

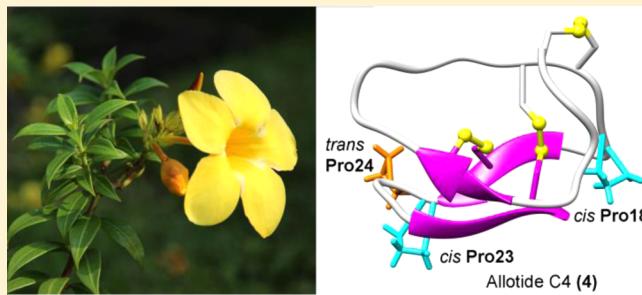
Allotides: Proline-Rich Cystine Knot α -Amylase Inhibitors from *Allamanda cathartica*

Phuong Q. T. Nguyen, Thuy T. Luu, Yang Bai, Giang K. T. Nguyen, Konstantin Pervushin, and James P. Tam*

School of Biological Sciences, Nanyang Technological University, Singapore 637551

Supporting Information

ABSTRACT: Cystine knot α -amylase inhibitors belong to a knottin family of peptidyl inhibitors of 30–32 residues and contain two to four prolines. Thus far, only four members of the group of cystine knot α -amylase inhibitors have been characterized. Herein, the discovery and characterization of five cystine knot α -amylase inhibitors, allotides C1–C5 (Ac1–Ac5) (**1–5**), from the medicinal plant *Allamanda cathartica* are reported using both proteomic and genomic methods. Proteomic analysis showed that **1–5** are 30 amino acids in length with three or four proline residues. NMR determination of **4** revealed that it has two *cis*- and one *trans*-proline residues and adopts two equally populated conformations in solution. Determination of disulfide connectivity of **2** by differential S-reduction and S-alkylation provided clues of its unfolding process. Genomic analysis showed that allotide precursors contain a three-domain arrangement commonly found in plant cystine knot peptides with conserved residues flanking the processing sites of the mature allotide domain. This work expands the number of known cystine knot α -amylase inhibitors and furthers the understanding of both the structural and biological diversity of this type of knottin family.



Cystine knot α -amylase inhibitors (CKAIs) belong to a knottin family of plant-derived α -amylase inhibitors comprising 30 to 32 amino acid residues.^{1,2} As a family, CKAIs is the smallest in size among the seven known families of the proteinaceous α -amylase inhibitors, which have received considerable attention for their therapeutic potential in the treatment and management of obesity and diabetes.³ Similar to other knottins, CKAIs are ribosomally synthesized peptides with a knotted disulfide topology. However, unlike other knottins, they are rich in proline, with at least one of the proline residues existing in a *cis* configuration.^{2,4}

CKAIs are found in the Amaranthaceae and Apocynaceae families, but only four CKAIs have been characterized thus far.^{1,2} The prototypic member, amaranth α -amylase inhibitor (AAI), was first discovered by Chagolla et al. in 1994 from the crop plant *Amaranthus hypochondriacus* (Amaranthaceae).¹ The 32-residue AAI has remained the only known member of this family for the past 20 years. An interest in the possible occurrence of cystine knot peptides in herbal medicine has led us to study the occurrence of CKAIs in medicinal plants of Southeast Asia. Recently, three CKAIs, named wrightide α -amylase inhibitors Wr-AI1 to Wr-AI3, have been reported from the apocynaceous plant *Wrightia religiosa*.² All three wrightides are 30 residues in length, two residues shorter than the prototypic AAI, and thus represent the smallest proteinaceous α -amylase inhibitors found to date.

Structural determination of CKAIs using NMR and X-ray crystallography has revealed their cystine knot (CK) arrange-

ment of three interlocked disulfide bonds, with one disulfide piercing the ring formed by two other disulfide bonds and the intervening backbone segments.^{2,4–7} This knotted arrangement is known to contribute to high resistance to heat and endoproteolytic degradation.^{8,9} In addition to the knotted arrangement, CKAIs possess disulfide bonds located at the ultimate or penultimate Cys residues locking the backbone, which renders them resistant to exopeptidase degradation. Owing to their tolerance to exopeptidase in a manner somewhat similar to the cyclic cystine knot peptides such as cyclotides, CKAIs have been referred to as pseudocyclic cystine knot peptides.^{2,10} Recently, it has been shown that cyclic cystine knot peptides are useful scaffolds for engineering metabolic stable peptide therapeutics.^{11,12} The pseudocyclic cystine knot peptides of the CKAI series may also have the potential to be such useful scaffolds.

NMR experiments showed that AAI and Wr-AI1 adopt well-defined and similar structures in solution.^{2,4,6} Both contain a highly twisted β -sheet formed by three short β -strands composed of two to four residues. These β -strands are connected by four β -turns (Wr-AI1) or a 3_{10} helix turn and three β -turns (AAI). The compact folded structures of AAI and Wr-AI1 are abundant in intramolecular hydrogen bonds with 30–50% exhibiting a slow amide H/D exchange rate. All known CKAIs contain two to four proline residues. Wr-AI1

Received: November 2, 2014

Published: April 2, 2015

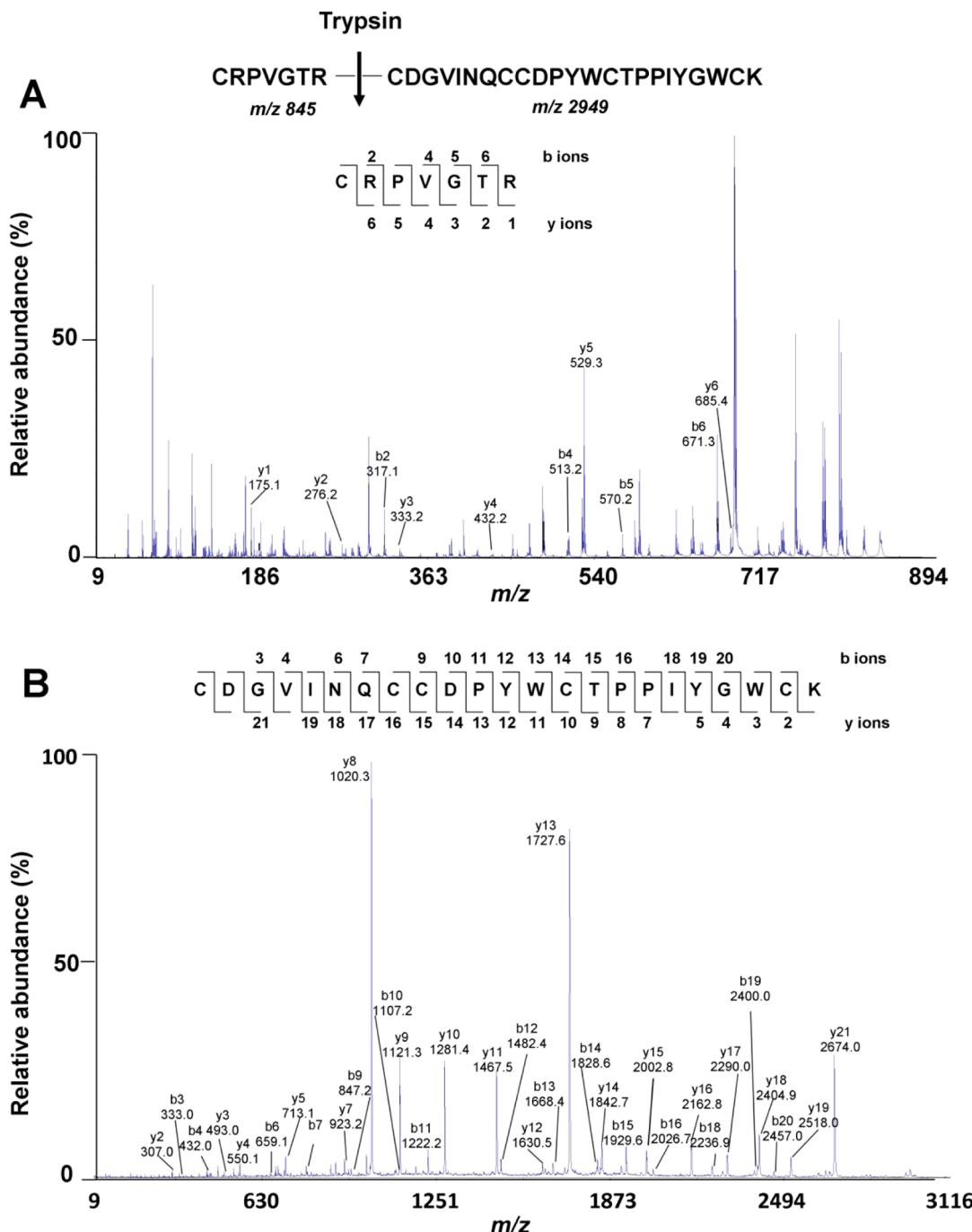


Figure 1. Tandem MS/MS profiles of the two tryptic fragments (m/z 845, A, and 2949, B) used to provide the full sequence of 2. Ile/Leu assignment was determined by amino acid analysis.

contains two proline residues, of which one exists in the *cis* configuration. The solution structure of Wr-AI1 has been confirmed from its X-ray crystal structure at 1.25 Å resolution.² In contrast, AAI contains four prolines, two of which exist in the *cis* configuration.⁴

Both AAI and the wrightides display inhibition against the yellow mealworm *Tenebrio molitor* α -amylase (TMA) but not against human or fungal α -amylases.^{1,2} Previous analysis of a TMA:AAI crystal by Pereira et al. showed that Arg7 (AAI), forming the only salt bridge with the catalytic residue Asp287 (TMA) in a TMA:AAI complex, is the key residue in the inhibitory mechanism of AAI.⁷ However, Arg7 is not found in the wrightides. An equivalent positively charged residue is also

not present in a TMA:Wr-AI1 simulated model.² Molecular dynamics simulation of the modeled TMA:Wr-AI1 complex suggested that AAI and Wr-AI1 interact with TMA in its active site via different interaction networks.² Thus, there is a need to discover additional CKAs from other plant species to further our understanding of their inhibitory mechanisms.

Apocynaceae species have been used in folkloric medicine for treating fever, diarrhea, skin diseases, tumors, cardiopathy, and other diseases.^{13–15} *Allamanda cathartica* L., belonging to the Apocynaceae family, is a small shrub with yellow trumpet flowers, commonly known as Golden Trumpet or Yellow Bell. All plant parts contain the antileukemic iridoid lactone allamandin.¹⁶ The leaves of *A. cathartica* are a source of other

Table 1. Cystine Knot α -Amylase Inhibitors from the Families Apocynaceae and Amaranthaceae

peptide	sequence ^a					no. of Pro	approach ^b	ref.
		loop	1	2	3	4		
allotide C1	Ac1 (1)	CIAHYG-KCDGIIN--Q <u>CCD</u> PWL <u>C</u> TPPIIG-ICI				3	P+G	this work
allotide C2	Ac2 (2)	CRP-VGTR <u>C</u> DGVIN--Q <u>CCD</u> PYW <u>C</u> TPPIYG-WCK				4	P+A	this work
allotide C3	Ac3 (3)	CRP-YGTR <u>C</u> DGVIN--Q <u>CCD</u> PYW <u>C</u> TPPIYG-WCK				4	P+G	this work
allotide C4	Ac4 (4) ^c	CIAHYG-KCDGIIN--Q <u>CCD</u> <u>P</u> WL <u>C</u> TPPIIG-FCL				3	P+G	this work
allotide C5	Ac5 (5)	CVSHYG-KCDGIIN--Q <u>CCD</u> PWL <u>C</u> TPPIIG-FCL				3	P+G	this work
wrightide R1	Wr-AI1 ^c	CAQKGE-Y <u>C</u> SVYLQ--C <u>CD</u> <u>P</u> YH <u>C</u> TQPVIGGI <u>A</u>				2	P+G	²
wrightide R2	Wr-AI2	CAQKGE-Y <u>C</u> SVYLQ--C <u>CK</u> <u>P</u> Y <u>Q</u> C <u>T</u> QPVIGGI <u>A</u>				2	P+G	²
wrightide R3	Wr-AI3	CAQKGE-Y <u>C</u> SVYLQ--C <u>CK</u> <u>P</u> Y <u>R</u> C <u>T</u> QPVIGGI <u>A</u>				2	G	²
AAI ^{cd}		CIPKWN-R <u>C</u> GPKMDGV <u>P</u> CCE <u>P</u> YT <u>C</u> TS <u>D</u> YYG-N <u>C</u> S				4	P	¹

^aDisulfide connectivity is shown by the solid lines. ^bApproach to obtain α -amylase inhibitor sequence: proteomic (P), genomic (G), or amino acid analysis (A). ^cThe *cis*-proline residues in the reported NMR spectroscopic structures of **4** (this work), Wr-AI1, and AAI are highlighted in underlined and italic letters (P), and the *trans*-proline residues are highlighted in italic letters (P). ^dAAI, the α -amylase inhibitor from *Amaranthus hypochondriacus*.

bioactive iridoids such as plumieride, plumiericin, and isoplumiericin.^{17,18} This plant has applications against jaundice and malaria in traditional medicine in Suriname.¹⁴ An aqueous extract of *A. cathartica* leaves exhibited wound-healing effects in a rat model, whereas its ethanolic extract has been found to show anti-inflammatory and antipyretic activities.^{14,19}

In the present study, five new α -amylase inhibitors, members of the CKAI family found in *A. cathartica*, named allotides C1–C5 and abbreviated as Ac1–Ac5 (**1–5**), are reported. A combination of a proteomic and a genomic study showed that allotides are 30-residue pseudocyclic CKAIs that exhibit a high tolerance to degradation by heat and endopeptidase treatment. An NMR spectroscopic study of **4** showed that it contains two *cis* and one *trans* proline bonds within a 30-residue backbone. The cystine knot motif of **2** was also confirmed by a top-down disulfide mapping. Four full-length precursors of allotides were determined by molecular cloning. Their precursors contain three domains and conserved processing sites, a feature common to many plant cysteine-rich peptides. Altogether, this work expands the list of CKAIs from four to nine peptides and furthers our understanding of their structure–function relationship to enable potential applications as a stable scaffold in engineering peptidyl pharmaceuticals.

RESULTS AND DISCUSSION

Isolation of CK α -Amylase Inhibitors from *A. cathartica*. Mass spectrometry (MS)-guided screening of fresh *A. cathartica* leaves (1 g sample) showed the presence of cysteine-rich peptides with a molecular weight of 3–5 kDa and six cysteine residues after reductive S-alkylation.¹⁰ These peptides were extracted in 50% ethanol and fractionated by sequential strong cation-exchange (SCX) and reversed-phase (RP) HPLC as described previously.² Five new peptides, allotides C1–C5 (**1–5**), were isolated in a scale-up purification procedure (800 g of fresh sample). To determine their

sequences, each purified allotide was fully reduced and digested with trypsin and chymotrypsin and subjected to MS/MS. Figure 1 provides an example of the b- and y-ion series of two tryptic fragments of **2** used to derive its amino acid sequence. Ambiguities between isobaric residues such as Ile/Leu were resolved by either amino acid analysis (for **2**, Table S1, Supporting Information) or genomic analysis (for **1** and **3–5**).

The allotides all contain 30 amino acids, which are similar lengthwise to the wrightides, and represent the smallest proteinaceous α -amylase inhibitors known to date.² They share high sequence homology (>75%; Table 1) and cysteine spacing. Pairwise sequence comparison showed that **2** and **3** differ from each other by only one or two residues. Loop 1 (based on N-terminal intercysteinyl residues, Table 1) is the most variable, with only one Gly in the middle of this loop being conserved. In contrast, most residues in the other three loops are highly conserved. The consensus sequence for these three loops among **1–5** is CDGXINQCCDP-XXCTPPIXGXCX, where X represents a variable residue. In comparison with other CKAIs, BLAST analysis revealed a moderate to high sequence homology of 32–50% and 65–75% to AAI and wrightides, respectively.

Disulfide Connectivity and Reductive Unfolding of **2**.

Mapping the disulfide of CK peptides is generally challenging because of the tightly knotted disulfide core. A previous study on the reductive unfolding of AAI failed to detect any intermediates during the reduction with dithiothreitol (DTT).²⁰ Only native or fully reduced AAI species were found, suggesting an all-or-none mechanism. The reported all-or-none model is in contrast with recent findings showing that the unfolding of CK peptides follows a sequential reduction pathway with discrete disulfide intermediates. In this study, the presence of discrete intermediates in the reductive unfolding of **2** using a differential S-reduction and S-alkylation method to map its disulfide connectivity was shown successfully.

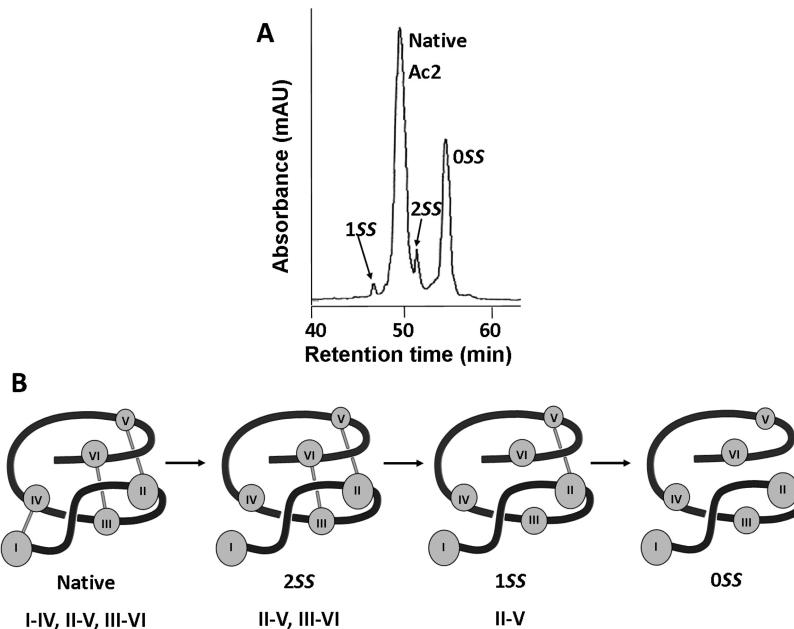


Figure 2. Disulfide mapping of **2**. (A) RP-HPLC profiles of partially S-reduced and NEM-alkylated **2**. (B) Schematic representation of the unfolding process of **2** showing that the Cys I–IV bond is likely the first to be broken, followed by the Cys III–VI bond, and finally the Cys II–V bond.

Purified **2** was partially reduced with tris(2-carboxyethyl)-phosphine to generate intermediates with one (1SS) or two disulfide bonds (2SS) remaining intact. The reduced disulfide was immediately S-alkylated with an excess of *N*-ethyl-maleimide (NEM). Both steps were performed under acidic conditions (pH 3.5) to prevent the scrambling of disulfide bonds. The partially NEM-tagged species were purified by RP-HPLC (Figure 2A) and analyzed by MS. The NEM-tagged 1SS and 2SS species were subsequently fully reduced with DTT and S-alkylated with iodoacetamide (IAA). These S-tagged peptides were sequenced to determine the positions of NEM- and IAA-modified cysteines, which can be used to deduce their disulfide connectivity (Figure S2, Supporting Information). The disulfide linkage was found between Cys I–IV in the 2SS species and between Cys II–V in the 1SS species. The third SS bond between Cys III–VI was obtained by deduction. Together with the NMR spectroscopic results discussed below, the chemical determination of disulfide connectivity of **2** confirmed unequivocally that allotides contain a CK motif involving disulfide bonds between Cys I–VI, Cys II–V, and Cys III–VI.

Figure 2B summarizes the unfolding process during the disulfide reduction steps for **2** by analyzing its partially reduced, discrete intermediate species. Only one 2SS intermediate with a reduced Cys I–IV disulfide linkage was detected in our experiment, suggesting that this S–S bridge is the first to be broken. Likewise, the Cys II–V bond appeared to be the most stable among the three disulfide linkages since only one 1SS species having the intact Cys II–V bond was isolated. Thus, the order of reductive unfolding for **2** could be summarized as Cys I–IV, Cys III–VI, and Cys II–V. This finding is in contrast to those of other CK peptides with a different cysteine motif belonging to the cyclotide family. For both linear cyclotides, such as hedyotide B2,¹⁰ and cyclic cyclotides, such as kalata B1 and clotide T2,^{21,22} the penetrating Cys III–VI bond is the last disulfide to be broken because it is buried and shielded inside the cystine core. The more solvent-exposed outer disulfide bonds forming the embedded ring are normally the labile disulfide in these peptides, and in turn, each will be the first to

be reduced. This result is interesting, considering that **2** and cyclotides are all constrained by a cystine knot arrangement and share a similar size.

Solution Structure of **4 by NMR Spectroscopy.** In agreement with the chemical determination of disulfide connectivity, 2D NMR spectroscopy confirmed that **4** adopts a CK topology with disulfide connectivity of Cys I–IV, Cys II–V, and Cys III–VI (Figure 3A). The penetrating disulfide linkage is Cys III–VI, a conserved feature with AAI, Wr-AI1, and many cystine knot peptides in similar size range of 30–40 residues such as the cyclotides and defensins.^{2,4,23} Three short β -strands were observed for Lys7-Cys8, Leu20-Thr22, and Gly27-Leu30 in the structure of **4**, highly similar to three β -strands featured in Wr-AI1 (Tyr7-Cys8, His19-Cys20, and Gly27-Ala30) or AAI (Arg7-Cys8, Thr22-Thr24, and Tyr28-Cys31).^{2,4} Unlike Wr-AI1, **4** has Lys7 oriented in a conserved position with Arg7 of AAI. Therefore, Lys7 could interact with TMA residues in the active site by forming a similar salt bridge with Asp287, mimicking the inhibitory mechanism of AAI.

The results using NMR spectroscopy also revealed a high content of *cis*-proline (two out of three residues) in the structure of **4**. Pro18 and Pro23 were assigned unambiguously in the *cis* configuration and Pro24 in the *trans* configuration. No mixture of *cis* and *trans* conformers was observed for any of these three proline residues. Precedent literature showed that *cis*-proline occurs at low rate in naturally occurring proteins (6–8%).²⁴ When preceded by an aromatic residue, *cis*-proline exists at significantly higher rate (12–16%) because parallel stacking of the pyrrolidine ring of proline and the aromatic side chain stabilizes the energetically unfavorable *cis* configuration state. In the Wr-AI1 structure, the proximity of a *cis* Pro17 pyrrolidine ring and a Tyr18 phenol ring stabilizes the *cis* Pro17 form.² Hydrogen bonding between Cys15 and Tyr18 also contributes to the stability of *cis* Pro17 in Wr-AI1. In the structure of **4**, parallel stacking of the proline and His4 imidazole rings could be the main stabilizing factor for *cis* Pro18. The distance between the two ring centers varies from 4.8 to 5.1 Å in 18 out of 20 structures in the NMR

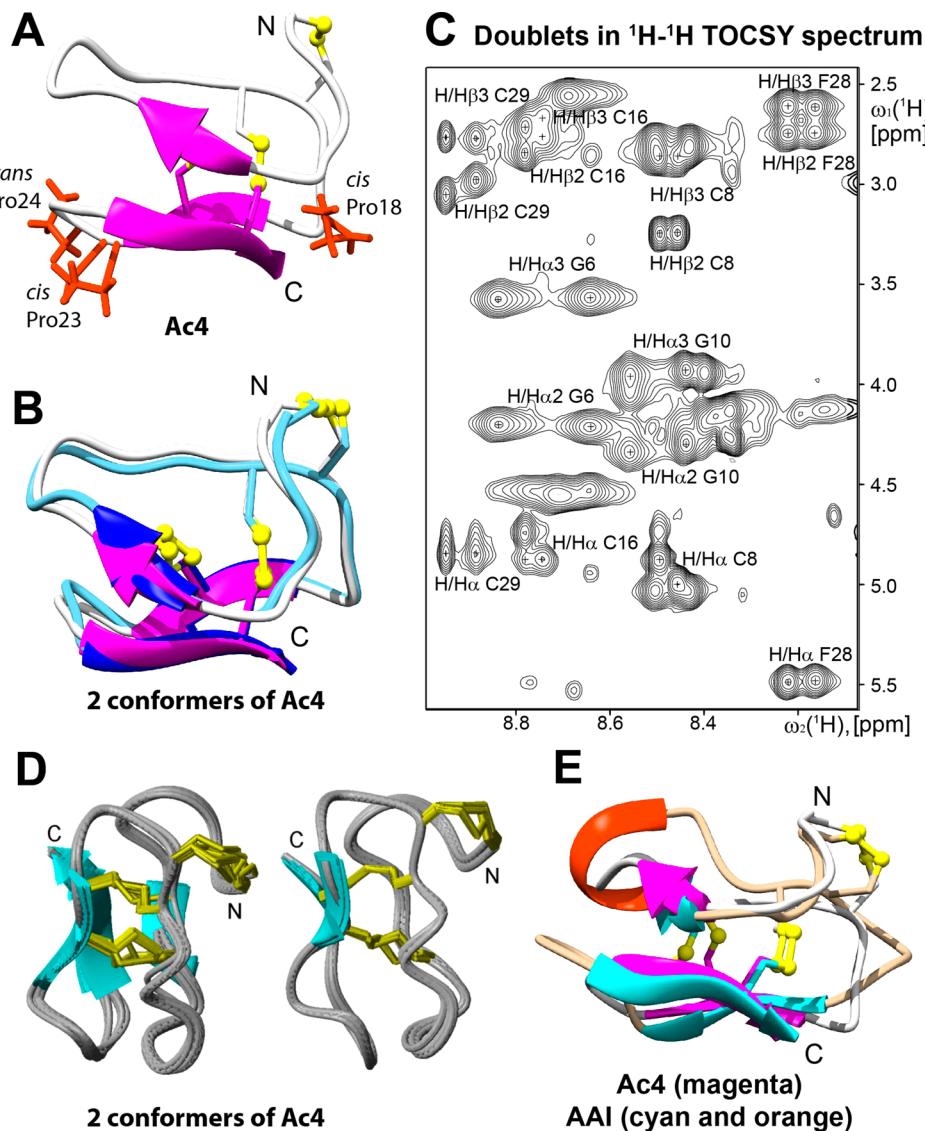


Figure 3. Three-dimensional structure in solution of 4. (A, B, and C) Solution structure and cross-peak doublets of two conformers of 4. The N- and C-termini of the peptide are indicated. Cysteine disulfide bonds are represented as yellow sticks. (D) Ensemble for 20 NMR spectroscopic structures of 4 in solution. Two equally populated conformations are shown on the left and right separately. The backbone of the peptide is shown in a gray color, and short β-sheet regions are labeled in cyan. (E) Alignment of 4 (magenta) and AAI (cyan and orange, PDB code: 1QFD). Mean global backbone root-mean-square deviations (RMSDs) are $0.30 \pm 0.13 \text{ \AA}$ ($0.04\text{--}0.56 \text{ \AA}$) and $0.25 \pm 0.11 \text{ \AA}$ ($0.02\text{--}0.50 \text{ \AA}$).

spectroscopic ensemble. On the other hand, *cis* Pro23 likely gains energetic favorability via loop-stabilizing hydrogen bonding between Thr21 and Phe28.

Despite a high similarity in the overall three-dimensional fold of 4, Wr-AI1 and AAI (RMSDs of superimposition of 4 with Wr-AI1 and AAI are 0.8 and 1.2 Å, respectively), and a conserved *cis*-proline in loop 3, significant secondary structure variations were observed. First, their NMR spectroscopic structures showed a significant backbone difference at the N-terminus, which might result from a notable N-terminal sequence divergence, especially in loop 2. Second, native AAI was found to exist in 70% *trans*:30% *cis* populations.⁴ Structural variation between these *cis/trans* populations was mostly observed in the Pro10-Pro20 segment of AAI, which is attributed to the *cis/trans* isomerization of the Val15-Pro16 peptide bond. In contrast, the solvated 4 adopts two equally populated structural forms, as evident by the cross-peak doublets (Figure 3C) in both the TOCSY and NOESY spectra,

demonstrating slightly different chemical shifts due to backbone resonance. Interestingly, the doubled resonances of 4 span almost the entire sequence (Ile2-Gly6, Cys8, Gly10-Asn13, Cys16, Ile25, Ile26, Phe28, and Cys29) but involve no proline residues.

Each of the two populations of 4 exhibited both right-hand hook and left-hand spiral configurations of three disulfide bridges in their ensemble of 20 final structures. Since the doubled resonances involve no proline, the structural heterogeneity (the presence of two equally populated conformations) is likely attributed to spatial alterations in the disulfide bridge configurations. In the six-class taxonomy of disulfide geometry proposed by Ozhogina and Bominaar, the distance between two C^α atoms of the involved Cys residues spans from <3.8 Å to >6.8 Å.²⁵ This means that the rotation and twisting of disulfide bonds, when propagated along the backbone, can lead to significant rearrangement of local and even distant residues. There have been reports on the

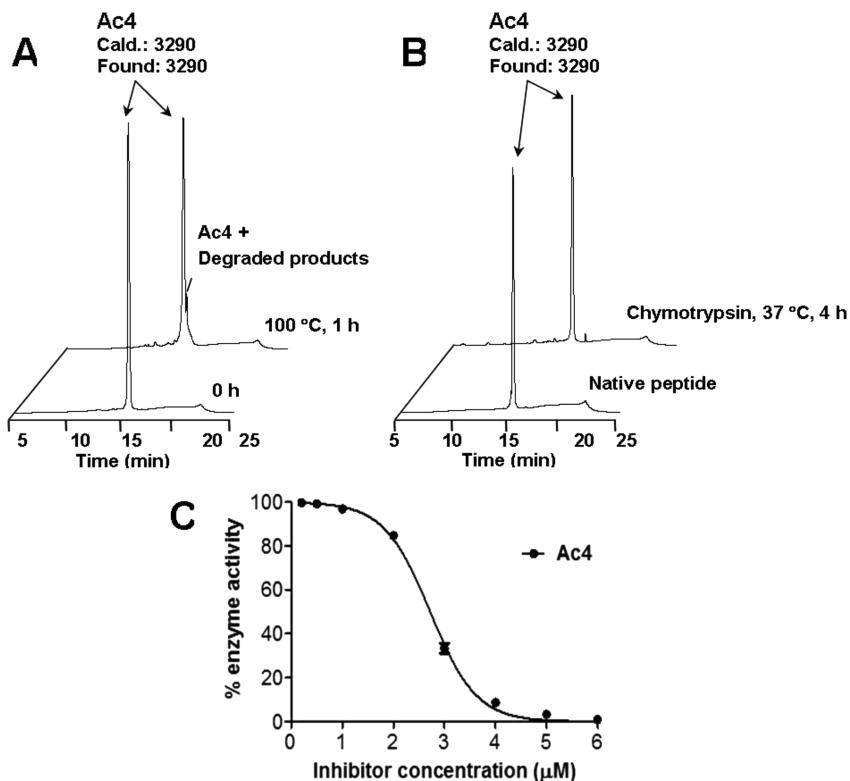


Figure 4. Stability and α -amylase inhibition activity of allptides. (A and B) Thermal and enzymatic stability of Ac4 (4), respectively. The peptide was heated at 100 °C for 1 h or incubated with chymotrypsin at 37 °C for 4 h. (C) Inhibition of *Tenebrio molitor* α -amylase by allptides. The peptide was preincubated with TMA for 20 min at 37 °C. Hydrolysis was started by adding 1% starch, proceeded for 5 min, and was stopped by adding a color reagent containing dinitrosalicylic acid. TMA without peptide treatment was used as the control, and its absorbance at 540 nm was scored 100%.

conformational changes due to disulfide rotation in solvated and crystallized proteins.^{4,26–28} An example is the disulfide conformational switch of the integrin epidermal growth factor 2 (I-EGF2) domain of the $\beta 2$ -subunit of the integrin protein.²⁷ Using X-ray crystallography, Shi et al. showed that the disulfide bond between Cys461 and Cys492 in the I-EGF2 domain assumes different conformations when its dihedral angle switches from right-handed in the bent resting state to left-handed in the extended active state. This leads to substantial rearrangement of neighboring residues, resulting in the displacement of the distal end of the proximal I-EGF1 domain over a distance of >25 Å in the active state. Thus, a similar conformational switch of disulfide bonds may occur in solution, giving rise to the two populations of 4 and their doubling cross-peaks in the NMR spectra.

Heat and Proteolytic Stability. CK peptides with a tightly knotted disulfide core are known for their high tolerance to heat and enzymatic treatments. Such stability is an important attribute for cysteine-rich peptides as putative peptidyl biologics to survive decoction in traditional medicinal preparations and retain both bioavailability and therapeutic efficacy by oral administration. As shown in this work, similar to Wr-AII, 4 with pseudocyclic CK topology can survive boiling water for 1 h, as well as chymotrypsin treatment for 4 h. The treated peptide was monitored by UPLC and MS, both of which showed that >90% of 4 (m/z 3290) remained intact after heat and proteolytic incubation, whereas the fully reduced 4 (as a control) was completely hydrolyzed by chymotrypsin after 4 h (Figure 4A). An 18 Da mass increase of the minor peak in heat-treated samples suggested the incorporation of one water

molecule in the degraded product, which likely results from breakage of the labile Asp17-Pro18 bond.

α -Amylase Inhibitory Activity. To determine the α -amylase inhibitory activities of 4, inhibition assays with α -amylases from the yellow mealworm (*Tenebrio molitor*), human saliva, the porcine pancreas, and a fungus (*Aspergillus oryzae*) were performed using the Bernfeld method.⁹ The results showed that 4 exhibited inhibitory activities against TMA in a dose-dependent manner with an IC_{50} of 2.6 μ M (Figure 4C). Like wrightides and AAI, 4 did not inhibit fungal or mammalian α -amylases under the experimental conditions used.

Other Biological Activities of the Amylase Inhibitors. Biological activities of 4 in cytotoxicity, hemolysis, and antibacterial assays were performed. To evaluate the effect of 4 on cell survival, BHK, Huh-7, and Vero cell viability was measured using Presto blue reagent. The peptide was found to be nontoxic to all three cell lines at concentrations up to 100 μ M. Hemolytic effects were investigated on human type O erythrocytes. Cliotide T4, a cyclic peptide with strong hemolytic activity, was used as a positive control.²² The result showed that 4 did not exhibit a hemolytic effect at concentrations up to 100 μ M. Finally, its antibacterial activity against *E. coli* was evaluated using a radial diffusion assay with D4R, a peptide dendrimer with broad-spectrum antimicrobial activity (synthesized in-house), as a positive control.²⁹ In the experiment performed, 4 did not exhibit bactericidal effects against *E. coli* at concentrations up to 100 μ M.

Biosynthesis of Allptides and Other CKAs. Rapid amplification of cDNA ends (RACE) polymerase chain reaction (PCR) was used to determine allotide precursors to provide an insight into their biosynthesis. 3'-RACE PCR using a

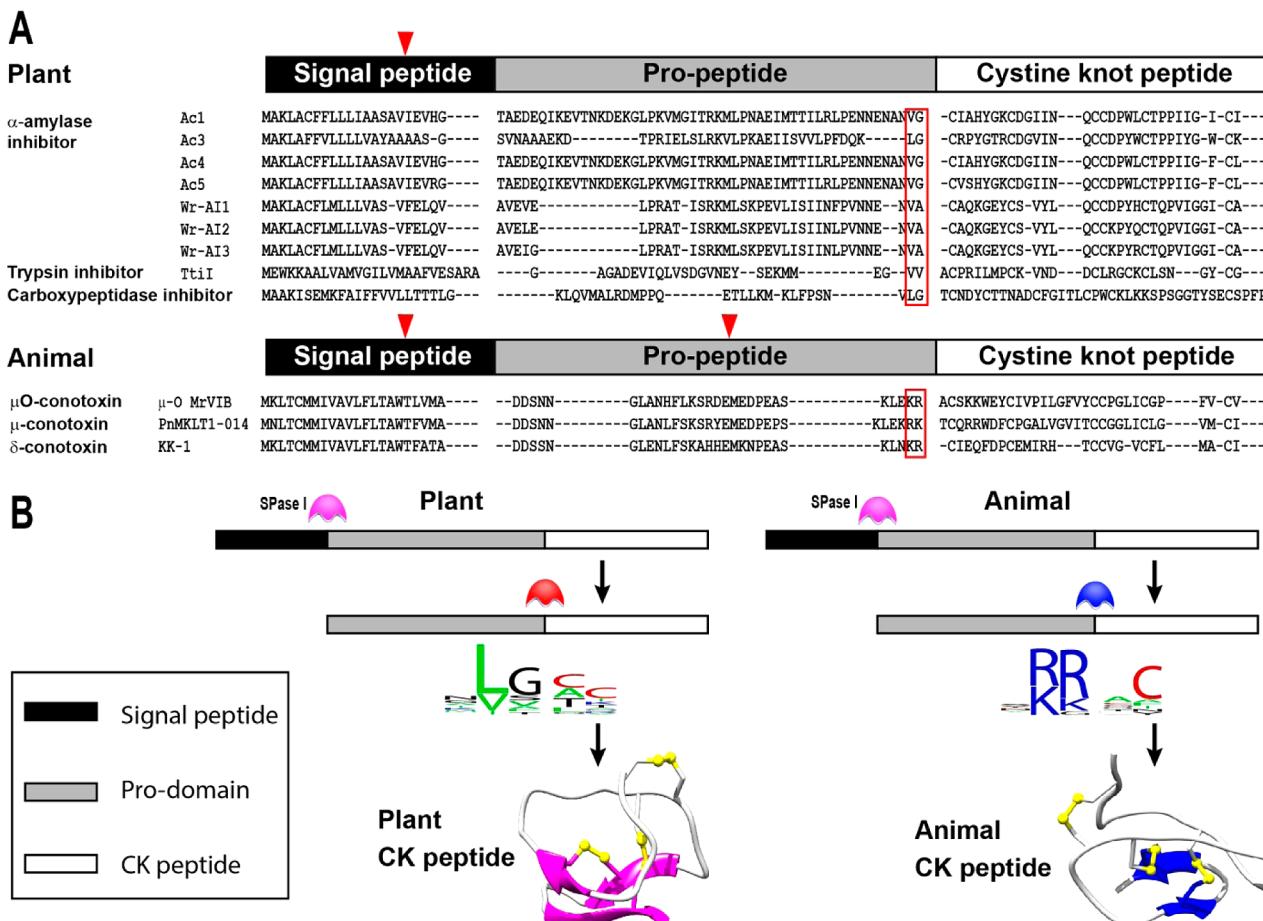


Figure 5. Gene alignment and biosynthesis pathway of cystine knot peptides from plants and animals. (A) Alignment of three-domain precursors of CKAIs and other cystine knot peptides. The red triangle represents an intron in the DNA sequence. Conserved processing sites at the C-terminus of the pro-domains are highlighted in the box. (B) Comparison of the biosynthesis pathways between plant and animal CK peptides. Signal peptide is removed from the full precursors most likely by SPase I. The pro-domain is then cleaved by processing enzymes to release mature CK peptides. These processing enzymes recognize different conserved sequences in plants and animals, as highlighted in the sequence logos. Sequences apart from CKAIs used to prepare the sequence logos were obtained from the KNOTTIN database (<http://knottin.cbs.cnrs.fr/>).^{43,44}

degenerate primer targeting the INQCCDPW sequence resulted in 3'-end partial *acc4/5* genes. For 5'-RACE PCR, a specific primer designed against the 3'-untranslated region of *acc4/5* gene was used. 3'-RACE and 5'-RACE clones were assembled to yield three full-length cDNAs named *acc1*, *acc4*, and *acc5*, encoding 1, 4, and 5, respectively. Similarly, the full transcript of *acc3* using a degenerate primer against the INQCCDPY sequence was obtained. To gain insight into the DNA sequence of novel allotypes, a pair of gene-specific primers targeting 3'- and 5'-untranslated regions were used in PCRs on a genomic DNA extract. The result revealed a phase-one intron with consensus splice junction GT-AG located in the middle of the ER signal peptide for allotype precursors.

Figure 5 presents the alignment of four allotide precursors in comparison with the wrightides and other CK inhibitors, including the towel gourd trypsin inhibitor TGTI-II, potato metallocarboxypeptidase inhibitor PCI, the μ O-conotoxin MrVIB, the μ -conotoxin PnMKLT1-014, and the δ -conotoxin King Kong KK-1.^{30–33} These precursors contain three domains: an ER signal sequence, a pro-peptide, and a mature CK peptide at their C-terminus. The common three-domain structure suggests that the biosynthesis of encoded CK inhibitors undergoes secretory protein processing, which generally involves removal of the signal peptide by SPase I

and subsequent enzymatic pro-peptide cleavage to release the mature functional peptide.^{34–36} Such a biosynthesis pathway is also proposed for many members of the cyclotide family, which is currently the largest cyclic cysteine-rich peptide family from plants. Examples include cyclotides from the Rubiaceae, Violaceae, and Solanaceae families as well as squash trypsin inhibitors from the Cucurbitaceae family.^{37–39}

It has been hypothesized that the evolution of CK peptides involves both convergent and divergent processes.⁴⁰ Convergent evolution is suggested to occur between phyla (animal, plant, fungus), whereas divergence happens within phyla. Convergence theory explains why different organisms across phyla adopt the same advantageous CK fold for functional domain during evolution. These functional peptides are thus conferred with favorable stability due to the same CK fold while their inherent diverse functions are conserved. For example, CK peptides from animals reported so far mainly target voltage-gated ion channels (e.g., Na⁺, K⁺, Cl⁻, and Ca²⁺), whereas CK peptides from plants, also belonging to the knottin family, generally inhibit digestive enzymes (e.g., trypsin, α -amylase, and metallocarboxypeptidase).⁴¹ Other domains, particularly the pro-domain or C-tail, are not targeted for this convergence and thus remain different between phyla. Examples include the common presence of one or two C-terminal basic residues (Arg

or Lys) in animal CK pro-peptides as opposed to one uncharged hydrophobic residue (Leu or Val) at position P2' in plant CK inhibitor precursors (Figure 5). In addition, animal CK peptide precursors usually contain two introns, while those from plants have only one, as shown in the precursors of allotides.

Divergence within phyla of CK peptides explains why selective mutations in a common ancestor precursor give rise to a wide range of defense compounds with fine-tuned functions. For example, many plant CK inhibitors inhibit insect proteases and carboxypeptidases, whereas allotides and their homologues target insect α -amylases.^{30,42} The selective mutations occur at higher rates in the functional domain, generating functional diversification. An example is that three allotide precursors, *acc1*, *acc4*, and *acc5*, contain an absolutely conserved ER signal domain and pro-domain but different functional domains. Altogether, this work provides supporting evidence for hypotheses on the convergent evolution across phyla and divergent evolution within phyla of animal and plant CK peptides.

Potential Applications of Pseudocyclic CKAIs. Compared to other families of plant cysteine-rich peptides such as defensins, heveins, and cyclotides, the number of CKAIs is small, and much remains to be elucidated. This work not only expands this family from four to nine members but also provides their defining features, features commonly found in plant cysteine-rich peptides and features especially associated with CKAIs.

Common features of CKAIs and other CK peptides include the following: (1) They display a well-defined structure in both solution and crystal, with two to three short antiparallel β -strands forming the backbone of a cystine knot as reported for cyclotides, squash trypsin inhibitors, and the plant leginsulins;^{45–47} (2) they are resistant to heat and proteolytic degradation due to a compact structure rigidified by a knotted cystine scaffold; and (3) they have three-domain precursor structures: a signal peptide, a pro-domain, and a mature peptide domain, an arrangement that is common to hevein, carboxypeptidase inhibitors, and several linear cyclotides from *Panicum laxum* and *Chassalia chartacea*.^{9,39,42,48}

Apart from their conserved cysteine motif, all nine members of CKAIs from their two plant families of origin share the following features. (1) They contain 30–32 residues with 2–4 prolines and display high sequence homology (Table 1). (2) Pro18 in loop 3 and Gly27 in loop 4 (numbering based on the sequence of 4) are conserved among nine CKAIs. Pro18 exists in the *cis* configuration for all three reported structures of CKAIs. In apocynaceous plants, Pro24 (numbering based on the sequence of 4) is conserved in both 4 and Wr-AI1 as a *trans* configuration. (3) Structural studies suggest that CKAIs interact with TMA via different interaction networks despite their high sequence homology.

Similar to other cystine knot peptides, CKAIs could provide a useful scaffold for engineering metabolic-stable pharmaceuticals. Previously, a cyclotide scaffold was used to engineer bradykinin antagonists as orally active analgesics.¹¹ The discovery of novel CKAIs from a wide range of sources suggests that the pseudocyclic CKAIs scaffold has high sequence tolerance and thus is promising for such a grafting strategy. Furthermore, the pseudocyclic CKAIs scaffold, rich in proline residues for which the secondary amide bonds are generally not solvated and more readily partition into the lipid membrane, could be advantageous for designing orally active peptides.

EXPERIMENTAL SECTION

General Experimental Procedures. NMR experiments were performed using a Bruker Avance II 700 MHz spectrometer equipped with a TXI cryoprobe at 25 °C. Mass spectrometry analysis for the crude extract and HPLC fractions were performed on the ABI 4800 MALDI-TOF/TOF system (Applied Biosystems, Framingham, MA, USA). HPLC and UPLC were performed on Shimadzu systems. Grace Vydac C₁₈ columns (particle size 5 μ m, pore size 300 Å, Columbia, MD, USA) with dimensions of 250 \times 22 mm, 250 \times 10 mm, and 250 \times 4.6 mm were used for preparative, semipreparative, and analytical RP-HPLC. PolyLC polysulfethyl A columns (250 \times 9.4 mm and 250 \times 4.6 mm) were used for SCX-HPLC. Absorbance in cytotoxicity, hemolysis, and α -amylase assays was acquired using the Infinite@ 200 PRO Tecan microplate reader (Tecan Group Ltd., Männedorf, Switzerland). Chemical reagents used in this study were of analytical grade and purchased from Sigma-Aldrich.

Plant Material. *Allamanda cathartica* was collected from April 2012 to January 2015 at Nanyang Technological University, Singapore. The plant was identified by Hui J. Lam and Elango Velautham at Singapore Botanic Garden, Singapore. A voucher specimen (SING-2015-045) was deposited at the Singapore Botanic Garden Herbarium.

Isolation of α -Amylase Inhibitors. A quantity (800 g) of the leaves of *A. cathartica* was homogenized and extracted twice in 50% ethanol (v/v). The clear supernatant after centrifugation (8000g, 10 min, 4 °C) was partitioned with dichloromethane. The filtered upper layer was loaded onto a C₁₈ flash column and eluted with increasing concentrations of EtOH (20%, 40% to 70% with 10% interval). To purify individual CKAIs, several dimensions of SCX and RP-HPLC were employed. SCX utilized a linear gradient from buffer A (20% acetonitrile, 20 mM KH₂PO₄, pH 2.8) to buffer B (20% acetonitrile, 0.5 M KCl, 20 mM KH₂PO₄, pH 2.8). CKAIs-containing fractions were subsequently purified using RP-HPLC with buffer A (0.1% trifluoroacetic acid in water) and buffer B (100% acetonitrile, 0.1% trifluoroacetic acid). The yields for 1–5 were approximately 10, 3, 5, 15, and 10 mg, respectively. De novo sequencing of the purified allotides were essentially performed as described previously.²

Top-Down Disulfide Mapping. Peptide 2 (0.2 mg) was partially reduced in 100 mM citrate buffer (pH 3.0) with 20 mM tris(2-carboxyethyl) phosphine at 37 °C for 30 min. NEM was added to a final concentration of 40 mM and incubated at 37 °C for 40 min. The reaction was quenched by loading the sample into an analytical C₁₈ column, and NEM-alkylated intermediate species were separated with a linear gradient of 10–60% buffer B. Intermediate species were fully reduced with 20 mM DTT (37 °C, 40 min) and S-alkylated with 40 mM IAA (37 °C, 30 min) before purification by HPLC. S-Alkylated peptides were then subjected to trypsin digestion and MS/MS sequencing.

NMR Study of the Solution Structure of 4. The peptide, concentrated to 1 mM, was exchanged with 20 mM phosphate buffer saline at pH 6. Two-dimensional homonuclear ¹H–¹H TOCSY, NOESY, and ROESY spectra were acquired using mixing times of 80, 150, and 100 ms, respectively. Residue-specific resonance assignment and assignment of NOESY cross-peaks were performed using CARA (www.nmr.ch). Three-dimensional structures of 4 were reconstructed using CYANA 2.0.⁴⁹ The refined structures of the two conformers of 4 (Tables S2 and S3, Supporting Information) are available in the Protein Data Bank database under accession codes 2M4D and 2M4C, respectively.

Cloning of Allotide Genes. Total RNA was extracted using PureLink Mini RNA purification kit (Invitrogen, Carlsbad, CA, USA) and converted to 3' and 5' RACE cDNA libraries using the 3' RACE System for Rapid Amplification of cDNA Ends (Invitrogen) and SMARTer RACE cDNA amplification kit (Clontech, Takara Biotechnology, Dalian, People's Republic of China), respectively. The degenerate primers targeting the sequences INQCCDPY (5'-ATTAATCArTGyTGyGArCCnTA-3') and INQCCDP (5'-ATTAAYCArTGyTGyGAyCC-3') were used in RACE PCRs. The products of expected size were cloned into pGEM-T Easy Vector

(Promega, Madison, WI, USA) and sequenced. Based on the partial genes, specific reverse primers targeting the 3' untranslated region were designed to amplify the 5' end of the transcripts in 5' RACE PCR. To study intron location, genomic DNA was extracted from the *A. cathartica* leaves using the PureLink Plant Total DNA purification kit (Invitrogen). Two specific primers designed against untranslated regions of each allotide were used in PCRs on a DNA template. The RNA sequences for **1** and 3–5 and DNA sequences for **1**, **4**, and **5** reported in this paper have been deposited in the GenBank database under GenBank accession numbers JX437181, JX437182, JX437183, JX437184, JX437178, JX437179, and JX437180, respectively.

Stability Tests and Biological Assays. Heat and endoproteolytic stability tests as well as biological assays (α -amylase inhibition, cytotoxicity, hemolysis, and antimicrobial assays) were performed essentially as reported previously.² Enzyme TMA from *Tenebrio molitor* larvae was extracted as described by Strobl and his co-workers.⁵⁰

ASSOCIATED CONTENT

Supporting Information

MS/MS spectra for sequence determination of **3**, amino acid analysis result for **2**, structure statistics of conformers of **4**, tandem MS/MS profiles for disulfide mapping, and the aliphatic region of the NOESY spectrum for **4**. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Tel: 65-6316 2863. Fax: 65-6515 1632. E-mail: jptam@ntu.edu.sg.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported in part by the Competitive Research Grant by National Research Foundation in Singapore (NRF-CRP8-2011-05). We thank Ms. H. J. Lam and Mr. E. Velautham at Singapore Botanic Garden, Singapore, for identifying the plant material.

REFERENCES

- (1) Chagolla-Lopez, A.; Blanco-Labra, A.; Pathy, A.; Sanchez, R.; Pongor, S. *J. Biol. Chem.* **1994**, *269*, 23675–23680.
- (2) Nguyen, P. Q. T.; Wang, S.; Kumar, A.; Yap, L. J.; Luu, T. T.; Lescar, J.; Tam, J. P. *FEBS J.* **2014**, *281*, 4351–4366.
- (3) Svensson, B.; Fukuda, K.; Nielsen, P. K.; Bønsager, B. C. *Biochim. Biophys. Acta* **2004**, *1696*, 145–156.
- (4) Martins, J. C.; Enassar, M.; Willem, R.; Wierzeski, J. M.; Lippens, G.; Wodak, S. J. *Eur. J. Biochem.* **2001**, *268*, 2379–2389.
- (5) Carugo, O.; Lu, S.; Luo, J.; Gu, X.; Liang, S.; Strobl, S.; Pongor, S. *Protein Eng.* **2001**, *14*, 639–646.
- (6) Lu, S.; Deng, P.; Liu, X.; Luo, J.; Han, R.; Gu, X.; Liang, S.; Wang, X.; Li, F.; Lozanov, V.; Pathy, A.; Pongor, S. *J. Biol. Chem.* **1999**, *274*, 20473–20478.
- (7) Pereira, P. J.; Lozanov, V.; Pathy, A.; Huber, R.; Bode, W.; Pongor, S.; Strobl, S. *Structure* **1999**, *7*, 1079–1088.
- (8) Colgrave, M. L.; Craik, D. J. *Biochemistry* **2004**, *43*, 5965–5975.
- (9) Nguyen, G. K. T.; Lian, Y.; Pang, E. W. H.; Nguyen, P. Q. T.; Tran, T. D.; Tam, J. P. *J. Biol. Chem.* **2013**, *288*, 3370–3380.
- (10) Nguyen, G. K. T.; Zhang, S.; Wang, W.; Wong, C. T. T.; Nguyen, N. T. K.; Tam, J. P. *J. Biol. Chem.* **2011**, *286*, 44833–44844.
- (11) Wong, C. T.; Rowlands, D. K.; Wong, C. H.; Lo, T. W.; Nguyen, G. K.; Li, H. Y.; Tam, J. P. *Angew. Chem., Int. Ed.* **2012**, *51*, 5620–5624.
- (12) Chan, L. Y.; Gunasekera, S.; Henriques, S. T.; Worth, N. F.; Le, S.-J.; Clark, R. J.; Campbell, J. H.; Craik, D. J.; Daly, N. L. *Blood* **2011**, *118*, 6709–6717.
- (13) Arulmozhi, S.; Mazumder, P.; Ashok, P.; Narayanan, S. *Pharmacogn. Rev.* **2007**, *1*, 163–170.
- (14) Nayak, S.; Nalabothu, P.; Sandiford, S.; Bhogadi, V.; Adogwa, A. *BMC Complementary Altern. Med.* **2006**, *6*, DOI: 10.1186/1472-6882-6-12.
- (15) Perez, R. M.; Zavala, M. A.; Perez, S.; Perez, C. *Phytomedicine* **1998**, *5*, 55–75.
- (16) Kupchan, S. M.; Dessertine, A. L.; Blaylock, B. T.; Bryan, R. F. *J. Org. Chem.* **1974**, *39*, 2477–2482.
- (17) Tiwari, T. N.; Pandey, V. B.; Dubey, N. K. *Phytother. Res.* **2002**, *16*, 393–394.
- (18) Abdel-Kader, M. S.; Wisse, J.; Evans, R.; van der Werff, H.; Kingston, D. G. *J. Nat. Prod.* **1997**, *60*, 1294–1297.
- (19) Saranya, S.; Chitra, M. *VRI Phytomed.* **2013**, *2*, 13–16.
- (20) Cemazar, M.; Zahariev, S.; Pongor, S.; Hore, P. J. *J. Biol. Chem.* **2004**, *279*, 16697–16705.
- (21) Goransson, U.; Craik, D. J. *J. Biol. Chem.* **2003**, *278*, 48188–48196.
- (22) Nguyen, G. K.; Zhang, S.; Nguyen, N. T.; Nguyen, P. Q.; Chiu, M. S.; Hardjojo, A.; Tam, J. P. *J. Biol. Chem.* **2011**, *286*, 24275–24287.
- (23) Chen, B.; Colgrave, M. L.; Daly, N. L.; Rosengren, K. J.; Gustafson, K. R.; Craik, D. J. *J. Biol. Chem.* **2005**, *280*, 22395–22405.
- (24) Milner-White, E. J.; Bell, L. H.; Maccallum, P. H. *J. Mol. Biol.* **1992**, *228*, 725–734.
- (25) Ozhogina, O. A.; Bominaar, E. L. *J. Struct. Biol.* **2009**, *168*, 223–233.
- (26) Pierson, N. A.; Chen, L.; Russell, D. H.; Clemmer, D. E. *J. Am. Chem. Soc.* **2013**, *135*, 3186–3192.
- (27) Shi, M. L.; Foo, S. Y.; Tan, S. M.; Mitchell, E. P.; Law, S. K. A.; Lescar, J. *J. Biol. Chem.* **2007**, *282*, 30198–30206.
- (28) Fass, D. *Annu. Rev. Biophys.* **2012**, *41*, 63–79.
- (29) Tam, J. P.; Lu, Y. A.; Yang, J. L. *Eur. J. Biochem.* **2002**, *269*, 923–932.
- (30) Ling, M. H.; Qi, H. Y.; Chi, C. W. *J. Biol. Chem.* **1993**, *268*, 810–814.
- (31) Villanueva, J.; Canals, F.; Prat, S.; Ludevid, D.; Querol, E.; Avilés, F. X. *FEBS Lett.* **1998**, *440*, 175–182.
- (32) McIntosh, J. M.; Hasson, A.; Spira, M. E.; Gray, W. R.; Li, W. Q.; Marsh, M.; Hillyard, D. R.; Olivera, B. M. *J. Biol. Chem.* **1995**, *270*, 16796–16802.
- (33) Woodward, S. R.; Cruz, L. J.; Olivera, B. M.; Hillyard, D. R. *EMBO J.* **1990**, *9*, 1015–1020.
- (34) Gillon, A. D.; Saska, I.; Jennings, C. V.; Guarino, R. F.; Craik, D. J.; Anderson, M. A. *Plant J.* **2008**, *53*, 505–515.
- (35) Mergaert, P.; Nikovics, K.; Kelemen, Z.; Maunoury, N.; Vaubert, D.; Kondorosi, A.; Kondorosi, E. *Plant Physiol.* **2003**, *132*, 161–173.
- (36) Davids, B. J.; Reiner, D. S.; Birkeland, S. R.; Preheim, S. P.; Cipriano, M. J.; McArthur, A. G.; Gillin, F. D. *PLoS One* **2006**, *1*, e44.
- (37) Jennings, C.; West, J.; Waine, C.; Craik, D.; Anderson, M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10614–10619.
- (38) Dutton, J. L.; Renda, R. F.; Waine, C.; Clark, R. J.; Daly, N. L.; Jennings, C. V.; Anderson, M. A.; Craik, D. J. *J. Biol. Chem.* **2004**, *279*, 46858–46867.
- (39) Mylne, J. S.; Chan, L. Y.; Chanson, A. H.; Daly, N. L.; Schaefer, H.; Bailey, T. L.; Nguyencong, P.; Cascales, L.; Craik, D. J. *Plant Cell* **2012**, *24*, 2765–2778.
- (40) Zhu, S. Y.; Darbon, H.; Dyason, K.; Verdonck, F.; Tytgat, J. *FASEB J.* **2003**, *17*, 1765–1767.
- (41) Kolmar, H. *FEBS J.* **2008**, *275*, 2684–2690.
- (42) Villanueva, J.; Canals, F.; Prat, S.; Ludevid, D.; Querol, E.; Aviles, F. X. *FEBS Lett.* **1998**, *440*, 175–182.
- (43) Gracy, J.; Le-Nguyen, D.; Gelly, J.-C.; Kaas, Q.; Heitz, A.; Chiche, L. *Nucleic Acids Res.* **2008**, *36*, D314–D319.
- (44) Gelly, J. C.; Gracy, J.; Kaas, Q.; Le-Nguyen, D.; Heitz, A.; Chiche, L. *Nucleic Acids Res.* **2004**, *32*, D156–D159.

- (45) Craik, D. J.; Daly, N. L.; Bond, T.; Waine, C. *J. Mol. Biol.* **1999**, *294*, 1327–1336.
- (46) Heitz, A.; Hernandez, J.-F.; Gagnon, J.; Hong, T. T.; Pham, T. T. C.; Nguyen, T. M.; Le-Nguyen, D.; Chiche, L. *Biochemistry* **2001**, *40*, 7973–7983.
- (47) Yamazaki, T.; Takaoka, M.; Katoh, E.; Hanada, K.; Sakita, M.; Sakata, K.; Nishiuchi, Y.; Hirano, H. *Eur. J. Biochem.* **2003**, *270*, 1269–1276.
- (48) Broekaert, I.; Lee, H. I.; Kush, A.; Chua, N. H.; Raikhel, N. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7633–7637.
- (49) Guntert, P.; Mumenthaler, C.; Wuthrich, K. *J. Mol. Biol.* **1997**, *273*, 283–298.
- (50) Strobl, S.; GomisRuth, F. X.; Maskos, K.; Frank, G.; Huber, R.; Glockshuber, R. *FEBS Lett.* **1997**, *409*, 109–114.