Discovery and Characterization of Novel Cyclotides Originated from Chimeric Precursors Consisting of Albumin-1 Chain a and Cyclotide Domains in the Fabaceae Family

Giang Kien Truc Nguyen1, Sen Zhang1, Ngan Thi Kim Nguyen1, Phuong Quoc Thuc Nguyen1, Ming Sheau Chiu1, Antony Hardjojo1 and James P Tam1*

1From School of Biological Sciences, Nanyang Technological University, Singapore 637551

Running Head: Chimeric Precursors Consisting of Cyclotide and A1a domains

* Address correspondence to: James P Tam. Fax: 65- 6515 1632; Email: jptam@ntu.edu.sg

The tropical plant Clitoria ternatea is a member of the Fabaceae family well-known for its medicinal values. Heat extraction of C. ternatea revealed that the bioactive fractions contained heat-stable cysteine-rich peptides (CRPs). The CRP family of A1b (Albumin-1 chain b/leginsulins) which is a linear cystine-knot CRP has been shown to present abundantly in the Fabaceae. In contrast, the cyclotide family which also belongs to the cystine-knot CRPs but with a cyclic structure is commonly found in the Rubiaceae, Violaceae, and Cucurbitaceae families. In this study, we report the discovery of a panel of 15 heat-stable CRPs, of which 12 sequences (cliotide T1-T12) are novel. We show unambiguously that the cliotides are cyclotides and not A1bs, as determined by their sequence homology, disulfide connectivity, and membrane-active properties indicated by their antimicrobial activities against E. coli and cytotoxicities to HeLa cells. We also show that cliotides are prevalent in C. ternatea and are found in every plant tissue examined, including flowers, seeds, and nodules. In addition, we demonstrate that their precursors are chimeras, half from cyclotide and the other half from Albumin-1, with the cyclotide domain displacing the A1b domain in the precursor. Their chimeric structures likely originate from either horizontal gene transfer or convergent evolution in plant nuclear genomes which are exceedingly rare events. Such atypical genetic arrangement also implies a different mechanism of biosynthetic processing of cyclotides in the Fabaceae and provides new understanding of their evolution in plants.

Clitoria ternatea is a perennial climber well-known for its butterfly-shaped vivid-blue flowers. It belongs to the Fabaceae family and displays a broad-spectrum of medicinal usages in almost every plant part (1). It is often cited in the Ayurvedic system of Indian medicines as an effective antidepressant, antimicrobial, antipyretic and nerve tonic for enhancing memory (1). It is also used as an alternative medicine in America and other tropical Asian countries (1). In Cuba, decoctions of roots and flowers are reported to have emmenagogue properties which promote menstruation and uterine contraction. Studies on animals have shown that the aqueous extracts of the flowers and leaves display antihyperglycemic effects in rats (2). Decoctions of roots and leaves elicit a wide spectrum of activities on the central nervous system and have been shown to enhance acetylcholine content in rat hippocampus (3-5).

Preliminary phytochemical screenings of C. ternatea extracts have shown that the biologically active fractions are rich in peptides and proteins while showing negative tests for alkaloids, saponins, flavonoids, coumarins, and lignans (3, 6). In these studies, plant extracts were prepared by boiling the pulverized plants with hot water. Although the exact chemical components have not been identified, it is plausible to speculate that the active principles are heat-stable proteins. The view of proteins as viable bioactive herbal components, however, is conceptually contradictory to our current knowledge of traditional medicines which have been biased toward small molecules with molecular weights less than 500 Da. Peptides and proteins whose molecular weights are considerably larger than 500 Da have generally not been considered as active principles with the common perception that they are unstable and unavailable as a source of active principles in decoctions. This bias is partly attributed to the intrinsic instability of peptides and proteins against heat during decoction preparation or their susceptibility to enzymatic and acidic hydrolysis during digestion. However, recent literature precedents suggest otherwise.

Cumulative evidence shows that several classes of cysteine-rich peptides (CRPs) in
plants such as defensins, A1bs (also known as leginsulins), knottins, and cyclotides are heat-stable (7-11). Although their primary sequences, biochemical properties and functions may differ greatly, CRPs possess multiple disulfide bridges that cross brace their structures, often conferring thermal, chemical, and enzymatic stability (8, 12). Of these, A1bs are well documented CRPs characteristic of the Fabaceae family and have been identified in several legume species (13, 14). They consist of 35 to 40 amino acid residues in length and contain three disulfide bridges (13). A1bs are highly stable and able to survive the acids and digestive enzymes in the porcine stomach and intestine (15, 16). They have been shown to possess several biological activities such as insecticidal and hormonal functions in plants (17, 18). They also affect mammalian physiological functions such as regulation of glucose metabolism in mice (19).

Cyclotides are another class of CRPs which has recently gained interest owing to their extraordinary stability and diverse biological functions (20, 21). They are macrocyclic peptides from plants of Rubiaceae, Violaceae, and Cucurbitaceae families (22, 23). They contain 28-37 residues and are structurally distinguished from the conventional linear CRPs such as A1bs by being cyclic. They possess a circular peptide backbone fortified by a cystine-knot motif (24, 25). Such structural elements produce their stability to heat, chemical, and enzymatic degradation (26). These advantages enhance the potential of cyclotides as active principles in traditional medicines.

Very recently, Poth et al. reported the isolation of 12 novel cyclotides from the seeds of *C. ternatea*, which marked the first discovery of cyclotides in the Fabaceae family (27). Their finding prompts us to report our independent work on the discovery of cyclotides in this new plant family. Here, we describe and report the isolation and characterization of novel cyclotides as heat-stable CRPs in every plant part of *C. ternatea* including leaves, stems, roots, shoots, nodules, flowers, pods, and seeds. Fifteen cyclotides were isolated, of which 12 sequences are novel and designated as cliotide T1-T12. Our results show unequivocally that cliotides may be the active ingredients responsible for the indications of *C. ternatea* in traditional medicines as anti-infectives and anti-tumorigenic agents (28). Interestingly, molecular cloning of novel cliotide genes revealed that they are derived from an entirely new genetic arrangement differing from the known cyclotides of the Rubiaceae and Violaceae families. Cliotides, at the gene level, have chimeric structures containing genetic elements from both Albumin-1 (A1) of the Fabaceae and cyclotides of the Rubiaceae and Violaceae families. This unusual arrangement suggests a novel mode of biosynthetic processing of cyclotides in the Fabaceae family and provides hints about their evolution in plants.

**Experimental Procedures**

**Screening for Heat-stable Biologics in *C. ternatea***- Fresh flowers (~50 mg) were macerated with 500 µL of water and incubated at 100 °C for 30 min. The heat-stable fraction was separated from the denatured protein precipitates and other plant debris by centrifugation. For comparison, a control experiment was performed without the heat treatment. The aqueous extracts with or without heating were monitored by MS and HPLC.

**Peptides Isolation and Purification**- Fresh plant materials (~1 kg, whole plant) were collected and homogenized with 5 L of water and incubated at 100 °C for 1 hr. After filtration, the aqueous extract was subjected to a flash chromatography column packed with 150 g of C18 material (Grace Davison). The column was washed with 20% ethanol and eluted with 80% ethanol to obtain the CRP-enriched fraction. This fraction was concentrated and subjected to preparative HPLC on Shimadzu system with a C18 Vydac column (250 x 21 mm) at a flow rate of 8 mL/min. A linear gradient of 1% min⁻¹ of 0-80% buffer B (100% acetonitrile, 0.05% trifluoroacetic acid) was applied and the eluants were monitored by UV detection at 220, 254, and 280 nm. The fractions obtained were re-purified by semi-preparative HPLC with a C18 Vydac column (250 x 10 mm) at a flow rate of 3 mL/min with the same gradient. The approximate yields obtained for cliotide T1, T2, T3, T4, T7, T8, T9, T10, and Cter B were 5, 10, 70, 50, 80, 80, 80, and 70 mg respectively.
S-reduction and S-alkylation- Around 20 µg of each peptide was dissolved in 100 µL of NH₄HCO₃ buffer (100 mM, pH 7.8) containing 10 mM dithiothreitol (DTT), and incubated for 1 hr at 37 °C. Two-fold excess of iodoacetamide (IAA) over the total thiol was added and incubated for 1 hr at 37 °C. S-alkylated peptides were purified by HPLC.

Enzymatic Digestion and Sequence Determination- Lyophilized S-alkylated peptides were dissolved in 5 µL NH₄HCO₃ buffer (20 mM, pH 7.8) and incubated with endoproteinase Glu-C (EndoGlu-C), trypsin or chymotrypsin (Roche) at a final peptide-to-enzyme ratio of 50:1. The digestions were allowed to proceed for 1 hr. Peptide fragments resulting from the digestions were examined by MALDI-MS followed by MALDI-TOF/TOF mass spectrometer. Primary peptide sequences were obtained by interpreting the b- and y-ions formed during the MS/MS fragmentation. Assignments of isobaric residues Ile/Leu and Lys/Gln were based on enzymatic digestion patterns, quantitative amino acid analysis, cDNA sequences, and homology to known cyclotides. Amino acid analysis was performed for five cliotides without complete gene sequences including cliotide T1, T5, T10, T11, and Cter D (Table S2-S6). Due to sample limitation, the Ile/Leu assignment of cliotide T6 was based solely on chymotryptic digestion and sequence homology.

Disulfide Mapping- Cliotide T2 (0.2 mg) was partially reduced in 500 µL of 100 mM citrate buffer pH 3.0, 20 mM TCEP at 37 °C for 35 min. Subsequently, NEM (N-ethylmaleimide) powder was added directly to a final concentration of 50 mM and incubated at 37 °C for another 15 min. The reaction was quenched by immediate injection of samples into a Vydac C18 column (250 x 4.6 mm) at a flow rate of 1 mL/min. Intermediate species were separated with a linear gradient of 0.3% min⁻¹ of 10-60% buffer B and analyzed with MALDI-TOF MS to verify the number of NEM-alkylated cysteines. NEM-alkylated intermediate species were then fully reduced with 20 mM DTT, incubated at 37 °C for 60 min. The reduced peptides were S-alkylated with 40 mM IAA, incubated at 37 °C for 30 min before stopping the reaction by injection into HPLC. S-alkylated peptides were lyophilized and re-dissolved in 20 µL of 10% ACN, 20 mM NH₄HCO₃, pH 7.8 for trypsin digestion. Fragments obtained were sequenced by MS/MS.

Antibacterial Assay- Five bacterial strains from the American Type Culture Collection (ATCC) were used. Gram-positive bacteria included Staphylococcus Aureus ATCC12600 and Enterococcus faecalis ATCC 47077. Gram-negative bacteria included Escherichia coli ATCC 700926, Pseudomonas aeruginosa ATCC 39018, and Klebsiella pneumonia ATCC 13883. All the strains were cultured in trypticase soy broth (TSB). The antimicrobial activities of cliotide T1-4 were examined by using the radial diffusion assay described by Lehrer et al. (29). D4R, a synthetic antimicrobial peptide, was used as the positive control (30).

Cytotoxicity Assay- HeLa cells were diluted in Dulbecco's modified Eagle's medium to a density of 20,000 cells per mL. 100 µL of the cell suspension was added into each well of a 96-well plate (Nunc) and incubated for 24 hr. Cyclotides were serially diluted in phosphate buffered saline (PBS), and 5 µL was dispensed into each well. After incubation for 72 hr, 20 µL of 5 mg/mL Thiazolyl Blue Tetrazolium Bromide (MTT) solution was added to each well. The plate was incubated for 3.5 hr at 37 °C. The media was removed and 150 µL of dimethyl sulfoxide was added for cell lysis. The absorbance at 590 nm was measured with a reference filter at 620 nm. Survival Index (SI) was calculated as percentage of growth inhibition in comparison with the control wells (added PBS only). Cyclotides’ cytotoxicites were indicated by IC₅₀ values (concentration that gives an SI of 50%).

Hemolytic Assay- Blood type A was taken from a healthy volunteer. Red blood cells (RBCs) were washed three times with PBS and resuspended in PBS to give a final 1% suspension. 95 µL of the suspension was added to each well of a 96-well plate. Cyclotides were serially diluted in PBS, and 5 µL was dispensed into each well. Each cyclotide concentration was diluted in PBS, and 5 µL was dispensed into each well. Each cyclotide concentration was tested in triplicate. The plate was incubated at 37 °C for 1 hr and centrifuged at 1,000 rpm for 6 min. Aliquots of 60 µL were transferred to a new 96-well plate. The absorbance was measured at 415 nm. The level of hemolysis was calculated as the percentage of maximum lysis (1% Triton X-100 control) after adjusting for minimum lysis.
from A1b/leginsulin sequences used included four of A1bs/leginsulins, cyclotides, and cliotides. 

**Phylogenetic Tree Construction-** Total RNA was extracted from fresh plant materials (leaves, flowers, roots) using PureLink™ RNA Purification System (Invitrogen). cDNA libraries were prepared from total RNA using Smarter™ RACE cDNA Amplification Kit (Takara). First, 3' RACE PCR was conducted using degenerate primers designed against cliotide T2 (5'-GGGAAAGTTGT-3' encoding IPCGESC). The PCR products were cloned into pGEM-T Easy vector (Promega) and sent for sequencing. The 5'-end partial gene was obtained from the same procedure using the UPM primer from the kit and a specific primer against the 3' untranslated region of the newly found gene. The full transcript was then assembled from the two partial genes. This strategy was also applied to other cliotides. In addition, we also used the specific primers against loop 1 of some newly obtained sequences in 3' RACE PCR to target random cyclotides, i.e. 5'-TGCGGCGAGAGTTGC-3' encoding CGESC and 5'-ATCCCATGTG-3'

**Cloning of Cliotide Genes-** Total RNA was extracted from fresh plant materials (leaves, flowers, roots) using PureLink™ RNA Purification System (Invitrogen). cDNA libraries were prepared from total RNA using Smarter™ RACE cDNA Amplification Kit (Takara). First, 3' RACE PCR was conducted using degenerate primers designed against cliotide T2 (5'-GGGAAAGTTGT-3' encoding IPCGESC). The PCR products were cloned into pGEM-T Easy vector (Promega) and sent for sequencing. The 5'-end partial gene was obtained from the same procedure using the UPM primer from the kit and a specific primer against the 3' untranslated region of the newly found gene. The full transcript was then assembled from the two partial genes. This strategy was also applied to other cliotides. In addition, we also used the specific primers against loop 1 of some newly obtained sequences in 3' RACE PCR to target random cyclotides, i.e. 5'-TGCGGCGAGAGTTGC-3' encoding CGESC and 5'-ATCCCATGTG-3'.

To identify intron locations, genomic DNA from *C. ternatea* flowers was extracted. PCR reactions on DNA templates were then conducted with specific primers designed against 5' and 3' untranslated regions of each cliotide. PCR products were purified and subjected to the procedure described above to obtain the DNA gene sequences.

**Phylogenetic Tree Construction-** Phylogenetic tree was constructed using translated precursors of A1bs/leginsulins, cyclotides, and cliotides. A1b/leginsulin sequences used included four from *Pisum sativum* (GenBank: CAB82859, CAE00465, CAE00467, CAE00468), two from *Medicago truncatula* (AC233561, CAE00461, CAE00462, ACJ85909), one from *Glycine max* (BAA04219), and one from *Glycine soja* (CA09880). Cyclotide sequences included two from *Viola baoshanensis* (DQ851860, ABI48957), two from *Viola odorata* (P58440, AY630565), one from *Glloeospermum blakeanum* (GQ438777), and one from *Hedyotis centranthoides* (FJ211188). Cliotide precursors of cliotide T1, T2, T3, T7, and T11 were used. The precursor protein sequences were aligned using COBALT multiple alignment (31). Phylogenetic tree was generated using fast minimum evolution method and visualized by Tree View software (32).

**RESULTS**

**Screening of Heat-stable Biologics in *C. ternatea* Flowers.** Heat-stable CRPs were extracted from *C. ternatea* flowers by incubation with boiling water at 100 °C for 30 min. The aqueous soluble fraction was separated from the denatured protein precipitates and other plant debris by centrifugation. A control experiment without the heat treatment was performed in parallel. The aqueous extracts were subjected to MALDI-TOF MS to detect the presence of putative biologics within the 1 to 10 kDa mass range. The mass spectrum revealed a group of peptides with strong m/z intensity at around 3 kDa (Figure 1A). Comparison of the RP-HPLC profiles before and after the heat treatment revealed that >80% of these peptides survived the boiling water incubation (Figure 1B, 1C). Their disulfide contents were subsequently analyzed by S-reduction with DTT and S-alkylation with IAA followed by MS analysis. The number of disulfide bonds was deduced by comparing the mass difference before and after reductive S-alkylation. Each S-alkylated half-cystine residue caused a mass shift of 58 Da. Most of these compounds displayed a mass shift of 348 Da suggesting the presence of three disulfide bridges, a common structural feature found in many plant biologics.

**Peptide Isolation and Sequencing: Cyclotides as Heat-stable Biologics in *C. ternatea*.** To isolate sufficient CRP samples for *de novo* sequencing and biological activities studies, 1 kg of fresh plant materials (whole plant) were collected and extracted with boiling water (100 °C). Using RP-HPLC, 15 heat-stable CRPs were isolated.

To determine their primary sequences, each HPLC-purified peptide was SS-reduced, S-carbamidomethylated, and digested with trypsin, chymotrypsin or EndoGlu-C. Generated fragments were sequenced by tandem mass spectrometry. As an example, MS/MS sequencing of cliotide T1 is shown in Figure 2. Cliotide T1 and its S-alkylated form had m/z of 3084 and 3432 Da, respectively. Enzymatic digestion of S-alkylated cliotide T1 by EndoGlu-C gave a single fragment with a mass increase of 18 Da suggesting the linearization of a cyclic amide backbone by hydrolysis. Its circular structure was confirmed by tryptic and...
chymotryptic digestions which yielded overlapping fragments in their entirety. Sequence analysis of cliotide T1 revealed 84% identity to circulin A, a prototypic cyclotide, further suggesting that it is a novel member of cyclotide family. Using this approach, we successfully elucidated the primary structures of 15 peptides, which all belong to the cyclotide family (Table 1). Twelve sequences (cliotide T1-T12) are novel, and three sequences (Cter A, Cter B, and Cter D) have been identified previously (27).

cDNA Cloning Revealed Chimeric Structures of Cliotide Encoding Genes. To determine their encoding genes and to provide insight about their biosynthesis, we designed degenerate primers for 3′ RACE PCR based on MS/MS determined sequences. A total of 11 partial clones were obtained. Afterward, a series of specific primers based on the newly obtained genetic sequences were used for 5′ RACE PCR to complete the full-length genes. Nine clones were named ctc1, ctc2, ctc3, ctc4, ctc5, ctc7, ctc8, ctc9, and ctc12 according to their cliotide inserts. Two clones encoding Cter A and Cter B were named ctc13 and ctc14, respectively.

The translated sequences of cliotide clones are shown in Figure 3. They feature an entirely different arrangement from previously published cyclotide precursors. In the Rubiaceae and Violaceae families, cyclotide genes such as oak1 and voc1 (33) contain an endoplasmic reticulum (ER) signal sequence, an N-terminal pro-domain (NTPP), an N-terminal repeat region (NTR), a cyclotide domain, and a short C-terminal tail. Sequence analysis revealed that cliotide precursors contained no NTPP or NTR regions. Instead, the ER signal sequence was followed directly by the cyclotide domain. In addition, a novel cysteine-rich domain was found at the C-terminal end of the precursor proteins separated from the cyclotide domain by a short putative linker region. A BLAST search showed that this novel domain was highly homologous to A1a domain (Albumin-1 chain a) of A1 genes found in many legume species (11, 34). Surprisingly, the gene architecture of cliotides displayed striking similarity to those of A1 genes (Figure 4) which contain a signal peptide followed by an A1b domain, a linker peptide, and an A1a domain (35). The A1b domain in pea seeds encodes for a 37-residue insecticidal peptide containing three disulfide bonds (11). The cliotide genes shared exactly the same arrangement with A1 genes except that the A1b domain was replaced by the cyclotide domain. The homology of cliotide genes to both A1 genes (Fabaceae) and cyclotide genes (Rubiaceae, Violaceae) makes them naturally occurring chimeric genes.

Cliotide Genes Contain a Single Intron at the ER Region. Known families of CRP genes are generally characterized by the presence of a single intron located within the ER domain (36-39). To compare the genetic structure of cliotides at both DNA and mRNA level, five cliotide precursor genes including cliotide T2, T3, T7, T8, and T10 were cloned from the leaf DNA using cDNA-derived sequences as primers. The DNA clones revealed a single intron located in the signal peptide region of the cliotide genes, which is similar to A1 genes from the Fabaceae and cyclotide genes from the Rubiaceae family (Figure 4). Thus, the single-intron architecture of cliotide DNA genes is consistent with genes from CRP families which include not only cyclotides and A1s but also plant defensins and thionins (36, 40). It is interesting to note that the cliotide T7 gene contains no intron, which is similar to those characterized from the Violaceae family. This suggests an intron loss event in some of the cliotide genes, possibly mediated through gene conversion with a reverse transcriptase product of a spliced mRNA.

Disulfide Mapping of Cliotide T2: New Precursor, Same connectivity. The disulfide connectivity of novel cliotides was sought to determine if they possessed a different arrangement from the cystine-knot motif found in cyclotides. Cliotide T2, present in high abundance in the floral tissue, was selected as a representative. The connectivity of cliotide T2 was determined by a differential S-tagging strategy through a sequential S-reduction and S-alkylation (41). The SS bonds of native cliotide T2 were first partially ruptured with TCEP to generate a series of isoforms with one or two disulfide bonds being reduced. The released thiols were immediately S-alkylated with an excess of NEM. The whole process was performed under acidic conditions at pH 3.5 to avoid scrambling of disulfide linkages. The partially S-reduced and S-alkylated peptides were purified by RP-HPLC.
Antimicrobial Activity. To assess their antimicrobial activity, four cliotides (cliotide T1-4) were tested against a panel of five different bacterial strains using a radial diffusion assay. Their selection was primarily based on abundance and varied structural characteristics, consisting of two Möbius and two bracelet cyclotides. Table 2 summarizes their antimicrobial activities. D4R, a synthetic antimicrobial peptide, was used as the positive control (30). The bracelet subgroup including cliotide T1 and cliotide T4 showed strong antimicrobial activity against three species of Gram-negative bacteria with the MIC values in the low µM range. Among these strains, E. coli was the most susceptible to cliotides’ bactericidal activities. Cliotide T1 and T4, however, were relatively ineffective against both tested Gram-positive bacterial strains S. aureus and E. faecalis. The Möbius subgroup comprising of cliotide T2 and cliotide T3 were inactive against all five tested bacterial strains at concentration up to 100 µM.

Hemolytic Activity. Hemolytic effects of cliotide T1-4 were conducted using human type A erythrocytes. Melittin, a known peptide from bee venom with strong hemolytic activity, was used as positive control. Cliotide T2 was relatively non-hemolytic and only lysed approximately 18% of the RBCs at 40 µM. The HD₅₀ of the other three cliotides ranged from 7.1 to 13.1 µM (Table 2), which were significantly less potent than melittin (3.2 µM).

Cytotoxicity against HeLa Cells. The cytotoxic properties of cliotide T1-4 was assessed by MTT assay, which monitored mitochondrial activity in living cells and hence indicated cell viability. HeLa cells were selected as a model for the assay. Five concentrations were tested for each cliotide and IC₅₀ values were determined from the survival curves. All of the four tested cliotides displayed sharp dose-response curves with IC₅₀ values ranging from 0.6 to 8.0 µM (Table 2). Cliotide T1 and T4 were the most potent with IC₅₀ values of 0.6 µM. Cliotide T3 had moderate cytotoxic effect with IC₅₀ at 2.0 µM and cliotide T2 was the least cytotoxic with IC₅₀ at 8.0 µM.

Tissue-specific Distribution. To determine if the novel cliotides are expressed universally or in a tissue-specific manner, MALDI-TOF MS was utilized to profile their expression in different plant parts. A total of eight different tissues were
collected including leaves, stems, roots, flowers, pods, seeds, nodules, and shoots. They were extracted separately using boiling water and analyzed by mass spectrometry. The mass spectra revealed that each tissue had a unique profile distinguished by the number and amount of expressed cliotides (Figure 6). Each tissue produced at least 10 different cliotides, but only a few specific cliotides dominated the expression profiles. For instance, cliotide T2 and cliotide T3 were highly expressed in the flowers and pods, accounting for more than 70% of total cyclotides in these tissues (Table S6). Similarly, cliotide T8, cliotide T9, and cliotide T10 were the major constituents in seeds. However, not all cliotides were tissue-specific. Certain cliotides such as cliotide T1 and cliotide T4 were found to be expressed in all examined tissues, suggesting their house-keeping role in plant defense.

DISCUSSION

In this study, we report the discovery of cyclotides as heat-stable biologics in *Clitoria ternatea*. Fifteen cyclotides have been characterized from plant extracts, of which 12 sequences are novel. Three cliotide sequences are identical to those isolated from the seeds of *C. ternatea* previously reported by Poth *et al.* (27) who identified a suite of 12 novel cyclotides. The lack of sequence overlaps could be attributed to several factors including methods of extraction, tissues examined, and geographical variations. Here, we show that cyclotides are not only present in seeds but in every plant part, including flowers, leaves, roots, nodules, pods, stems, shoots, and seeds. The presence of cyclotides in nodules is noteworthy because of their high abundance and possible roles in controlling the symbiosis of their host plant and nitrogen-fixing bacteria.

The identification of cliotides as cyclotides and not A1bs is unequivocal. A1bs previously discovered in the legumes are CRPs of similar sizes as cyclotides, but contain a linear structure. Cyclotides have a cyclic structure with a head-to-tail cyclized peptide backbone. Using a combination of enzymatic fragmentations and MS sequencing, all 15 cliotides are shown to be cyclic proteins. Disulfide mapping of cliotide T2 confirms that it retains a cystine-knot motif similar to cyclotides. In addition, sequence and cysteine-spacing comparisons also firmly place cliotides in the cyclotide family (Table 1).

Disulfide mapping of cliotide T2 provides useful information about its unfolding mechanism based on the intermediate species isolated. The abundance of 2SS species formed by the breaking of the Cys II-V bond suggests that this cystine linkage is most susceptible to reducing reagents. 1SS species was subsequently generated by reducing of Cys I-IV bond with the remaining of Cys III-VI bond as the most stable disulfide linkage. This is consistent with our understanding of cyclotides’ cystine knot structure that the penetrating disulfide bond of cystine-knot motif (Cys III-VI) is shielded and buried deep inside the cystine-core to account for its resistance against reducing reagents (42).

Functionally, cyclotides are characterized by their membrane-active properties for a wide variety of biological actions (23, 43). Among these, antimicrobial, hemolytic, and cytotoxic activities have been studied to dissect the membrane binding ability of cyclotides on three different membrane types of bacteria, erythrocytes, and cancer cells (23). These bioactivities were thus selected for functional evaluations of novel cliotides. Our results show that cliotides retain the characteristic functions of cyclotides with similar potency. Clidotides T1 and T4 are active against *E. coli* with MIC value of approximately 1 μM as compared to 1.5 μM for cyclopsychotride (21). They are cytotoxic to HeLa cells with an IC₅₀ within the 0.6-8.0 μM range, which is comparable to the cytotoxic effects of several cyclotides such as varv A (IC₅₀ 2.7-6.3 μM), varv F (IC₅₀ 2.6-7.4 μM), and cycloviolacin O2 (IC₅₀ 0.1-0.3 μM) (44). They also have similar hemolytic activities (HD₅₀ 7.9-13.4 μM) as compared to kalata B1 (11.7 μM) (45).

Extraction of cyclotides in our study was achieved by using boiling water, demonstrating their heat stability. Although numerous extraction techniques, often involving organic solvents, have been used to isolate active principles, decoction preparation by soaking the herbs in boiling water is the most relevant method in traditional medicines. Heating of the plant extract at 100 °C for 30 min showed nearly identical cliotide profiles before and after the treatments. Our results also demonstrate the thermal stability of cyclotides not only as pure compounds dissolved in water (26) but also in
the complex mixture of the plant extract which contains many herbal-derived, small molecules. Thus, our findings provide experimental support that cyclotides can survive heat treatment during the decoction preparation process and may be one of the active principles attributed to the medicinal values of *C. ternatea*.

Similar to many CRPs, the cliotide genes contain a single intron in the ER domain (40). Intriguingly, our results show unexpectedly that cliotide precursors display an entirely new arrangement differing from the known cyclotides from the Rubiaceae and Violaceae families. Cliotide genes contain a hybrid coding region consisting of peptide domains from two unrelated genes. In most legume species, the A1 genes encode three major domains: ER, A1b, and A1a (11). Cliotide genes share essentially the same arrangement except for the exchange of the A1b for the cyclotide domain. This raises the question whether A1b peptides are still expressed in *C. ternatea* or they have been totally replaced by cyclotides. Our attempts to identify A1b peptides in the *C. ternatea* extract have failed to identify similar peaks in the known mass range of A1bs. It is possible that they are present in low abundance that escapes our detection.

Characterization of cliotide genes provides new understanding about the biosynthetic mechanism of cyclotides in the Fabaceae family. Clitotide precursors, unlike their cyclotide counterparts, are devoid of the NTTP and NTR regions typically separating the ER signal from the mature cyclotide domain. This suggests that cliotides are N-terminally processed by signal peptidases (SPase I) (Figure 7). The absence of the NTTP and NTR regions also implies that these protein domains may not be essential for the biosynthesis of cyclotides in the Fabaceae family. They instead may have other endogenous functions in plants. At the C-terminus, the cyclotide domains are probably cleaved by AEP (asparaginyl endopeptidase) as in Rubiaceae and Violaceae species indicated by the presence of the highly conserved Asn at the processing site. AEP is a cysteine proteinase and forms with the C-terminal Asn a reactive thioester bond that leads a head-to-tail ligation to afford the cyclized structure of a cyclotide. Interestingly, the conserved Leu residue located at two residues downstream of the C-terminal Asn processing site is replaced by Val or Ile in cliotide genes. This reflects a difference in specificity at the P2' position of AEP in *C. ternatea*.

In addition to providing new understanding about the biosynthetic mechanism of cyclotides, the cliotide genes also support the natural occurrence of cliotide T1 and T12. These two peptides share identical primary sequences, except for the single amino acid difference at the ligation site, the Asn and Asp variation in loop 6. Five similar cyclotide pairs with such a variation have been identified in the previous study by Poth et al. from the extract of *C. ternatea* seeds (27). The origin of the Asn- and Asp-cyclotide variants has been a topic of debate. It is uncertain whether the Asp-variants are gene-encoded or formed as the result of the conversion of the Asn to Asp due to deamidation. The isolation of both cliotide T1 and T12 encoding genes (ctc1 and ctc12) in this study provides firm evidence that the Asp-variants in *C. ternatea* are gene-encoded. However, it remains to be determined whether the same molecular mechanism is responsible for the Asn and Asp variation observed in other species such as *O. affinis* and *V. odorata* (46, 47).

The genetic replacement of A1bs by cyclotides raises another question about the functional overlap between these two peptide families. The peptides are similar in size, both possessing a cystine-knot scaffold and exhibiting insecticidal activity (48). A1bs are known to exert physiological functions in plants such as growth, differentiation and cell proliferation by binding to a 43-kDa receptor located at the plant cell wall (34, 49, 50). In contrast, little is known about the physiological functions, intracellular location or receptor-binding capabilities of cyclotides in plants. The substitution of A1bs for cyclotides in *C. ternatea* suggests that they may share certain roles and provides hints about the physiological functions of cyclotides.

How does the chimeric structure of cliotide genes occur? The cyclotide-encoding domains in *C. ternatea* may arise by vertical heredity, horizontal gene transfer or convergent evolution. In vertical heredity, cyclotide genes would have to exist in the ancestral legume lineages and are inherited in *C. ternatea* by means of sexual reproduction. This hypothesis, however, cannot explain why only A1bs are detected in all the Fabaceae hitherto examined (11).

In a horizontal gene transfer (HGT) event,
C. ternatea acquires the cyclotide domains horizontally from other species. Such a scenario is supported by three lines of evidence. First, there is an unexpected clustering of C. ternatea with several Rubiaceae and Violaceae species instead of Fabaceae species using phylogenetic tree constructed with the cyclotide, A1, and cliotide genes (Figure 8). It should be noted that three different gene types were used for constructing the phylogenetic tree because of the chimeric nature of the cliotide precursors. Second, a DNA database search of 22 Fabaceae species, including two plants with full genome sequences Glycine max and Medicago truncatula, identified orthologs of A1 genes but none of them contains the cyclotide domain. Finally, our preliminary screening of two related Clitoria species, C. fairchildiana and C. laurifolia, did not detect the presence of cyclotides, but peaks at 3.7 kDa resembling A1b peptides were observed (data not shown). It is thus possible that C. ternatea may have received the cyclotide genes from donor species of the Rubiaceae or Violaceae families. If this hypothesis is shown to be valid, it will provide important evidence of HGTs between plant nuclear genomes. HGTs are known to occur mostly in plant mitochondria but rarely in nuclear and chloroplast DNA (51).

Intriguingly, the displacement of A1bs for cyclotide domains in clotide genes appears to interrupt their normal functions, an event that often leads to negative selection (52, 53). However, in this instance, the cyclotide domains were not only transferred but also correctly processed, folded and fully functional. Furthermore, they also display differential distribution in different plant parts, which suggests that they have been positively selected and fine-tuned for tissue-specific functions.

The third possibility is convergent evolution from A1bs to cyclotides. Both are similar in size, share a cystine knotted structure, and overlap in certain functions. However, for this to happen, it would require massive mutations and substitutions for complete conversion from A1bs to cyclotides. An interesting example of convergent evolution has recently been reported for the platypus (54). The toxin genes identified from this animal show striking homology to those found in other venomous species such as snakes, starfish and sea anemones. Genetic analysis reveals that these venoms have evolved independently from different origins but ended up developing similar families of molecules (55).

Genetic characterization of cyclotides in the Fabaceae may open up new avenues of research allowing genetic manipulation and production of cyclotide-based therapeutics in agriculturally important Fabaceae plants. There has been an increasing interest in using cyclotides as molecular scaffolds for grafting of bioactive epitopes. The current bottlenecks for chemical synthesis of these engineering peptides are high production cost and uncertainty in obtaining the correct folding. The use of plants as bioreactors can provide a cost-efficient solution and rapid production of cyclotides for medical and industrial applications.

In conclusion, our data provide the first genetic characterization of cyclotides in the Fabaceae. We show that their precursors are chimeras, half from cyclotide (Rubiaceae and Violaceae) and the other half from Albumin-1 (Fabaceae), with the cyclotide domain replacing the A1b domain in the precursor. Their chimeric structures suggest that the occurrence of cyclotides in the Fabaceae could be the result of a horizontal gene transfer between plant nuclear genomes or a convergent evolution from A1bs to cyclotides. Collectively, our findings provide new understanding about the biosynthetic mechanism of cyclotides in the Fabaceae and their evolution in plants.

REFERENCES

1. Mukherjee, P. K., Kumar, V., Kumar, N. S., and Heinrich, M. (2008) J. Ethnopharmacol. 120, 291-301


FOOTNOTES

This research was supported by Biomedical Research Council (BMRC 09/1/22/19/612) of A*STAR and Academic Research Fund (ARC21/08) of Ministry of Education in Singapore.

The nucleotide sequences reported in this paper have been deposited in the GenBank database under GenBank accession numbers: JF931988 (ctc1), JF931989 (ctc2), JF931990 (ctc3), JF931991 (ctc4), JF931992 (ctc5), JF931993 (ctc7), JF931994 (ctc8), JF931995 (ctc9), JF931996 (ctc12), JF931997 (ctc13), JF931998 (ctc14).
FIGURE LEGENDS

Figure 1. Thermal stability of biologics in the *C. ternatea* flower. (A) MS spectrum of the heat-stable aqueous extract of the *C. ternatea* flower. Peaks labeled K$^+$ indicate the potassium adduct (+38 Da). (B, C) HPLC profile of the flower extract before (B) and after (C) heat treatment at 100 °C for 30 min. The main peaks were labeled according to their cliotide contents.

Figure 2. MS/MS sequencing of cliotide T1. (A) Enzymatic digestion of S-alkylated cliotide T1 by trypsin yielded two fragments with m/z of 812 and 2657 Da; and digestion by chymotrypsin yielded 1395 and 2074 Da fragments. (B-E) MS/MS spectra of digested fragments. Analysis of these spectra gave overlapping sequences indicative of a circular amide backbone of cliotide T1.

Figure 3. Deduced amino acid sequences of cliotide precursors. They display a novel gene arrangement as compared to previously known cyclotide precursors from the Rubiaceae and Violaceae families. The precursors contain an ER signal region directly followed by a mature cyclotide domain, a short linker region, an A1a domain, and a short C-terminal tail. This gene organization resembles that of pea albumin-1 (PA1) precursor except that the A1b domain is replaced by the cyclotide domain. The inverted triangles indicate the location of introns. The *pal* gene can be found in GenBank with the accession number CAB82859.

Figure 4. Schematic comparison of the genetic arrangements of cyclotide (Rubiaceae and Violaceae), cliotide, and A1 (Fabaceae) genes (56, 57). Signal peptides are highlighted in black, N-terminal pro-domains in gray, N-terminal repeats in purple, A1b and cyclotide domains in green, and A1a domains in orange. The inverted triangles indicate the introns located in the signal peptide regions.

Figure 5. Disulfide mapping of cliotide T2. (A) RP-HPLC profile of partially S-reduced and S-NEM labeled cliotide T2. P1-4 indicate peaks 1-4, and labels in the brackets indicate the corresponding peptide species. Peaks 1 and 3 are 2SS and 1SS intermediate species with two and one intact disulfide bonds, respectively. Peak 2 (N) is the native peptide; Peak 4 (RA) is the fully S-NEM labeled peptide. (B) Schematic presentation of disulfide mapping of cliotide T2. 1SS and 2SS species were isolated, reduced with DTT and subsequently S-alkylated with IAA. S-alkylated peptides were treated with trypsin and fragments were sequenced by MS/MS. (C, D) MALDI-TOF MS/MS spectra of trypsin-digested peptide fragments. Their interpreted sequences are shown in the top of the figures.

Figure 6. Tissue-specific distribution of cliotides in *Clitoria ternatea*. Eight different plant parts were collected including flower, stem, shoot, root, leaf, seed, pod, and nodule. The extracts were profiled by mass spectrometry. The peaks are labeled according to the masses of the corresponding cliotides. Peaks labeled K$^+$ indicate the potassium adducts (+38 Da).

Figure 7. Putative enzymes involved in the cyclotide bioprocessing. In *C. ternatea* (Fabaceae), the signal peptidase I (SPase I) responsible for ER excision is also the N-terminal processing enzyme of cliotides. In the Rubiaceae and Violaceae families, an unknown enzyme(s) is required for the N-terminal cleavage of the cyclotide domain. The C-termini of cyclotides in these three families are processed by asparaginyl endopeptidases (AEPs) because of the highly conserved N/D residue at the processing site. The sequence logos were built using WebLogo server (58, 59).

Figure 8. Phylogenetic tree showing the relationship among the cyclotide, cliotide and A1 genes from different Rubiaceae, Violaceae and Fabaceae species. The precursor protein sequences were aligned using COBALT multiple alignment. The phylogenetic tree shows that the cliotide genes cluster with cyclotide genes from the Rubiaceae and Violaceae but not the A1 genes from the Fabaceae.
Table 1. Novel cyclotides from *Clitoria ternatea*

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence</th>
<th>MW a</th>
<th>AA b</th>
<th>Subfamily c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cT1</td>
<td>GI-P-GESCVFIPC-ITGAIGSSCKSKVCYRN</td>
<td>3083</td>
<td>30</td>
<td>B</td>
</tr>
<tr>
<td>cT2</td>
<td>GFLKCGESCVQGEC-YT--P~GISCDWFICKKN</td>
<td>3259</td>
<td>30</td>
<td>M</td>
</tr>
<tr>
<td>cT3</td>
<td>GL-PTGEGETCTLGTG--YV--PCS~SVPICMKN</td>
<td>3057</td>
<td>29</td>
<td>M</td>
</tr>
<tr>
<td>cT4</td>
<td>GI-P-GESCVFIPC-ITAAIGSSCKSKVCYRN</td>
<td>3097</td>
<td>30</td>
<td>B</td>
</tr>
<tr>
<td>cT5</td>
<td>GI-P-GESCVFIPC-ISTVIGSSCKNKVCYRN</td>
<td>3168</td>
<td>30</td>
<td>B</td>
</tr>
<tr>
<td>cT6</td>
<td>SI-P-GESCVYIPC-ITTVGSSCKNSVCYSN</td>
<td>3118</td>
<td>30</td>
<td>B</td>
</tr>
<tr>
<td>cT7</td>
<td>GI-P-GESCVFIPC-CTVALLGSSCKDKVCYKN</td>
<td>3226</td>
<td>31</td>
<td>B</td>
</tr>
<tr>
<td>cT8</td>
<td>GI-P-GESCVFIPC-ISSVVGSSCKSKVCYNN</td>
<td>3071</td>
<td>30</td>
<td>B</td>
</tr>
<tr>
<td>cT9</td>
<td>GI-P-GESCVFIPC-ITTVGSSCKNKVCYNN</td>
<td>3126</td>
<td>30</td>
<td>B</td>
</tr>
<tr>
<td>cT10</td>
<td>GI-P-GESCVFIPCTVALLGSSCKDKVCYKN</td>
<td>3242</td>
<td>31</td>
<td>B</td>
</tr>
<tr>
<td>cT11</td>
<td>GI-P-GESCVFIPCTTALLGSSCKDKVCYKN</td>
<td>3240</td>
<td>31</td>
<td>B</td>
</tr>
<tr>
<td>cT12</td>
<td>GI-P-GESCVFIPC-ITGAIGSSCKSKVCYRD</td>
<td>3084</td>
<td>30</td>
<td>B</td>
</tr>
<tr>
<td>Cter A</td>
<td>d</td>
<td>GVIP-GESCVFIPC-ISTVIGSSCKNKVCYRN</td>
<td>3267</td>
<td>31</td>
</tr>
<tr>
<td>Cter B</td>
<td>d</td>
<td>GV-P-CAESCVWIPCTVALLGSSCKDKVCYLN</td>
<td>3250</td>
<td>31</td>
</tr>
<tr>
<td>Cter D</td>
<td>d</td>
<td>GI-P-CAESCVWIPCTVALLGSSCKDKVCYLN</td>
<td>3264</td>
<td>31</td>
</tr>
<tr>
<td>A1b e</td>
<td>ASCNGVCSPFEMPFGTSACRCIPVGLVIGYCRNPSG</td>
<td>3738</td>
<td>37</td>
<td>A1b</td>
</tr>
</tbody>
</table>

a Molecular weights are reported as monoisotopic mass. b Number of amino acids (AA).

c Subfamily: Bracelet (B), Möbius (M). Cysteine residues are highlighted in yellow, negatively charged residues in red and positively charged residues in blue. d Cter A, Cter B, and Cter D have been previously reported by Poth et al. (27). e A1b (pea albumin-1 chain B) is shown for comparison (57).
Table 2. Antibacterial, hemolytic and cytotoxic activities of selected cliotides

| Peptide | MIC<sup>a</sup> (µM) |  |  |  |  |
|---------|----------------|----------------|----------------|----------------|
|         | <i>E. coli</i> | <i>K. pneumonia</i> | <i>P. aeruginosa</i> |  |  |
| cT1     | 1.1            | 2.7            | 4.7            | 7.1            | 0.6            |
| cT2     | >100           | >100           | >100           | >100           | 8.0            |
| cT3     | >100           | >100           | >100           | 13.1           | 2.0            |
| cT4     | 1.0            | 5.5            | 7.5            | 8.4            | 0.6            |
| D4R<sup>d</sup> | 1.3           | 1.9            | 1.5            | -              | -              |
| Melittin<sup>e</sup> | -                | -                | -                | 3.2            | -              |

<sup>a</sup> Minimal Inhibitory Concentration.  
<sup>b</sup> HD<sub>50</sub> refers to the peptide concentration that causes 50% lysis of RBCs.  
<sup>c</sup> IC<sub>50</sub> refers to the peptide concentration that causes 50% death of HeLa cells.  
<sup>d</sup> D4R is a synthetic antibacterial peptide that was used as the positive control for antibacterial assay.  
<sup>e</sup> Melittin was used as the positive control for hemolytic assay.
Figure 1
Figure 2
Figure 3

Clotide precursor

CTC1

ER

clotide

A1a

Tail

pa1

ER

A1b

A1a

Tail

Figure 3

Pea albumin precursor

17
**Figure 4**

![Diagram showing gene structures from different plant families](image-url)

- **Fabaceae**
  - *pa1* from *P. sativum*
  - *ctc3* from *C. ternatea*
  - *ctc7* from *C. ternatea*

- **Rubiaceae**
  - *Oak1* from *O. affinis*
  - *Oak2* from *O. affinis*

- **Violaceae**
  - *Voc1* from *V. odorata*
  - *Mra4* from *M. ramiflorus*
Figure 5
Figure 5
Figure 6
Figure 7
Figure 8
Discovery and characterization of novel cyclotides originated from chimeric precursors consisting of albumin-1 chain a and cyclotide domains in the fabaceae family

Giang Kien Truc Nguyen, Sen Zhang, Ngan Thi Kim Nguyen, Phuong Quoc Thuc Nguyen, Ming Sheau Chiu, Antony Hardjojo and James P. Tam

J. Biol. Chem. published online May 19, 2011

Access the most updated version of this article at doi: 10.1074/jbc.M111.229922

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2011/05/19/M111.229922.DC1

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2011/05/19/jbc.M111.229922.full.html#ref-list-1