Supporting Information

Poth et al. 10.1073/pnas.1103660108

Text S1. Detailed Materials and Methods

Plant Extraction. Leaf material (approximately 3.5 g) from *C. ternatea* plants (grown in St. Lucia, Brisbane, Australia) was ground using a mortar and pestle prior to solvent extraction with 20 mL of 50% (v/v) acetonitrile, 2% (v/v) formic acid. Crude extract was centrifuged for 4 min at 4,000 × *g*, and the supernatant passed through a 0.45 micron syringe filter prior to lyophilization.

MALDI-TOF MS. MALDI-TOF analyses were conducted using an Applied Biosystems 4700 TOF-TOF Proteomics Analyser. Samples were prepared through 1:1 dilution with matrix consisting of 5 mg mL^{-1} CHCA in 50% (v/v) acetonitrile, 1% (v/v) formic acid prior to spotting on a stainless steel MALDI target. MALDI-TOF spectra were acquired in reflector positive operating mode with source voltage set at 20 kV and Grid1 voltage at 12 kV, mass range 1,000–5,000 Da, focus mass 1,500 Da, collecting 2,000 shots using a random laser pattern and with a laser intensity of 3,500. External calibration was performed by spotting CHCA matrix 1:1 with Applied Biosystems Sequazyme Peptide Mass Standards Kit calibration mixture diluted 1:400 as described previously (1).

Enzymatic Digestion. Prior to tandem MS analyses, cyclotides were cleaved to produce linearized fragments following reduction and alkylation to prevent reoxidation. 1 mg lyophilized crude leaf extract was reconstituted in 100 mM NH_4HCO_3 (pH 8) and reduced by addition of 15 μL of 100 mM dithiothreitol and incubation at 60 °C for 30 min in a nitrogenous atmosphere. Incubation with a further 15 μL of 1 M iodoacetamide followed for 60 min at RT. Alkylated sample was enzymatically digested following addition of 20 μL of 400 ng μL^{-1} endoproteinase Glu-C (Sigma P2922) and incubation at 37 °C for 18 hours. Samples were quenched with formic acid and stored at 4 °C until further analysis.

NanoLC-MS/MS. Reduced, carbamidomethylated, and enzymatically digested (linearized) cyclotide-containing crude leaf extract (5 μL) was injected onto an Agilent 1100 Binary HPLC system (Agilent). Reverse phase separation of samples was achieved using a Vydac MS C18 300 Å, column (150 mm × 300 μm) with a particle size of 5 μm (Grace Davison). The mobile phase consisted of solvent A (0.1% formic acid (aq)) and solvent B (90/10 acetonitrile/0.1% formic acid (aq)) at a flow rate of 4 μL min^{-1}. The column eluate was directly interfaced to a QStar® Elite hybrid LC-MS/MS system (Applied Biosystems/MDS SCIEX) equipped with a nano-electrospray ionization source. MS/MS data-dependent acquisition mode was used in which survey MS spectra were collected (m/z 350 to 1,800) for 1 second followed by three MS/MS measurements on the three most intense parent ions (20 counts/second threshold, +2 to +5 charge state, and m/z 100 to 1,600 mass range for MS/MS), using the manufacturer's "smart exit" and "smart CE" settings. Parent ions previously targeted were excluded from repetitive MS/MS acquisition for 12 seconds (mass tolerance of 250 mDa). Data were acquired and processed using Analyst QS 2.0 software.

RNA Extraction and cDNA Synthesis. Total cellular RNA was extracted from 97 mg leaf tissue of adult butterfly pea (*Clitoria ternatea*) using TRIzol® LS reagent (Invitrogen) according to the manufacturer's instructions. Tissue was resuspended and homogenized in 1 mL TRIzol® LS (Invitrogen) using a Savant FastPrep FP120 beater (Q-BIOgene). The extracted RNA was resuspended in 50 μL DEPC-treated water, quantified by spectrophotometry (A260 nm) using a 2100 bioanalyzer (Agilent Technologies) and stored at −80 °C. DNA contamination in the extracted RNA was removed using a DNA-free kit (Ambion) according to manufacturer's instructions and all RNA samples were subjected to minus reverse transcriptase reactions to validate removal of genomic DNA. Complementary DNA was generated using random hexamers and Superscript III reverse transcriptase (Invitrogen).

Amplification of Initial Products. After alignment of known and newly identified cyclotide protein sequences, a degenerate primer *Ct-For1A* (5′ CCiACNTGYGGNGARACNTG 3′) was designed based upon the PTCGETC motif, frequently observed in Möbius cyclotides. This primer was then used together with an oligo-dT primer (5′ GCCCGGG T_{20} 3′) to amplify products from cDNA using GoTaq® DNA polymerase (Promega). PCR reactions were performed using a PCT200 thermocycler (MJ Research). The reaction profile was 94 °C for 2 min, 40 cycles of 94 °C for 15 s, annealing at 45 °C for 20 s and extension at 72 °C for 1 min, and a final extension at 72 °C for 7 min. PCR products were cloned into pGEM-T Easy (Promega), transformed into DH5α and grown on Luria-Bertani (LB) agar plates containing 20 mM glucose and 100 μg mL^{-1} Ampicillin. Plasmid preparations were performed as per manufacturer's instructions (Qiaprep® miniprep kit, Qiagen), and sequences were confirmed with M13 forward and reverse primers using BIG DYE 3.1 terminator mix on an ABI 3130XL Sequencer (PE Applied Biosystems). Sequence files were analyzed using Sequencher™ 4.1 (GeneCodes).

Rapid Amplification of cDNA Ends (RACE). RACE was performed using the FirstChoice® RLM-RACE kit (Applied Biosystems) according to manufacturer's instructions. First strand cDNA synthesis was performed on leaf-derived RNA. Sequence-specific primers *Cter M-RACE-Rev1* (5′ GGAAACACCAACCAAAATGGATGT 3′) and *Cter M-RACE-Rev2* (5′ TCACTGTTTTTGCATTAGCTGCAA 3′) were designed using Primer 3 (v.0.4.0) program (http://frodo.wi.mit.edu/), and then used for primary and nested PCR amplifications respectively. PCR products were cloned and sequenced.

Confirmation of the Full-Length cDNA Sequence. To confirm that the assembled sequence corresponded to a unique expressed gene transcript, primers *Cter M-SpecFor* (5′ TCCTTATTTTCATCAACTATGGCTTA 3′) and *Cter M-SpecRev* (5′ TCATACATGATCACTTTTAGTTGG 3′) were designed and used to amplify the leaf-derived cDNA. The reaction profile was 94 °C for 3 min, 35 cycles of 94 °C for 15 s, annealing at 60 °C for 20 s and extension at 72 °C for 1 min, and a final extension at 72 °C for 7 min. PCR products were cloned and sequenced.

Peptide Synthesis. Synthetic Cter M was synthesized at 0.5 mmol scale using BOC chemistry with cyclization achieved via a C-terminal thioester as described previously (2). Cleavage of the peptide from the resin was achieved using hydrogen fluoride (HF) with *p*-cresol and *p*-thiocresol as scavengers [9:0.8:0.2 (vol/vol) HF:*p*-cresol:*p*-thiocresol] at −5 to 0 °C for 1.5 h. RP-HPLC with a C_{18} column was used to purify the crude peptide using a gradient of 0–80% solvent B (solvent A: H_2O/0.05% trifluoroacetic acid; solvent B: 90% CH_3CN/10%H_2O/0.045% trifluoroacetic acid) by monitoring the absorbance at 215 nm. The peptide was oxidized in 0.1 M ammonium bicarbonate pH 8, 1 mM reduced glutathione in 50% isopropanol (v/v). The oxidized peptide was

purified by RP-HPLC and the molecular mass and purity determined by ES-MS and RP-HPLC respectively. Synthetic Cter M was coeluted with the native peptide to confirm the correct disulfide connectivity had formed.

NMR Spectroscopy. Spectra were recorded at 600 or 900 MHz (Bruker Avance NMR spectrometers) on a sample containing 1 mM Cter M in 10% D_2O/90% H_2O. The two-dimensional spectra, including TOCSY, COSY, and NOESY, were recorded as previously described (3). Distance restraints were obtained from a NOESY spectrum recorded with a 200-ms mixing time at 290 K. The sequential assignment procedure was used to sequence specifically assign the amino acids (4). The three-dimensional structure of Cter M was calculated by deriving distance and angle restraints from the NOESY and COSY spectra, respectively. A family of structures that are consistent with the experimental restraints was calculated using the programs CYANA (5) and CNS (6). A set of 50 structures was calculated and the 20 lowest energy structures selected for further analysis. Structures were analyzed using the programs PROCHECK_NMR (7) and PROMOTIF (8). MolMol (9) and PyMol (10) were used to display the structural ensembles and surfaces of the peptides, respectively.

Enzymatic Stability. Wild-type Cter M, and reduced and alkylated Cter M, were incubated with trypsin and chymotrypsin at a 20:1 peptide:enzyme ratio. Aliquots were taken after 1 min and after 8 h and analyzed by LC-MS. The peak area and intensity of the cyclic form of each peptide was compared to that at the commencement of the incubation. Linear peptide fragments resulting from proteolytic digestion were also examined.

Preparation of Lipid Vesicles. Small unilamellar vesicles (SUVs) with a diameter of 50 nm were used for surface plasmon resonance (SPR) studies, as described in an earlier study (11). Palmitoyloleoylphosphatidylcholine (POPC) was used to prepare model membranes. Lipid mixtures were prepared of pure POPC. Peptide and lipid samples were prepared in buffer 10 mM HEPES, pH 7.4, containing 150 mM NaCl to represent the physiological pH and ionic strength.

Surface Plasmon Resonance Studies. The binding of the peptides Cter M and kB1 to lipid bilayers was evaluated by means of SPR. All solutions were freshly prepared and filtered (0.22 μm pore size). L1 sensor chips and a BIAcore 3000 system (Biacore, GE Healthcare) were used. Lipids were deposited on the L1 chip, as described previously (12). Association of Cter M and kB1 to the lipid bilayer was evaluated by injection of peptide sample over the lipid surface for 180 s (5 μL min^{-1}), whereas the dissociation was followed for 600 s after the injection. The chip surface was regenerated after each peptide injection cycle, as described previously (12). Measurements were conducted at 25 °C.

Hemolytic Activity Assay. Peptides were dissolved in water and serially diluted in PBS to give 20 μL test solutions in a 96-well U-bottomed microtitre plate (Nunc). Human type A RBCs (red blood cells) were washed with PBS and centrifuged at 1500 g for 60 s in a microcentrifuge several times until a clear supernatant was obtained. A 0.25% suspension of washed RBCs in PBS (100 μL) was added to the peptide solutions. The plate was incubated at 37 °C for 1 h and centrifuged at 150 g for 5 min. Aliquots of 100 μL were transferred to a 96-well flat-bottomed microtitre plate (Falcon) and the absorbance was measured at 405 nm with an automatic Multiskan Ascent plate reader (Labsystems). The amount of hemolysis was calculated as the percentage of maximum lysis (1%Triton X-100 control) after adjusting for minimum lysis (PBS control). Synthetic melittin (Sigma) was used for comparison. The hemolytic dose necessary to lyse 50% of the RBCs (HD_{50}) was calculated using the regression constant from the linear portion of the hemolytic titration curve (Graphpad Prism v5.04).

Insecticidal Assay. *H. armigera* larvae were obtained from the Queensland Department of Employment, Economic Development & Innovation. A feeding trial was conducted for 48 h with larvae maintained at 25 °C throughout the experiment. Larvae were given diets consisting of wheat germ, yeast, and soy flour. Test diets contained Cter M or kalata B1 (used as a positive control) (13), and the control diet did not have any added peptide. Larvae were weighed at 0, 24, and 48 h. Following this, the larvae were photographed. Statistical differences were analyzed using a paired *t*-test or ANOVA test.

Other Supporting Information Files Dataset S1 (PDF)

1. Saska I, Colgrave ML, Jones A, Anderson MA, Craik DJ (2008) Quantitative analysis of backbone-cyclised peptides in plants. *J Chromatogr B* 872:107–114.
2. Clark RJ, Daly NL, Craik DJ (2006) Structural plasticity of the cyclic-cystine-knot framework: implications for biological activity and drug design. *Biochem J* 394:85–93.
3. Rosengren KJ, Daly NL, Plan MR, Waine C, Craik DJ (2003) Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. *J Biol Chem* 278:8606–8616.
4. Wüthrich K (1986) *NMR of Proteins and Nucleic Acids.* (Wiley-Interscience, New York).
5. Guntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278:353–378.
6. Brunger AT (2007) Version 1.2 of the crystallography and NMR system. *Nat Protoc* 2:2728–2733.
7. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486.
8. Hutchinson EG, Thornton JM (1996) PROMOTIF—A program to identify and analyze structural motifs in proteins. *Protein Sci* 5:212–220.
9. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graph* 14:51–55, 29–32.
10. DeLano WL (2002) *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA).
11. Huang YH, et al. (2009) The biological activity of the prototypic cyclotide kalata B1 is modulated by the formation of multimeric pores. *J Biol Chem* 284:20699–20707.
12. Henriques ST, Pattenden LK, Aguilar MI, Castanho MA (2008) PrP(106-126) does not interact with membranes under physiological conditions. *Biophys J* 95:1877–1889.
13. Jennings C, West J, Waine C, Craik D, Anderson M (2001) Biosynthesis and insecticidal properties of plant cyclotides: The cyclic knotted proteins from *Oldenlandia affinis*. *Proc Natl Acad Sci USA* 98:10614–10619.

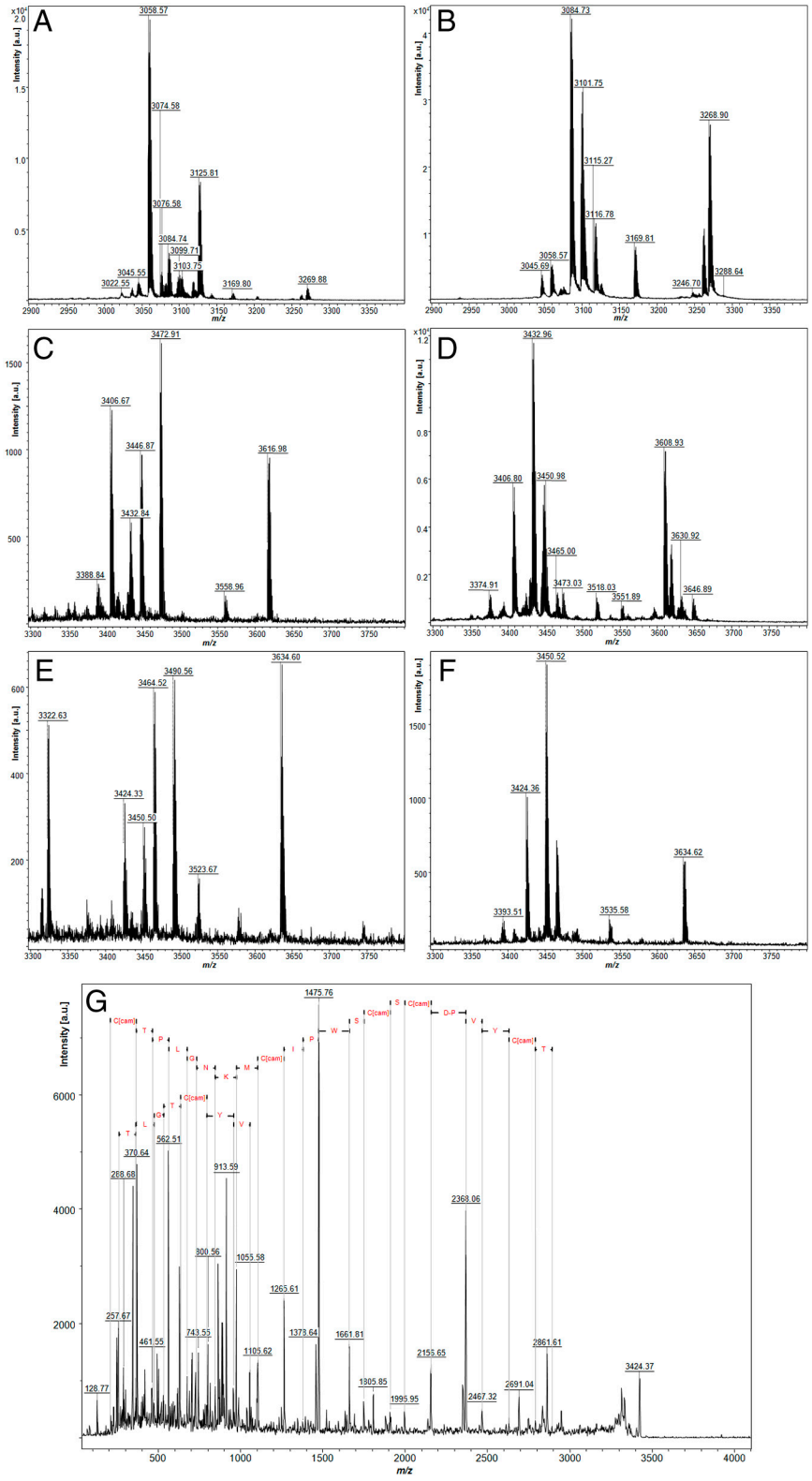

Fig. S1. Detection and sequencing of cyclotides via MALDI-TOF MS. MALDI-TOF mass spectra of crude leaf (*A*) and flower (*B*) extracts. Reduction and alkylation resulted in carbamidomethyl modification of cysteine residues with an increase in mass of 348 Da (leaf, *C*; and flower, *D*). Digestion with endoproteinase Glu-C linearizes the cyclic peptide resulting in a mass increase of 18 Da with cleavage occurring primarily at the conserved Glu in loop 1 (leaf, *E*; flower, *F*). (*G*) MALDI-TOF/TOF mass spectrum of Cter M showing assignment of b- and y-sequence ions. The sequence of Cter M (native mass 3057 Da, alkylated linear mass 3423 Da) was determined to be: TCTLGTCYVPDCSCSWPICMKNGLPTCGE, where methionine was unmodified. The form with oxidized methionine was also detected with mass 3073 Da.

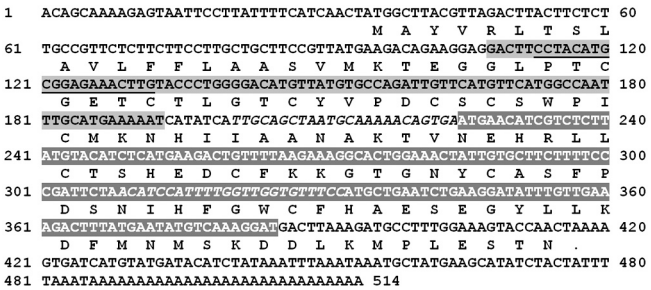

Fig. S2. Complete cDNA sequence and putative translated protein sequence for Cter M isolated from leaf tissue of butterfly pea (*Clitoria ternatea*). The site of initial degenerate primer *CterM-For1A* is underlined, and gene-specific primers used for 5′ RACE are italicized. The cyclotide peptide is highlighted in light gray, and putative albumin-like domain is highlighted in dark gray.

Fig. S3. ClustalW2 alignment of Cter M with BLASTP- and TBLASTN-matched Fabaceae albumin-1 precursor proteins. Orange-highlighted amino acids delineate mature PA1 chain-b peptide sequence in Fabaceae albumins. Red-highlighted amino acids delineate the mature sequence of cyclotide Cter M. Purple-highlighted amino acids delineate predicted mature PA1 chain-a peptide sequence. Sequences originate from: (*A*) *Alysicarpus ovalifolius* (AJ784942.1), (*B*) *Bituminaria bituminosa* (CAH05243.1), (*C*) *Medicago truncatula* (ACJ85909.1), (*D*) *Astragalus monspessulanus* (CAH05241.1), (*E*) *Canavalia brasiliensis* (AJ784944.1), (*F*) *Trigonella foenum-graecum* (CAH05250.1), (*G*) *Medicago truncatula* (CAE00461.1), (*H*) *Medicago truncatula* (CAE00462.1), (*I*) *Glycine soja* (Q9ZQX0.1), (*J*) *Glycine max* (CAA11040.1), (*K*)*Vigna radiata* (AB052881.1), (*L*) *Phaseolus angularis* (Q9FRT9.1), (*M*) *Pisum sativum* (P62929.1), and (*N*)*Clitoria ternatea* (GenBank ID: JF501210). The alignment within the cyclotide domain of the Cter M was manually adjusted to highlight alignment of the six conserved Cys residues with those of the albumin domain.

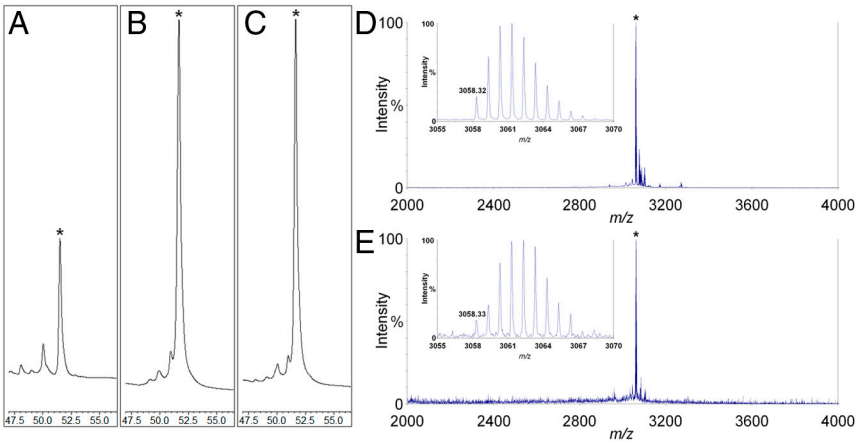

Fig. S4. Analytical HPLC and mass spectrometric analyses show that native and synthetic Cter M are identical. (*A*) Native Cter M; (*B*) synthetic Cter M; and (*C*) coelution of native and synthetic Cter M; MALDI-TOF mass spectra of (*D*) native Cter M extracted from *Clitoria ternatea* leaf material and (*E*) synthetic Cter M. In all panels the asterisk indicates the Cter M peak.

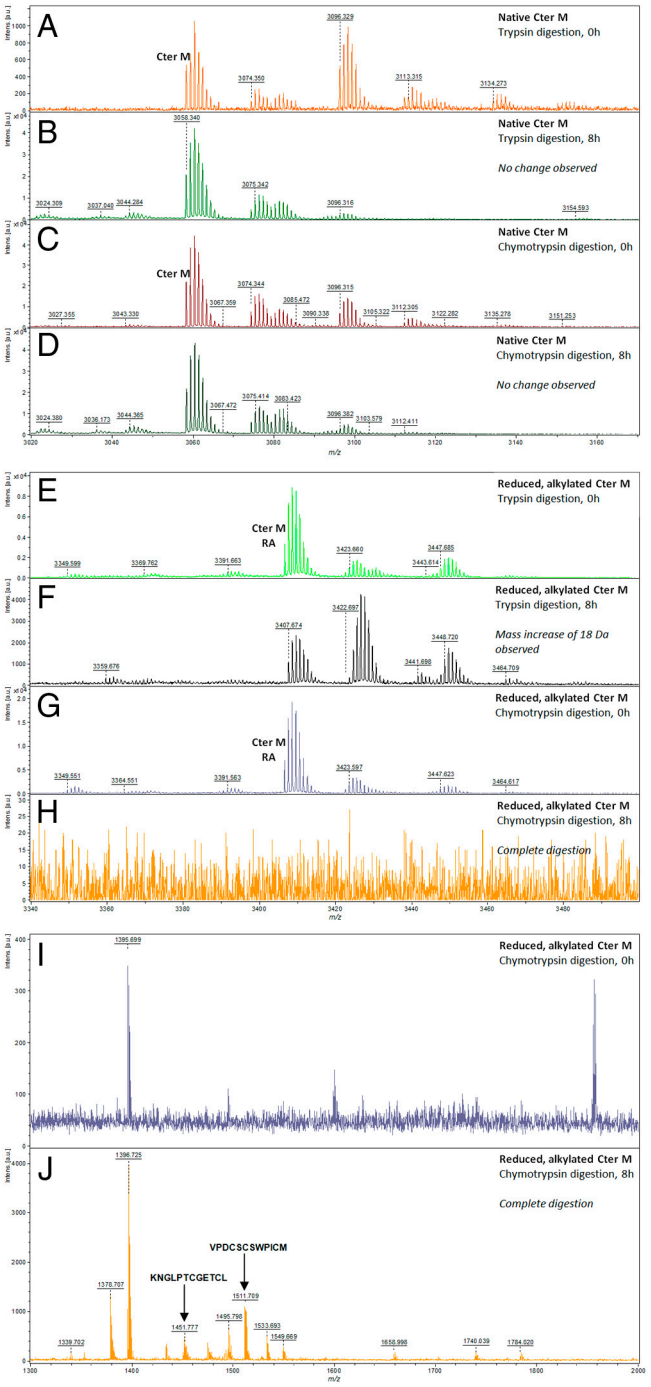

Fig. S5. Cter M is resistant to proteolysis by trypsin and chymotrypsin. Leaf extract showing native Cter M at m/z 3058.3 (*A* and *C*) was subjected to trypsin (*B*) and chymotrypsin (*D*) digestion with no observed hydrolysis. The reduced and alkylated peptide, m/z 3407.6 (*E*, *G*, and *I*) underwent proteolytic cleavage by trypsin (*F*) and chymotrypsin (*H* and *J*). Because there is only a single tryptic site, the trypsin digestion product is observed at m/z 3424.6. Cleavage at the three chymotryptic sites resulted in the formation of major products at m/z 1450.7 and 1511.7 corresponding to KNGLPTCGETCL and VPDCSCSWPICM, respectively.

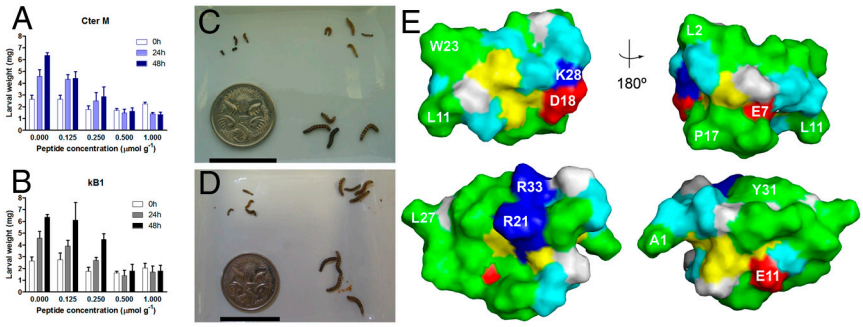

Fig. S6. The effect of Cter M on the growth of *Helicoverpa armigera* and surface representations of Cter M and PA1b. The weight of larvae at 0, 24, and 48 h is plotted versus peptide concentration for Cter M (*A*) and kB1 (*B*) and the size of control larvae alongside larvae fed at medium (0.25 μmol g^{-1} diet) and high (1.0 μmol g^{-1} diet) peptide concentrations at 48 h is depicted for Cter M (*C*) and kB1 (*D*). Scale bar = 20 mm. (*E*) Hydrophobic residues are in green, polar in cyan, negatively charged in red, positively charged in blue, Cys in yellow, and Gly in gray. Cter M (top) and PA1b (bottom).

Table S1. Sequences of novel and known peptides derived from leaf and flower of *Clitoria ternatea*

Peptide Name	Peptide Sequence*,†	Expt *m/z* RAE‡	Expt mass (Da) RAE§	Theor mass (Da) RAE§	Error (ppm)	Mass (Da)¶	Subfamily	Source	Reference
Cter M	TCTLGTCYVPDCSCSWPICMKNGLPTCGE	3424.36	3423.36	3423.25	30.9	3057.25	Möbius	Leaf, Flower	*this work*
Cter N	TCVLGTCYTPDCSCTALVCLKNGSAFCGE	3301.43	3300.43	3300.20	29.2	2934.20	Bracelet	Leaf	*this work*
Cter O	SCVFIPCITGIAGCSCKSKVCYRNGIPCGE	3450.50	3449.50	3449.42	23.9	3083.42	Bracelet	Leaf, Flower	*this work*
Cter P	SCVFIPCITAAIGCSCKSKVCYRNGIPCGE	3464.51	3463.51	3463.43	23.7	3097.43	Bracelet	Leaf, Flower	*this work*
Cter Q	SCVFIPCISTVIGCSCKNKVCYRNGIPCGE	3535.58	3534.58	3534.47	30.2	3168.47	Bracelet	Flower	*this work*
Cter R	SCVFIPCTVTALLGCSCKDKVCYKNGIPCGE	899.17	3592.66	3592.50	43.8	3226.50	Bracelet	Flower	*this work*
Cter A	SCVFIPCISTVIGCSCKNKVCYRNGVIPCGE	3634.62	3633.62	3633.54	21.1	3267.54	Bracelet	Leaf, Flower, Seed‖	(1)
Psyle F	SCVFIPCITAAVGCSCKNKVCYRDGVIPCGE	895.19	3576.74	3576.48	71.3	3210.48	Bracelet	Flower	(2)

*Underlined sequence has been confirmed by PCR.
†Ile/Leu were distinguished by chymotrypsin digests, from gene sequence, or based on homology with known cyclotides.
‡Experimental *m/z* values are provided from either MALDI-TOF MS or ESI-TOF-MS spectra.
§Monoisotopic masses are provided for reduced, alkylated (carbamidomethylated Cys), and EndoGlu-C digested (RAE) peptides.
¶Monoisotopic masses of fully folded native (wild-type) peptides are given.
‖Peptide identified in *C. ternatea* seed material in ref. 1.

1 Poth AG, et al. (2011) Discovery of cyclotides in the Fabaceae plant family provides new insights into the cyclization, evolution and distribution of circular proteins. *ACS Chem Biol* 10.1021/cb100388j.
2 Gerlach SL, Burman R, Bohlin L, Mondal D, Göransson U (2010) Isolation, characterization, and bioactivity of cyclotides from the Micronesian plant *Psychotria leptothyrsa*. *J Nat Prod* 73:1207–1213.

Table S2. NMR and refinement statistics for Cter M

NMR distance and dihedral constraints	
Distance constraints	
Total NOE	398
Intraresidue	84
Sequential ($\|i-j\|=1$)	149
Medium-range ($\|i-j\|<4$)	51
Long-range ($\|i-j>5$)	114
Total dihedral angle restraints	14
Structure Statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.02 ± 0.001
Dihedral angle constraints (°)	0.5 ± 0.08
Max. dihedral angle violation (°)	2
Max. distance constraint violation (Å)	0.2
Deviations from idealized geometry	
Bond lengths (Å)	0.003 ± 0.0001
Bond angles (°)	0.60 ± 0.03
Impropers (°)	0.49 ± 0.03
Average pairwise rmsd* (Å)	
Backbone	0.2 ± 0.11
Heavy	0.58 ± 0.20
Ramachandran statistics	
% in most favored region	76.1
% in additionally allowed region	20.7
% in generously allowed region	3.2

*Pairwise rmsd was calculated among 20 refined structures.