5, which introduces a conceptual 180° twist in the circular peptide backbone.[3] This proline residue is missing in the bracelet and the trypsin inhibitor subfamilies.

Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia. E-mail: d.craik@imb.uq.edu.au, m.trabi@uq.edu.au, j.mylne@uq.edu.au, lillian.sando@csiro.au; Fax: +61 7 3346-2029; Tel: +61 7 3346-2019

† Current address: Queensland Institute of Medical Research, Herston Rd, Herston QLD 4006, Australia

‡ Current address: CSIRO Livestock Industries, Queensland Bioscience Precinct, 306 Carmody Rd, St Lucia, QLD 4067, Australia

Table 1 Cyclotides isolated from various Violaceae, Rubiaceae and Cucurbitaceae species

Species	Family	Cyclotide	Plant part	Reference
Oldenlandia affinis	R[a]	Kalata B1 to B17	Aerial	3,13,16,40,44
Viola arvensis	V	Violapeptide I	Aerial	20
		Varv peptide A to H	Aerial	21,45
Viola tricolor	V	Violapeptide 1	Aerial	20
Psychotria longipes	R	Cyclopsychotride A	Aerial	18
Chassalia parvifolia	R	Circulin A to F	Stem	9,10
Viola odorata	V	Cycloviolacin O1 to O11	Aerial	3
		Vodo M and N	Aerial	23
		Cycloviolacin O13 to O25	Whole	46
Viola hederacea	V	Cycloviolacin H1	Aerial	3
		Cycloviolacin H2 and H3	Several	34
		Cycloviolacin H4	Underground	47
		Vhr1	Roots	24
		Vhl 1 and 2	Leaves	34
Leonia cymosa	V	Cycloviolin A to D	Bark	26
Hybanthus parviflorus	V	Hypa A	Aerial	25
Hybanthus floribundus	V	HyflA-C	Aerial	36
Palicourea condensata	R	Palicourein	Bark	19
Viola cotyledon	V	Vico A and B	Aerial	22
Viola tricolor	V	Tricyclon A	Flowers	31
Viola baoshanensis	V	Viba 1 to 17	Whole	38
Momordica cochinchinensis	C	MCoTI-I	Seeds	29
	C	MCoTI-II	Seeds	29
Melicytus ramiflorus	V	Mra1-5	Leaves	This work

[a] C: Cucurbitaceae, R: Rubiaceae, V: Violaceae.

As indicated in Fig. 1, in *O. affinis* cyclotides are synthesised as prepropeptides consisting of a signal sequence, a variable proregion, an N-terminal repeat region and between one and three mature cyclotide domains, separated by N-terminal repeats.[13] The first cyclotide precursor sequences discovered showed that the sequences associated with the cyclotide domain were flanked at each end by a GLP peptide triad. The mature cyclotide retains just one copy of the GLP sequence, which could in principle be derived entirely from one of the flanking triads, or partially from both, depending on the initial processing site; so this creates four possible processing sites. However, in the clone encoding the cyclotide kalata B2, the cyclotide sequence is flanked N-terminally by a GLP triad, but C-terminally by an SLP triad (Fig. 1b), suggesting processing N-terminally of the GLP and SLP elements. Recently, analogous prepropeptides have been isolated from two species of the genus Viola, namely *V. odorata*[30] and *V. tricolor*,[31] showing a similar modular structure with between one and three mature cyclotide domains. A key feature of all precursor sequences examined so far is the presence of an Asn residue (or less frequently Asp) at the C-terminal processing site. Consistent with this observation, two recent reports provided evidence that the cyclisation event is facilitated by an asparaginyl endoproteinase.[32,33] Although the amount of sequence data from different species are growing, an understanding of the processing, folding and cyclisation events necessary to produce mature circular cyclotides will be facilitated by a larger number of sequences.

In the current study we screened a range of species of the genus Melicytus (Violaceae) to shed more light on the distribution of cyclotide-expressing species in the Violaceae, and the mechanism of cyclotide biosynthesis in this family. Most cyclotides reported until now from Violaceae plants were derived from herbaceous species. The genus Melicytus, however, contains deciduous shrubs and trees native to New Zealand, Norfolk Island, Fiji and the Solomon Islands. *Melicytus ramiflorus* Forster & Forster (common name mahoe or whitey wood) has been used by the Maori against scabies and is still sold by a New Zealand company as a natural remedy for diarrhoea, rheumatism and various other illnesses—with the probably very appropriate warning not to use it during pregnancy, given the early indication of uterotonic activity of kalata B1 in native medicine applications.[17] Here we report the presence of cyclotides in eight Melicytus species as well as in one Melicytus hybrid (for simplicity, the hybrid will be treated as a species in the remainder of the text). Furthermore, we describe the isolation and sequencing of cyclotide peptides and cDNA clones from the leaves of *M. ramiflorus.* The new sequences confirm an emerging view that cyclotides seem to be ubiquitous in the Violaceae (but sparse in the Rubiaceae) and provide additional insights into the biochemical basis of the processing mechanisms of circular proteins.

Results

We collected a diverse range of Melicytus species to explore the distribution of cyclotides in a genus within the Violaceae that had not previously been screened for their presence. The samples included shrubs (*e.g. Melicytus alpinus*, *Melicytus nova-zealandiae*) and trees (*e.g. Melicytus ramiflorus*, *Melicytus lanceolatus*), several of them (*e.g. Melicytus crassifolius*, *Melicytus flexuosus*) endemic to New Zealand. The samples were collected at several different locations in New Zealand (both North and South Island) and Australia. A methanol–DCM extraction procedure developed in our laboratory and described previously[3] proved to be fast, reliable and well suited for the extraction of rather hydrophobic peptides like the cyclotides from a range of tissues in the Melicytus species.

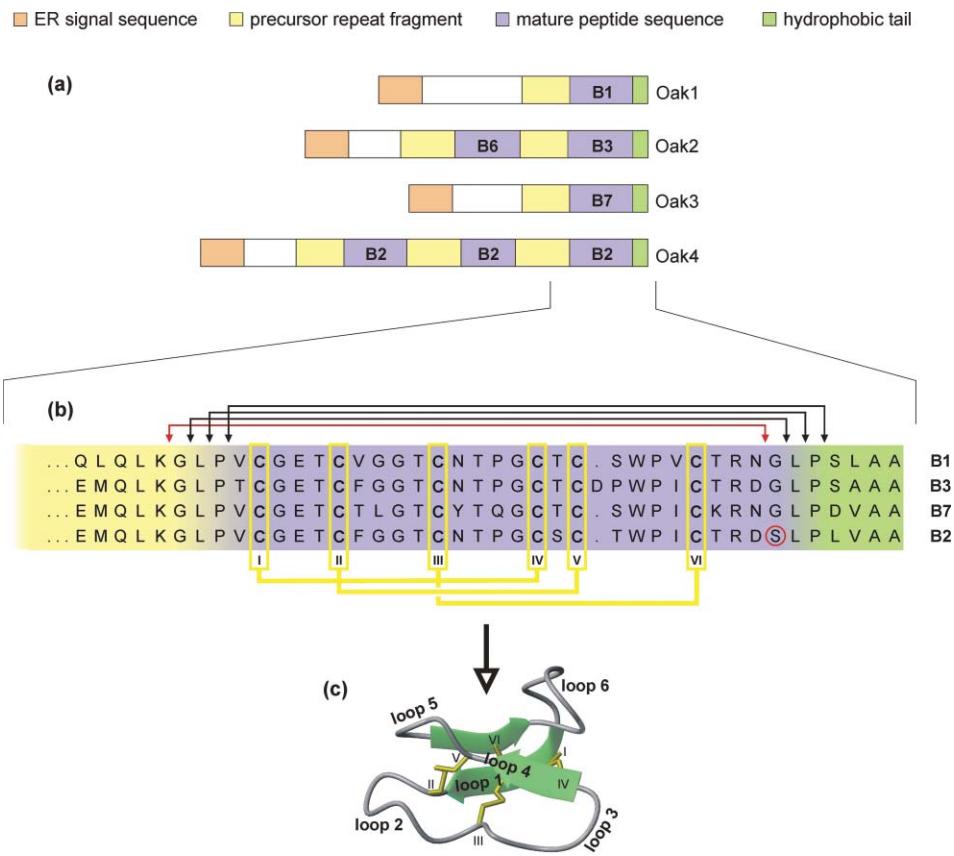

Fig. 1 Precursor organisation, processing and three dimensional structure of cyclotides. (a) Cyclotide precursor proteins deduced from *O. affinis* cDNA clones contain a signal sequence, a variable pro-region (given in white), between one and three mature cyclotide domains separated by highly conserved precursor repeat fragments, and a short hydrophobic tail. (b) Most mature cyclotide sequences are flanked by two GLP sequences, creating the four possible processing sites marked with arrows. However, the presence of a Ser residue (red circle) in *Oak4* suggests processing N-terminally of the GLP/SLP repeat (red arrow). The absolutely conserved Cys residues are boxed and numbered with Roman numbers. Disulfide connectivities are indicated by yellow lines below the sequences. (c) The solution structure of kalata B1 (PDB ID code 1NB1) shows the very compact cyclotide fold, stabilised by the cystine knot core. The backbone segments, or loops, between two adjacent Cys residues are numbered; Cys residues are again marked with Roman numbers.

An advantage of the method is that minimal handling of the samples is required and thus it is suited to small quantities (a few grams) of plant tissue. HPLC traces of the crude extracts generally showed no peaks other than cyclotides eluting between 50 and 70% solvent B (corresponding to 45–63% acetonitrile), facilitating both screening by LC-MS and further purification by preparative RP HPLC.

MS/MS studies showed that by changing the collision energy of the mass spectrometer it is possible to distinguish between backbone cyclised and open chain peptides.[34,35] Circular peptides such as cyclotides preferentially withstand higher collision energy without excessive fragmentation. Overall, the presence of cyclotides in crude extracts was deduced by a combination of late elution time, a mass within the range typical for cyclotides (~2100 to 3900 Da) and stability under harsh MS conditions.

Using this protocol, nine Melicytus species (*M. ramiflorus*, *M. angustifolius*, *M. flexuosus*, *M. alpinus*, *M. crassifolius*, *M. lanceolatus*, *M. micranthus* and *M. nova-zealandiae*, *M.* hybrid) were screened for expression of cyclotides. All plants investigated were found to contain cyclotides. In general the cyclotide profiles were rather complex (Fig. 2), suggesting the presence of up to 50 different cyclotides per plant species. As already seen in a number of Viola species,[24,35] the cyclotide profiles showed differences depending on the tissue tested (Fig. 3). In total we found 492 different masses, comprising those shown in Fig. 2 and 3 together with 261 masses not shown because they are below the nominal 30% intensity threshold. Of these, 388 were novel (*i.e.*, not corresponding to masses of known cyclotides currently listed in Cybase).[28]

We selected *Melicytus ramiflorus* for more detailed analysis for several reasons; it is the most prominent Melicytus species in New Zealand, it is cultivated, was available in kilogram quantities and has been used by the indigenous Maori for centuries as a natural medicine. We selected peaks from *Melicytus ramiflorus* where it was possible to purify single components for further analysis. To determine their primary sequences, the selected cyclotide peaks were reduced, alkylated and purified prior to enzymatic digest. The peptides were sequenced using Edman degradation and/or MS/MS. These studies led to the identification of five novel cyclotides, mra1–5 (mra for *Melicytus ramiflorus*), whose sequences are shown in Fig. 4. For comparison, a selection of

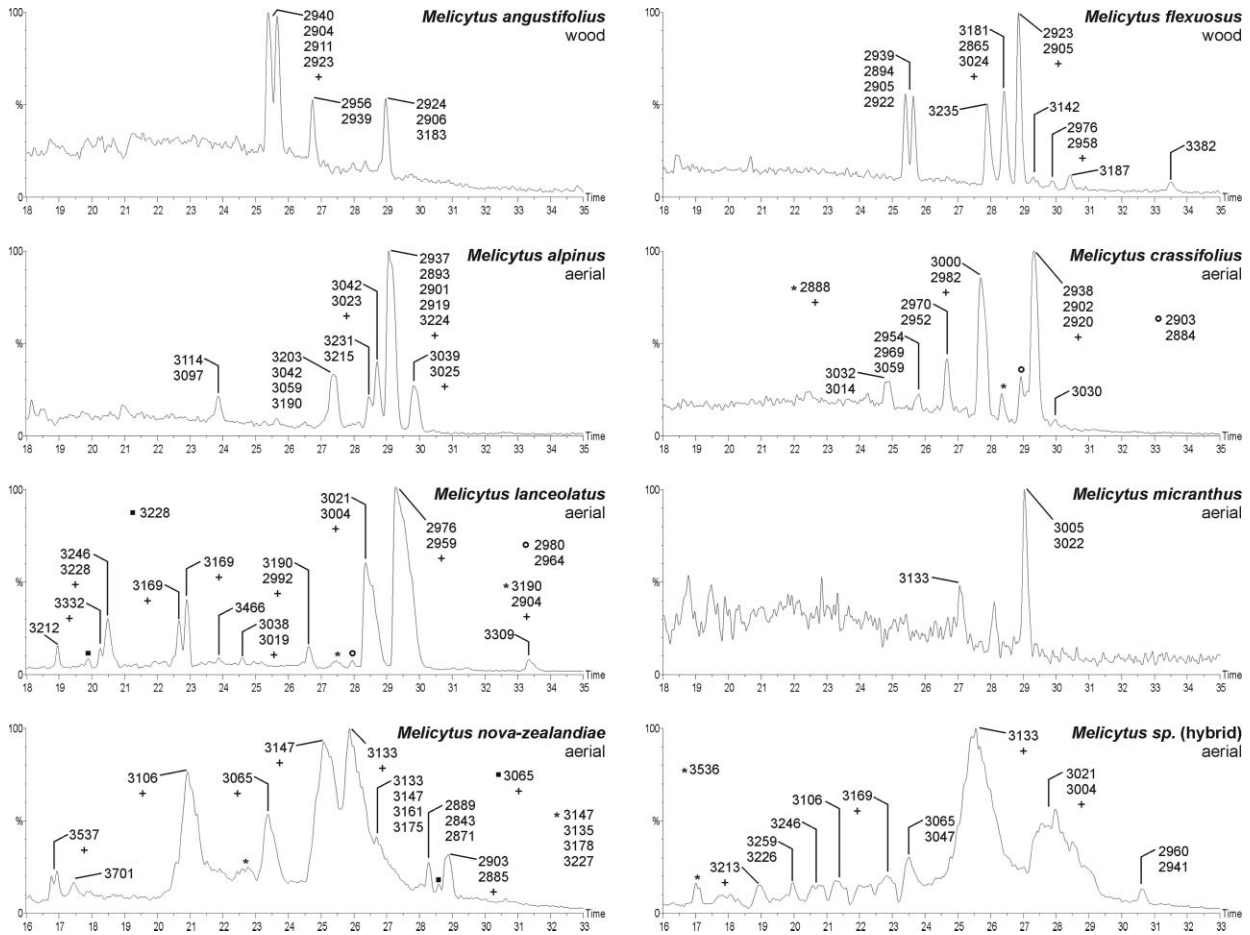

Fig. 2 LC-MS traces of various Melicytus species. The cyclotide-containing region of the chromatographic trace is shown with the peaks labelled with the corresponding peptide masses. The presence of one or more minor components with a signal intensity of less than 30% of that of the strongest corresponding signal is indicated by a plus ('+'); the masses of these components are not given in the figure, but approximately 231 such unique masses were observed. HPLC conditions: flow rate 200 µl per minute, 0–80% solvent B′ over 40 minutes. Additional traces of another species, *M. ramiflorus* are shown in Fig. 3.

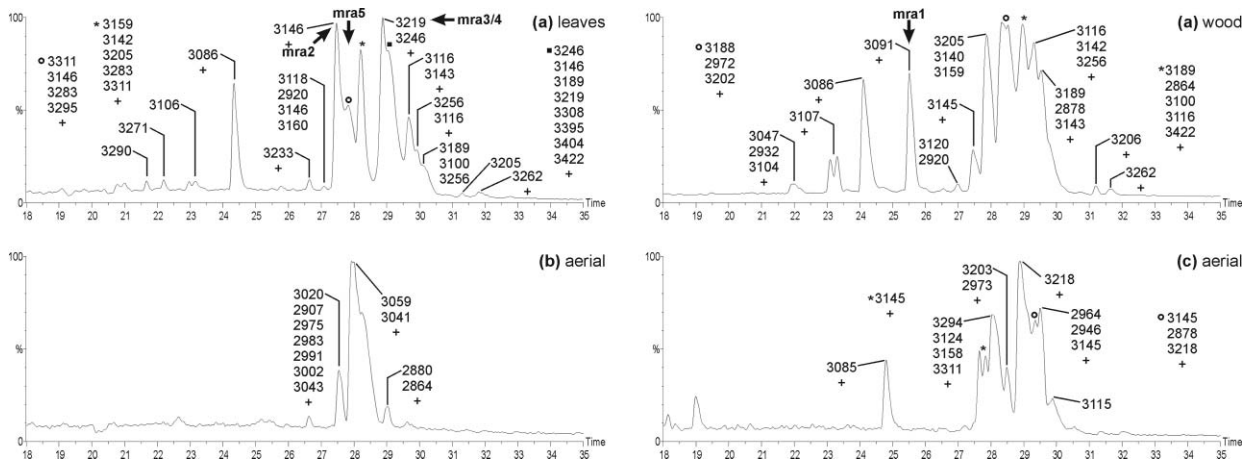

Fig. 3 LC-MS traces of *M. ramiflorus* showing the regional and tissue specific differences of crude extracts. Plant material was collected at the Royal Botanic Gardens Melbourne **(a)**, Auckland Botanic Gardens **(b)** and Christchurch Botanic Gardens **(c)**, respectively. The cyclotide region is shown, with the peaks labelled with the corresponding peptide masses. The presence of one or more minor components with a signal intensity of less than 30% of that of the strongest corresponding signal is indicated by a plus ('+'); the masses of these components are not given in the figure. HPLC conditions: flow rate 200 µl per minute, 0–80% B′ in 40 minutes.

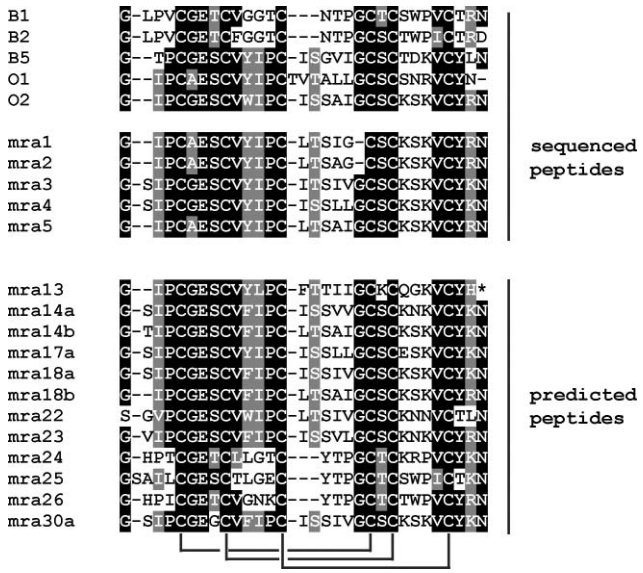

Fig. 4 Non-redundant *Melicytus ramiflorus* cyclotide sequences obtained through peptide sequencing or deduced from cDNA sequences. A selection of known cyclotides from *Oldenlandia affinis* (B1, B2, B5) and *Viola odorata* (O1, O2) is also shown. Mra1–5 were sequenced by Edman degradation and MS/MS. *Mra13* to *mra26* were deduced from translated cDNA clones. In clones with two cyclotide domains, the cyclotide closest to the amino terminus is given the suffix "a" and the one closer to the carboxyl terminus "b". The brackets on the bottom of the figure indicate the disulfide connectivities. Only unique cyclotide sequences are included; several sequences were determined both on peptide and cDNA level or cloned multiple times, respectively (mra4 = *mra4a* and *mra16a;* mra5 = *mra4b; mra14b = mra20* and *mra21; mra18a = mra19a*, *mra28* and *mra29a; mra18b = mra19b, mra28b, mra29b* and *mra30b*).

cyclotides from two other species, namely *Oldenlandia affinis* (Rubiaceae) and *Viola odorata* (Violaceae), is also shown.

To complement the discovery of cyclotide peptide sequences, we made cDNA from RNA isolated from *M. ramiflorus* leaves and probed it for cyclotide-like sequences using a range of different strategies to maximise sequence length and diversity. Conservation of the ER signal sequence AAFALPA among some Violaceae[36] prompted us to clone close to full length cyclotide precursor genes by 3′ RACE using a degenerate primer for the AAFALPA sequence. We also designed degenerate primers based on the newly discovered *M. ramiflorus* peptide sequences to target the corresponding cDNA clones. Furthermore, we used degenerate primers against conserved cyclotide sequences as well as specific primers designed against the 3′ UTR of 3′ RACE fragments in order to amplify close to full length clones. All the clones displayed the general architecture previously reported for clones from other cyclotide-bearing plant species.[13,30,31]

We isolated a total of 29 cyclotide precursor cDNA clones, 17 of which were unique. The 17 unique cDNA clones, summarised in Table 2, are comprised of two full length cDNAs (*Mra4*, *Mra13*) isolated by 5′ RACE using specific primers; six close to full-length cDNAs (*Mra18, Mra22, Mra23, Mra24, Mra25, Mra26*) isolated by 3′ RACE using a primer designed against the conserved ER signal sequence AAFALPA; two partial sequences (*Mra16* and *Mra17*) amplified with the AAFALPA primer and a specific reverse primer designed to complement to amino acids sequence PCGESCVYIPCIS; two partial sequences (*Mra20* and *Mra21*) amplified by 3′ RACE using a degenerate primer designed against amino acid sequence GIAGCSCKNK; three close to full-length cDNAs (*Mra28, Mra29, Mra30*) amplified with the AAFALPA primer and a specific primer designed against the 3′ UTR of *Mra20*; and two close to full-length cDNAs (*Mra14*, *Mra19*) amplified with the AAFALPA primer and a specific primer designed against the 3′ UTR of *Mra21.*

Of the 17 unique cyclotide precursor cDNAs, six encoded single cyclotide domain genes (*Mra13, Mra22, Mra23, Mra24, Mra25, Mra26*), seven represented two cyclotide domain genes (*Mra4, Mra14, Mra18, Mra19, Mra28, Mra29, Mra30*) and four of the partial cDNAs (*Mra16, Mra17, Mra20, Mra21*) encoded two or potentially more cyclotide domains. The cyclotide precursor proteins encoded by the 17 unique cDNAs contained 24 complete cyclotides, of which 14 sequences are unique. Included in these 14 unique cyclotides are mra4 and mra5, which are encoded by *Mra4*. There were no cDNA sequences isolated that supported the isolated peptides mra1, mra2 or mra3. Therefore, we have determined a total of 17 new cyclotide sequences on cDNA and peptide level. One full-length cDNA (*Mra13*) was unusual in that it contained one mature cyclotide domain, but lacked the typical C-terminal Asn/Asp as well as the hydrophobic tail region. On this basis, it is most likely a linear derivative of a cyclotide, much like the previously reported violacin A.[37]

All except for three of the new cyclotides discovered in this work belong to the bracelet subfamily. Three out of the 17 (18%) new cyclotides show an AES triad in loop 1, a slightly higher proportion (13%) than seen in previously published bracelet sequences.[28] The new cyclotides are designated mra (for their origin, *M. ramiflorus*) followed by sequential numbers. To distinguish between cyclotide sequences discovered on the peptide level and those deduced from cDNA clones, the names for the latter are given in italics (*e.g. mra25*) throughout the text. For precursors that contain two

Table 2 cDNA clones isolated from *M. ramiflorus*

Clone	Method	Primer	Completeness	Mature protein sequences
Mra4, *Mra13*	5′ RACE	Specific primers	Full length	2 (*Mra4*); 1 (*Mra13*)
Mra18, *Mra22-Mra26*	3′ RACE	AAFALPA[a]	Close to full length	2 (*Mra18*); 1 (*Mra22-Mra26*)
Mra16, *Mra17*		AAFALPA[a] + specific reverse primer[b]	Partial	2 or more
Mra20, *Mra21*	3′ RACE	Degenerate primer[c]	Partial	2 or more
Mra28-Mra30		AAFALPA[a] + specific primer[d]	Close to full length	2
Mra14, *Mra19*		AAFALPA[a] + specific primer[e]	close to full length	2

[a] Conserved ER signal sequence. [b] To complement the amino acid sequence PCGESCVYIPCIS. [c] Against the amino acid sequence GIAGCSCKNK. [d] Designed against the 3′ UTR of *Mra20.* [e] Designed against the 3′ UTR of *Mra21.*

cyclotide domains, the cyclotide closest to the amino terminus was given the suffix "a" and the one closer to the carboxyl terminus "b" (*e.g. mra18a*, *mra18b*). Although detected in previous studies,[13,30] no full length three-domain cyclotide precursor genes were found in the current study, although some or all of the partial cDNA clones identified here (*Mra16*, *Mra17*, *Mra20*, *Mra21*) could contain a third cyclotide domain in the unsequenced region.

Discussion

In this study we investigated several species of the previously unscreened genus Melicytus within the Violaceae and found that it, like all other genera within the Violaceae examined so far, contains cyclotides. Using a combination of peptide-based and nucleic-acid based screening we sequenced 17 new cyclotides, but based on mass and chromatographic profiling we suggest that there are at least several hundred unreported cyclotides present within the nine Melicytus species examined.

Cyclotide extraction and profiling

The main characteristics of cyclotides are their hydrophobic nature and resultant late elution time on RP HPLC, and their exceptionally stable CCK framework built around three absolutely conserved disulfide bonds. We used these characteristics to deduce the likely presence of cyclotides in the screening program on Melicytus species—in this way we would predict that there are potentially more than 492 different cyclotides (including 388 new ones) present in the nine species examined. Our extraction process typically eliminates non-cyclotide compounds that would give comparable elution times on RP HPLC. Plants that do not contain cyclotides typically show none or only a small number of peaks in the cyclotide-typical retention time regime. Furthermore, since all cyclotides discovered so far range in size from 2100 to 3900 Da, any false positives based on retention time can be easily distinguished by their masses. In general, the degree of fragmentation of proteins in MS studies depends largely on the collision energy applied. The circular backbone and to some extent the cysteine cross-bracing of the cyclotides inhibits this fragmentation, resulting in the same m/z independent of the collision energy applied, thus setting the cyclotides apart from other proteins of similar size. The combination of these four factors, *i.e.*, tailored extraction process, late elution time, characteristic mass range and exceptional stability, allows for a confident identification of cyclotides in the screening process, which can be augmented by chemical reduction studies.

Cyclotide profiles

All Melicytus species screened in this study contained multiple cyclotides. For some species, for example *M. micranthus*, only dormant wood was available, but even in these samples the presence of peptides with cyclotide characteristics was verified. Some Melicytus species contained comparatively early-eluting cyclotides and an unusually high number of peptides that mainly gave rise to M + 4 peaks in mass spectra, indicating the presence of more than the 'average' number of amino acids amendable to protonation, such as Lys or Arg residues. On average each species expresses approximately 50 cyclotides. Some cyclotides, such as kalata B1 or varv A, are found in multiple Violaceae members, while others appear to be unique to individual species.

Novel sequences

Although none of the cyclotides isolated from *M. ramiflorus* varies extensively from already known cyclotide sequences, there are a range of differences to previously described cyclotides. All cyclotides isolated here on the peptide level belong to the bracelet subfamily. In contrast to members of the Möbius subfamily, bracelet cyclotides lack the *cis*-proline residue in loop 5. Although both Rubiaceae and Violaceae plants express bracelet as well as Möbius cyclotides, it seems that bracelet cyclotides predominate in both plant families: including the sequences presented in this study, 70 bracelet cyclotides have been described to date, but only 39 Möbius cyclotides.

Typically, the amino acid sequence of the first loop in the cyclotide backbone is a highly conserved GE(S/T) triad, regardless of whether the peptide belongs to the Möbius or bracelet subfamily. In the newly discovered sequences we see some exceptions to this general rule, *e.g.* GEG in *mra30a.* A related exception also is found in recently discovered VbCP5 with an atypical GER sequence in loop 1.[38] Peptides mra1, mra2, and mra5 have an AES triad, which has been seen in other Violaceae family members. The newly discovered sequences support the previous proposal that a small amino acid (Ala or Gly) is needed at position 1 of loop 1 to allow the formation of the embedded disulfide/backbone ring that is at the core of the cystine knot.[39] The central Glu residue in loop 1 is absolutely conserved in all the sequences presented in this work, as has been seen in all previously published cyclotides, with the exception of kalata B12, in which this residue is replaced with an Asp.[40]

In mra1 and mra2, loop 3 consists of only five residues instead of the usual six to seven. In the cyclotides *mra24*, *mra25* and *mra26*, loop 3 comprises only four residues. The solution structures of two bracelet cyclotides, circulin A and cycloviolacin O1, have been determined and in both cases loop 3 contains a short helical segment.[3,39,41] In the solution structure of kalata B1, the prototypical Möbius cyclotide, the four residues in loop 3 form a type II turn.[39] It can be assumed that in mra1 and mra2, due to the smaller loop size, this region of the molecule would also form a turn rather than a helical segment, resembling a Möbius cyclotide in that respect.

In the cyclotides described here, loop 6 consists of between six and eight residues, including a common GSIP motif also found in vico A and vico B.[22] With 15 different loop sequences known for loop 6, this region can be regarded as rather variable, a surprising finding for a region that has been proposed to be involved in cyclisation.[13] However, in all cases there is a conserved Asn/Asp residue in this loop that has been shown to be vital for cyclisation *via* the action of an asparaginyl endoproteinase.[32,33] Thus, the new sequences strongly support a role for an endogenous asparaginyl endoproteinase in the cyclisation process.

Cyclotide clones

Using a range of cloning strategies we isolated 17 unique cDNA clones, including two full-length clones, eleven close-to-full-length and four partial cDNA clones from the leaves of *M. ramiflorus.*

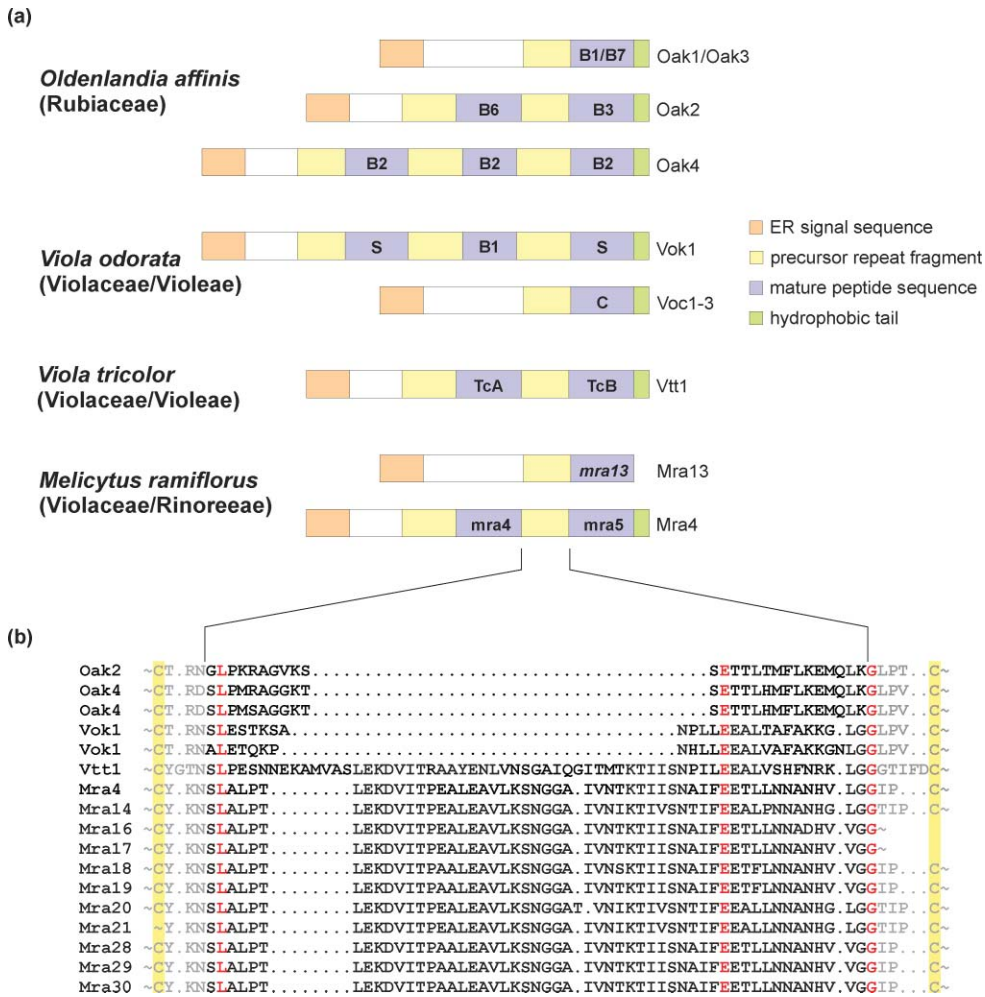

Fig. 5 (a) Organisation of the *M. ramiflorus* clones in comparison with clones isolated from *O. affinis*,[13] *V. odorata*[30] and *V. tricolor.*[31] (b) Deduced amino acid sequences of the internal precursor repeat fragments in clones containing more than one mature cyclotide domain. In this region, only very few residues, given in red, are absolutely conserved across both Rubiaceae and Violaceae clones. The flanking cyclotide domains are given in grey with the Cys residues on yellow background.

The general architecture of these clones is similar to previously reported genes from *O. affinis* (Rubiaceae),[13] *V. odorata* (Violaceae)[30] and *V. tricolor* (Violaceae),[31] as summarised in Fig. 5a. The endoplasmic reticulum signal sequence is followed by a precursor repeat fragment and one or more cyclotide domains separated again by the precursor repeat fragment. All full- and close-to-full-length clones except for one (*Mra13*, see below) end with a short hydrophobic tail region after the last mature cyclotide domain.

Interestingly, the precursor repeat fragments preceding the cyclotide domains differ significantly between species. Fig. 5b shows a comparison of the precursor repeat fragment sequences flanked by two mature cyclotide domains. The length of these repeats varies between 25 and 62 residues. Only one residue is absolutely conserved: a Leu residue two amino acids downstream from the first cyclotide domain. One attribute of the first residue after the cyclotide domain is conserved as in every case it has a small side chain (Ala, Ser, Gly).

Although the precursor repeat fragments from Melicytus clones are more than twice as long as those from *O. affinis* and *V. odorata* clones, the region about 10–15 residues upstream of the mature cyclotide domain seems to show some conservation with the consensus sequence ETTL for Rubiaceae and EEALL for Violaceae, respectively. NMR investigations of the precursor repeat fragments of *O. affinis* and *V. odorata* clones showed this conserved region to be situated towards the N-terminus of an amphipathic helix thought to be involved in precursor processing.[30]

Perhaps surprising was the lack of cDNAs for mra1, mra2 or mra3. These three peptides were some of the more abundant peptides isolated from *M. ramiflorus*, but despite using a variety of cloning strategies on the same leaf tissue we were unable to isolate their corresponding cDNAs. This could be due to our predominant use of the conserved AAFALPA primer, which was designed against three of the four cDNA sequences from *Viola odorata.*[36] Although this region is highly conserved and use of this primer has been very effective for isolating cyclotide precursor cDNA sequences from the Violaceae, it is unlikely that all cyclotide precursor genes have the AAFALPA ER signal sequence. *Mra13* for example, was isolated by 5′ RACE and encodes AAFAALD instead of AAFALPA. Furthermore, the AAFALPA primer is

not fully degenerate; it only accounts for 162 different DNA sequences which, when translated, give an AAFALPA amino acid sequence. There are over 8,000 possible cDNA sequences that encode AAFALPA and so even if a gene encodes AAFALPA, if it differs from the primer it may reduce the efficiency of amplification or not be amplified at all. The AAFALPA primer remains our most effective tool for rapidly isolating full length cDNA clones from Violaceae members, but it may be biasing the cDNA clones we isolate away from the most highly expressed cyclotide precursor cDNAs to those with AAFALPA primer-like sequences. Other workers found that using the AAFALPA primer with *Viola biflora* strongly favoured amplification of *Vbc5* which encodes cycloviolacin O9, despite this cyclotide being a minor component of the peptide profile.[42]

Alignment of the sequences from *M. ramiflorus* with a recent set from *Viola baoshanensis* submitted to GenBank as well as other previously isolated genes[30,31,37] reveals that two short regions in the precursor repeat fragment of Violaceae are very highly conserved; KT(I/V)SN and EE(A/T)LL (Fig. 5). Two of the three precursor repeat fragments from *V. odorata* are the only exception to this rule lacking the KT(I/V)SN region.

A Leu residue two residues upstream from the N-terminal side of the mature cyclotide sequence was previously found to be absolutely conserved,[30,31] but in *M. ramiflorus* it is frequently replaced by a Val residue (Fig. 6a). In one case this region is atypical; the sequence upstream of the predicted cyclotide domain in *Mra23* (KTDATN) does not contain any hydrophobic residues. Atypical regions immediately upstream of the predicted cyclotide domain have also been seen in *V. baoshanensis* genes *VbCP2* (LDNGRN) and *VbCP4* (ANLKAGN),[38] which lack the Leu residue two residues upstream of the predicted mature cyclotide. It is interesting to note that the three rare exceptions to a Leu/Val residue two residues upstream from the most N-terminal cyclotide residue have an Asn residue next to the starting Gly of the mature cyclotide sequence. Therefore these precursors may be N-terminally processed by an asparaginyl endoproteinase, *i.e.*, the new sequences suggest the possibility of more than one mechanism for N-terminal processing.

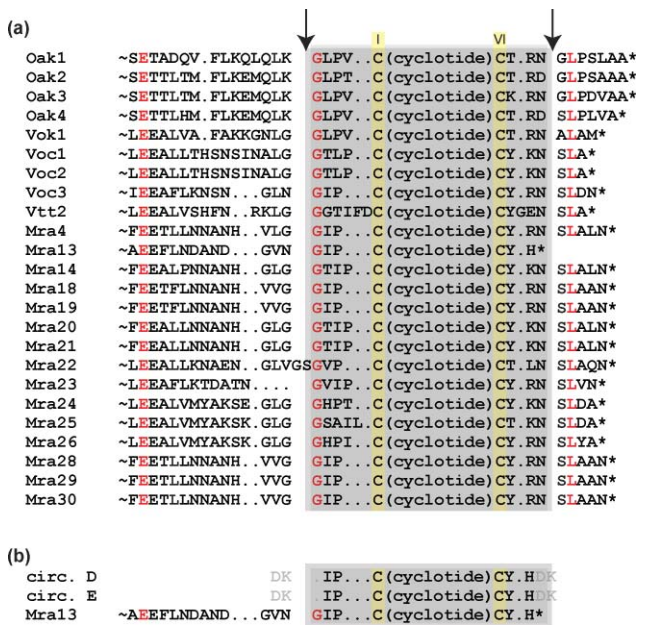

Fig. 6 (a) Summary of the processing sites around the mature cyclotide domains. The processing sites are indicated by arrows and absolutely conserved residues are red. (b) *Mra13* in comparison with circulins D (circ. D) and E (circ. E). The residues given in grey in the two circulin sequences indicate that the location of backbone cyclisation is not known for these two cyclotides and deduction of the processing site is impossible due to their sequences in loop 6 not conforming to the pattern seen in other cyclotides.

As mentioned before, cyclotide clones usually end in a short hydrophobic tail fragment. In clone *Mra13*, however, this hydrophobic tail is replaced by a premature stop codon at a position upstream of the putative precursor processing site. Two other cyclotides, circulins D and E isolated from *Chassalia parvifolia* (Rubiaceae),[10] show features similar to those of the deduced sequence of *mra13*, including a His residue in the same position as in *mra13*, a Lys residue in loop four and a somewhat unusual sequence in loop 6 (Fig. 6b). Since neither the cDNA for these two circulins nor the mature cyclotide mra13 have been isolated, it remains uncertain whether the two head-to-tail cyclised circulins are also encoded by a clone with premature stop codon or whether *Mra13* actually encodes for a cyclotide that is not backbone cyclised upon excision from the precursor protein. The latter is more likely and has some parallel with the linear cyclotide derivative violacin A from *V. odorata*.[37]

All Melicytus clones (except for *Mra13*) confirm the C-terminal processing site of the precursor to be after Asp/Asn. The absolutely conserved Leu residue one position away from the C-terminal processing site is also present in all Melicytus clones (again with the exception of *Mra13*). Jennings *et al.*[13] and Mulvenna *et al.*[31] suggested that an Asn-specific proteinase might be responsible for processing C-terminally of the mature cyclotide sequence and this has recently been confirmed,[32,33] but the role of the absolutely conserved Leu residue is still under investigation.

In conclusion, this study has helped to refine the structural elements typical for cyclotide precursor genes, namely a hydrophobic tail, an NTR region that is conserved across related species, an absolutely conserved Leu residue two residues downstream from the mature cyclotide domain, and an Asn residue at the end of the mature cyclotide sequence. In addition to more than doubling the number of cyclotide precursor sequences known, we have shown that cyclotides are found in shrubs and trees of the Violaceae family—not just herbaceous plants. We now believe that cyclotides are ubiquitous to the family Violaceae which includes some 900 species[43] that vary greatly in habitat and appearance, with two examples of its diversity being the commonly used garden edging sweet violet (*Viola odorata*) and whiteywood trees from the forests of New Zealand (*Melicytus ramiflorus*). This proposed ubiquitous distribution of cyclotides in the Violaceae is very different to the scattered appearance we find within the Rubiaceae and Cucurbitaceae families.

Experimental section

For several applications (*e.g.* HPLC, flash chromatography, redissolving of extracts) two standard solvents were used, referred to as solvent A (0.05% TFA in water) and solvent B (90% acetonitrile and 0.05% TFA in water) throughout the text.

Plant material

M. ramiflorus for cyclotide and RNA isolation was collected from the Royal Botanic Gardens Melbourne, Australia. Plant material for screening purposes was obtained from Auckland Regional Botanic Gardens (Manurewa), Wellington Botanic Garden and Christchurch Botanic Gardens (all New Zealand).

Peptide extraction

Plant material was extracted following a method developed in our laboratory and described earlier.[3,24] Briefly, the ground plant material was extracted overnight with methanol–DCM (1 : 1). After filtration through cotton wool the filtrate was mixed with chloroform and small amounts of water until a phase separation was visible. The lower (organic) phase, containing all the chlorophyll, was discarded, the upper (aqueous) phase, containing the cyclotides, was dried under vacuum first on a rotary evaporator (Büchi), then on a lyophyliser (Dynavac FD12), yielding what will be further referred to as crude extract.

Initial purification by flash chromatography

For large samples, *i.e.* extraction of more than 500 g of fresh plant material, an initial purification step was added to the procedure in order to minimise necessary HPLC runs. An open glass column (diameter approximately 3 cm) was filled with C18 reversed phase material (Matrex silica C18, pore diameter 60 Å, particle size 35–70 µm; Amicon, Switzerland) and pre-conditioned for about half an hour with 50% solvent B followed by 100% solvent A. The crude extract was dissolved in a minimal amount of solvent A and put onto the flash column without filtering. The components were eluted with 250 ml each of 0% B, 10% B, 20% B, 30% B, 50% B, 70% B and 100% B. Depending on the plant species, cyclotides were typically found in the 50 and 70% fraction. These fractions were lyophilised, redissolved in a minimal amount of up to 5% solvent B and subjected to purification by semi-preparative HPLC.

Purification and isolation of cyclotides

Cyclotides were isolated and purified using either preparative columns (Vydac 218TP or Phenomenex Jupiter, 250 × 22 mm, 5 µm, 300 Å) run at a flow rate of 8 ml per minute or semi-preparative columns (Vydac 218TP, 250 × 46 mm, 5 µm, 300 Å) run at 3 ml per minute. In both cases, the cyclotides were eluted with a 1% per minute gradient from 0 to 80% B and detected by UV at 230 nm. Two different HPLC systems were used for preparative HPLC: a Waters 600 HPLC system with Waters 600 MS controller, Waters 484 MS tuneable absorbance detector and Applied Biosystems Model 600 Data Analysis System software (version 1.0.1) running on an Apple Macintosh computer or an Agilent 1100 series system with degasser, quaternary pump, variable wavelength detector and Chemstation for LC software (Agilent Technologies) running on a PC.

Screening

LC-MS was performed with two different Phenomenex Jupiter columns (150 × 2 mm or 50 × 2 mm, 5 µm, 300 Å, both equipped with a SecuriGuard guard column) run at 200 µl per minute. Since TFA interferes with ionisation of the peptides in the ion source and therefore decreases the sensitivity of the mass spectrometric analysis, the solvents used for both LC-MS and MS/MS were A′ (0.05% formic acid in water, v/v) and B′ (90% acetonitrile and 0.05% formic acid in water, v/v). Peptides were eluted with a 2% per minute gradient from 0 to 80% solvent B′. The eluent was fed directly (*i.e.* without split system) into the ion spray source of a Micromass time-of-flight mass spectrometer controlled by a PC running MassLynx version 3.5 (Micromass Ltd., UK). Mass spectra were obtained in positive ion mode over a range m/z 700–2000.

Sequence determination

Because of the circular backbone and the knotted disulfide pattern, cyclotides are not directly amenable to sequencing by Edman degradation or mass spectrometry. The following steps for reduction and alkylation of the disulfide bonds and enzymatic digest were used for both isolated peptides and peptide mixtures.

The peptide or peptide mixture was dissolved in ammonium acetate buffer (0.1 M, pH 7.8) at a concentration of approximately 1 mg mL^{-1}. TCEP solution (0.5 M in water) was added and the reaction mixture incubated at 50 °C for one hour. After adding a ten-fold excess of maleimide solution (0.5 M in water) and incubating at 50 °C for another 30 minutes, the reaction mixture was purified by HPLC and the fractions lyophilised.

The Cys-reduced and -alkylated peptides were digested with endoproteinase GluC (EC 3.4.21.19, purchased from Sigma), cleaving peptide bonds almost specifically on the C-terminal side of Glu residues. The reduced and alkylated peptide sample was dissolved in ammonium acetate buffer (0.1 M, pH 7.7) at a concentration of approximately 10 mg mL^{-1} and incubated with reconstituted enzyme solution (10^{-4} g L^{-1}) at 37 °C for one to two hours. Progress of the reaction was monitored by mass spectrometry. The cleaved peptides were further purified by HPLC and subjected to sequencing by conventional Edman degradation or tandem mass spectrometry.

PCR analysis

Plant leaf material for RNA extraction was put into RNAlater solution (Ambion, Austin, TX) immediately after collection and RNA was isolated with the RNAqueous kit (Ambion, Austin, TX). Single stranded cDNA was prepared from RNA using either the OneStep RT-PCR kit (Qiagen, Hilden, Germany) or Superscript III (Invitrogen). Full length cDNAs were isolated first by using forward primers designed from the cyclotide sequences for mra3 and mra4 determined by Edman degradation (5′-CG ATC GAT TGY GGI GAR AGT TGY GT-3′) and oligo-dT (Proligo, Lismore, Australia). 5′ RACE using the FirstChoice® RLM-RACE kit (Ambion) was then performed using specific primers designed against the 3′ UTR and/or mature cyclotide sequences of the partial clones, namely mrl4o 5′-A AGA GCA CCC AAT AGC GCT AGT GA -3′, mrl4i 5′-C AAG AGC TTC AGG GGT GAT CAC AT-3′ and RB99 5′-TGG TAG CAA ACT TTG CCT TGG CAC-3′ for *Mra4* and *Mra13* respectively. *Mra18, Mra22, Mra23, Mra24, Mra25, Mra26* were isolated by 3′ RACE using anchor primer 5′-GAG CAA CGT CAC GAA AGA AGC GTT TTT TTT TTT TTT T-3′ and the aforementioned AAFALPA primer 5′-GGG CHG CHT TYG CHC TTC CHG

C-3′.[36] *Mra16* and *Mra17* were amplified with the AAFALPA primer and JM14 5′-AAT ACA AGG AAT ATA AAC ACA AGA TTC TCC RCA NGG DAT-3′ which targets sequences encoding the amino acid sequence PCGESCVYIPCIS. *Mra20* and *Mra21* were amplified by 3′ RACE using anchor primer and JM12 5′- TG GAA TTG CTG GAT GTT CTT GYA ARA AYA A-3′ which targets sequences encoding the amino acid sequence GIAGCSCKNK. *Mra28, Mra29* and *Mra30* were amplified with AAFALPA and JM40 5′-AGT TAA AAA GAG AGA TAG GAA GCA-3′. *Mra14* and *Mra19* were amplified with the AAFALPA primer and JM41 5′-TGG AAA GAC AGA TAT CTA GGA AGC A-3′. The resulting bands were excised from an agarose gel and cloned into pCR 2.1-TOPO vector using the TOPO Cloning kit (Invitrogen) or pGEM-T-easy (Promega) for sequencing.

GenBank accessions

The *Melicytus ramiflorus* cDNA sequences described in this work have been submitted to GenBank and possess the following accession numbers: *Mra4* EF103478; *Mra13* EF103479; *Mra14* EF103465; *Mra16* EF103466; *Mra17* EF103467; *Mra18* EF103469; *Mra19* EF103468; *Mra20* EF103463; *Mra21* EF103464; *Mra22* EF103472; *Mra23* EF103473; *Mra24* EF103471; *Mra25* EF103470; *Mra26* EF103474; *Mra28* EF103475; *Mra29* EF103476; *Mra30* EF103477. The corresponding protein accession numbers are: Mra4 ABO21629; Mra13 ABO21630; Mra14 ABO21616; Mra16 ABO21617; Mra17 ABO21618; Mra18 ABO21620; Mra19 ABO21619; Mra20 ABO21614; Mra21 ABO21615; Mra22 ABO21623; Mra23 ABO21624; Mra24 ABO21622; Mra25 ABO21621; Mra26 ABO21625; Mra28 ABO21626; Mra29 ABO21627; Mra30 ABO21628.

Abbreviations

CCK	cyclic cystine knot
HPLC	high performance liquid chromatography
RP	reverse phase
MS	mass spectrometry
LC-MS	liquid chromatography-mass spectrometry
DCM	dichloromethane
TCEP	triscarboxyethylphosphine
TFA	trifluoro acetic acid

Standard one and three letter codes are used for amino acids.

Acknowledgements

This work was supported in part by a grant from the Australian Research Council. DJC is an Australian Research Council (ARC) Professorial Fellow. JSM was funded by a University of Queensland Post-Doctoral Research Fellowship and is currently an ARC Queen Elizabeth II Fellow. We thank staff at Royal Botanic Gardens Melbourne, Australia, Auckland Regional Botanic Garden (Manurewa), Wellington Botanic Garden and Christchurch Botanic Garden (all New Zealand) for their help and cooperation. We are also indebted to Julie Dutton for her help in primer design, to Rekha Bharathi (University of Queensland) for technical assistance and to Marilyn Anderson (La Trobe University) for valuable discussions.

References

1 D. J. Craik, *Science*, 2006, **311**, 1563.
2 M. Trabi and D. J. Craik, *Trends Biochem. Sci.*, 2002, **27**, 132.
3 D. J. Craik, N. L. Daly, T. Bond and C. Waine, *J. Mol. Biol.*, 1999, **294**, 1327.
4 P. K. Pallaghy, K. J. Nielsen, D. J. Craik and R. S. Norton, *Protein Sci.*, 1994, **3**, 1833.
5 D. J. Craik, *Toxicon*, 2001, **39**, 1809.
6 M. L. Colgrave and D. J. Craik, *Biochemistry*, 2004, **43**, 5965.
7 N. L. Daly, K. R. Gustafson and D. J. Craik, *FEBS Lett.*, 2004, **574**, 69.
8 K. R. Gustafson, T. C. McKee and H. R. Bokesch, *Curr. Protein Pept. Sci.*, 2004, **5**, 331.
9 K. R. Gustafson, R. C. I. Sowder, L. E. Henderson, I. C. Parsons, Y. Kashman, J. H. I. Cardellina, J. B. McMahon, R. W. J. Buckheit, L. K. Pannell and M. R. Boyd, *J. Am. Chem. Soc.*, 1994, **116**, 9337.
10 K. R. Gustafson, L. K. Walton, R. C. I. Sowder, D. G. Johnson, L. K. Pannell, J. H. I. Cardellina and M. R. Boyd, *J. Nat. Prod.*, 2000, **63**, 176.
11 J. P. Tam, Y. A. Lu, J. L. Yang and K. W. Chiu, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 8913.
12 B. L. Barbeta, A. T. Marshall, A. D. Gillon, D. J. Craik and M. A. Anderson, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 1221.
13 C. Jennings, J. West, C. Waine, D. Craik and M. Anderson, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10614.
14 C. V. Jennings, K. J. Rosengren, N. L. Daly, M. Plan, J. Stevens, M. J. Scanlon, C. Waine, D. G. Norman, M. A. Anderson and D. J. Craik, *Biochemistry*, 2005, **44**, 851.
15 L. Gran, F. Sandberg and K. Sletten, *J. Ethnopharmacol.*, 2000, **70**, 197.
16 O. Saether, D. J. Craik, I. D. Campbell, K. Sletten, J. Juul and D. G. Norman, *Biochemistry*, 1995, **34**, 4147.
17 K. Sletten and L. Gran, *Medd. Nor. Farm. Selsk.*, 1973, **7–8**, 69.
18 K. M. Witherup, M. J. Bogusky, P. S. Anderson, H. Ramjit, R. W. Ransom, T. Wood and M. Sardana, *J. Nat. Prod.*, 1994, **57**, 1619.
19 H. R. Bokesch, L. K. Pannell, P. K. Cochran, R. C. Sowder, 2nd, T. C. McKee and M. R. Boyd, *J. Nat. Prod.*, 2001, **64**, 249.
20 T. Schöpke, M. I. Hasan Agha, R. Kraft, A. Otto and K. Hiller, *Sci. Pharm.*, 1993, **61**, 145.
21 P. Claeson, U. Göransson, S. Johansson, T. Luijendijk and L. Bohlin, *J. Nat. Prod.*, 1998, **61**, 77.
22 U. Göransson, A. M. Broussalis and P. Claeson, *Anal. Biochem.*, 2003, **318**, 107.
23 E. Svangard, U. Göransson, D. Smith, C. Verma, A. Backlund, L. Bohlin and P. Claeson, *Phytochemistry*, 2003, **64**, 135.
24 M. Trabi and D. J. Craik, *Plant Cell*, 2004, **16**, 2204.
25 A. M. Broussalis, U. Göransson, J. D. Coussio, G. Ferraro, V. Martino and P. Claeson, *Phytochemistry*, 2001, **58**, 47.
26 Y. F. Hallock, R. C. I. Sowder, L. K. Pannell, C. B. Hughes, D. G. Johnson, R. Gulakowski, J. H. I. Cardellina and M. R. Boyd, *J. Org. Chem.*, 2000, **65**, 124.
27 D. J. Craik, N. L. Daly, J. Mulvenna, M. R. Plan and M. Trabi, *Curr. Protein Pept. Sci.*, 2004, **5**, 297.
28 J. P. Mulvenna, C. Wang and D. J. Craik, *Nucleic Acids Res.*, 2006, **34**, D192.
29 J. F. Hernandez, J. Gagnon, L. Chiche, T. M. Nguyen, J. P. Andrieu, A. Heitz, T. Trinh Hong, T. T. Pham and D. Le Nguyen, *Biochemistry*, 2000, **39**, 5722.
30 J. L. Dutton, R. F. Renda, C. Waine, R. J. Clark, N. L. Daly, C. V. Jennings, M. A. Anderson and D. J. Craik, *J. Biol. Chem.*, 2004, **279**, 46858.
31 J. P. Mulvenna, L. Sando and D. J. Craik, *Structure (London)*, 2005, **13**, 691.
32 A. D. Gillon, I. Saska, C. V. Jennings, R. F. Guarino, D. J. Craik and M. A. Anderson, *Plant J.*, 2008, **53**, 505.
33 I. Saska, A. D. Gillon, N. Hatsugai, R. G. Dietzgen, I. Hara-Nishimura, M. A. Anderson and D. J. Craik, *J. Biol. Chem.*, 2007, **282**, 29721.
34 B. Chen, M. L. Colgrave, N. L. Daly, K. J. Rosengren, K. R. Gustafson and D. J. Craik, *J. Biol. Chem.*, 2005, **280**, 22395.
35 M. Trabi, E. Svangard, A. Herrmann, U. Göransson, P. Claeson, D. J. Craik and L. Bohlin, *J. Nat. Prod.*, 2004, **67**, 806.

36 S. M. Simonsen, L. Sando, D. C. Ireland, M. L. Colgrave, R. Bharathi, U. Goransson and D. J. Craik, *Plant Cell*, 2005, **17**, 3176.
37 D. C. Ireland, M. L. Colgrave, P. Nguyencong, N. L. Daly and D. J. Craik, *J. Mol. Biol.*, 2006, **357**, 1522.
38 J. Zhang, B. Liao, D. J. Craik, J.-T. Li, M. Hu and W.-S. Shu, *Gene*, 2009, **431**, 23.
39 K. J. Rosengren, N. L. Daly, M. R. Plan, C. Waine and D. J. Craik, *J. Biol. Chem.*, 2003, **278**, 8606.
40 M. R. R. Plan, U. Göransson, R. J. Clark, N. L. Daly, M. L. Colgrave and D. J. Craik, *ChemBioChem*, 2007, **8**, 1001.
41 N. L. Daly, A. Koltay, K. Gustafson, R. M. R. Boyd, J. R. Casas-Finet and D. J. Craik, *J. Mol. Biol.*, 1999, **285**, 333.
42 A. Herrmann, R. Burman, J. S. Mylne, G. Karlsson, J. Gullbo, D. J. Craik, R. J. Clark and U. Göransson, *Phytochemistry*, 2008, **69**, 939.
43 H. E. J. Ballard, K. J. Sytsma and R. R. Kowal, *Syst. Bot.*, 1998, **23**, 439.
44 N. L. Daly, R. J. Clark, M. R. Plan and D. J. Craik, *Biochem. J.*, 2006, **393**, 619.
45 U. Göransson, T. Luijendijk, S. Johansson, L. Bohlin and P. Claeson, *J. Nat. Prod.*, 1999, **62**, 283.
46 D. C. Ireland, M. L. Colgrave and D. J. Craik, *Biochem. J.*, 2006, **400**, 1.
47 B. Chen, M. L. Colgrave, C. Wang and D. J. Craik, *J. Nat. Prod.*, 2006, **69**, 23.

