The Evolution of *Momordica* Cyclic Peptides

Tunjung Mahatmanto,[1] Joshua S. Mylne,[2] Aaron G. Poth,[1] Joakim E. Swedberg,[1] Quentin Kaas,[1] Hanno Schaefer,[3] and David J. Craik*[,1]

[1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Qld, Australia

[2]The University of Western Australia, School of Chemistry and Biochemistry & The ARC Centre of Excellence in Plant Energy Biology, Crawley, Perth, WA, Australia

[3]Plant Biodiversity Research, Technische Universität München, Freising, Germany

***Corresponding author:** E-mail: d.craik@imb.uq.edu.au.

Associate editor: Brandon Gaut

Abstract

Cyclic proteins have evolved for millions of years across all kingdoms of life to confer structural stability over their acyclic counterparts while maintaining intrinsic functional properties. Here, we show that cyclic miniproteins (or peptides) from *Momordica* (Cucurbitaceae) seeds evolved in species that diverged from an African ancestor around 19 Ma. The ability to achieve head-to-tail cyclization of *Momordica* cyclic peptides appears to have been acquired through a series of mutations in their acyclic precursor coding sequences following recent and independent gene expansion event(s). Evolutionary analysis of *Momordica* cyclic peptides reveals sites that are under selection, highlighting residues that are presumably constrained for maintaining their function as potent trypsin inhibitors. Molecular dynamics of *Momordica* cyclic peptides in complex with trypsin reveals site-specific residues involved in target binding. In a broader context, this study provides a basis for selecting *Momordica* species to further investigate the biosynthesis of the cyclic peptides and for constructing libraries that may be screened against evolutionarily related serine proteases implicated in human diseases.

***Key words:* Momordica seeds, cyclic cystine knot peptides, serine protease inhibitors, evolution.**

Article

Introduction

Head-to-tail or backbone cyclization confers peptides with resistance to proteolysis and peptides bearing this trait presumably evolved from ancestral acyclic peptides (Trabi and Craik 2002). Backbone-cyclized peptides found in angiosperms are here categorized into three groups based on their structures (fig. 1). Group 1 consists of cyclic peptides with three disulfide bonds that form a knotted core, that is, the cyclotides (Craik et al. 1999). Group 2 consists of cyclic peptides with one disulfide bond, which includes SFTI-1 or sunflower trypsin inhibitor-1 (Luckett et al. 1999), SFT-L1 or SFTI-Like 1 (Mylne et al. 2011), and PDPs or PawS-derived peptides (Elliott et al. 2014). Members of both groups belong to the subclass homopolycyclopeptides type VIII of the plant cyclopeptides (Tan and Zhou 2006). Group 3 consists of cyclic peptides with no disulfide bonds referred to as orbitides (Arnison et al. 2013). Members of this group belong to the subclass homomonocyclopeptides type VI of the plant cyclopeptides (Tan and Zhou 2006).

All of the aforementioned cyclic peptides are gene-encoded and occur in phylogenetically distant families of angiosperms. Group 1 has been found in Rubiaceae, Violaceae, Cucurbitaceae, Fabaceae, and Solanaceae (Craik 2013); Group 2 in Asteraceae (Elliott et al. 2014); and Group 3 in Annonaceae, Caryophyllaceae, Euphorbiaceae, Lamiaceae, Linaceae, Phytolaccaceae, Rutaceae, Schizandraceae, and Verbenaceae (Arnison et al. 2013). Despite being phylogenetically distant, the biosynthesis of the first two groups appears to have evolved in parallel, being channeled through transpeptidation by asparaginyl endopeptidase (AEP) that joins their ends (Saska et al. 2007; Gillon et al. 2008; Mylne et al. 2011, 2012). This AEP-mediated cyclization requires a conserved proto-N-terminal Gly, proto-C-terminal Asx (i.e., Asn or Asp), small residue at P1′, and Xle (i.e., Leu or Ile) at P2′ (Mylne et al. 2011, 2012). On the other hand, cyclization of the third group is mediated by peptide cyclase (PCY1), a serine protease-like enzyme (Barber et al. 2013).

The seeds of *Momordica cochinchinensis* (Cucurbitaceae) contain cyclic peptides that belong to the cyclotide group. MCoTI-I and -II (*Momordica cochinchinensis* trypsin inhibitor-I and -II) were the first members of *Momordica* cyclic peptides to be discovered (Hernandez et al. 2000) and have been studied extensively, particularly for applications in the biomedical field. Interest in these peptides stems from the proteolytic stability conferred by their structural motif (Colgrave and Craik 2004) and the amenability of the residues comprising their intercysteine loops to mutation (Austin et al. 2009). Thus, in principle *Momordica* cyclic peptides can be used as highly stable grafting scaffolds. As both peptides have the ability to enter cells (Greenwood et al. 2007; Cascales et al. 2011; Contreras et al. 2011; D'Souza et al. 2014), they have been touted as potential vectors for the delivery of grafted epitopes with desired activities to intracellular targets (Ji et al. 2013). Examples of grafting applications include engineering of 1) MCoTI-I into an anti-HIV agent (Aboye et al. 2012) and an antagonist of intracellular proteins Hdm2 and HdmX for

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

© The Author 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

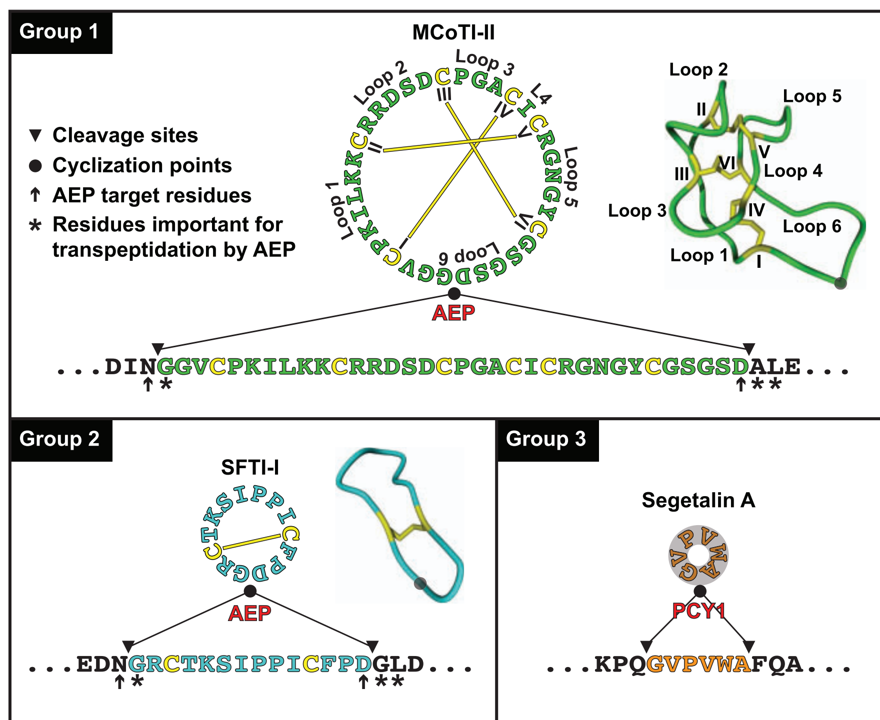

FIG. 1. Proposed classification for currently known backbone cyclized peptides in angiosperms. The cyclic peptide domains (colored letters) with neighboring residues are given below the structures. Group 1: Cyclic peptides with three disulfide bonds that form a knot, that is, Cys III–VI threading a ring formed by Cys I–IV, Cys II–V, and their interconnecting backbone. This structural motif is known as a cyclic cystine knot or CCK. The intercysteine residues that form the loops are labeled Loops 1–6. An example member of the group, that is, MCoTI-II, is shown. Group 2: Cyclic peptides with one disulfide bond, for example, SFTI-1 (PDB#1JBL). Cysteine connectivities are shown as yellow lines. AEP target residues, that is, Asp/Asn, are shown with arrows. Neighboring residues (P1′ and P2′ sites) important for transpeptidation by AEP are shown with asterisks. Group 3: Cyclic peptides with no disulfide bond, that is, the orbitides, for example, segetalin A. Cleavage sites are shown with triangles. The peptide cyclization points are shown with black dots. (For interpretation of the references to color, please refer to the web version of this article.).

suppressing tumor growth (Ji et al. 2013) and 2) MCoTI-II into a β-tryptase inhibitor (Thongyoo et al. 2009; Sommerhoff et al. 2010) and human leukocyte elastase inhibitor (Thongyoo et al. 2009) for inflammatory disorders, and a pro-angiogenic agent for wound healing (Chan et al. 2011). Despite these remarkable successes, the introduction of new activities onto the *Momordica* cyclic peptide scaffold remains challenging because the limiting structural and functional constraints are not yet fully understood. During the course of evolution, negative selection purges mutations that have deleterious effects to the structure of peptides, constraining their ability to acquire new function, which would otherwise be fixed under positive selection (Tokuriki and Tawfik 2009). Thus, knowledge of residues under selection should provide insights into the limitations of *Momordica* cyclic peptides to be engineered as scaffolds.

To understand the evolution of *Momordica* cyclic peptides, it is imperative that their distribution and diversity be traced and mapped. *Momordica* is a clade of approximately 60 tropical and subtropical climbers and creepers that diverged from a common ancestor around 35 Ma (Schaefer and Renner 2010) and underwent long-distance dispersal across Africa, Asia, and Australia (Schaefer et al. 2009). Being a crucial agent for plant dispersal, seeds play a key role in the speciation of plants, carrying within them genetic information for the establishment of new plants under spatiotemporally disparate environmental pressures. Tracing the distribution of cyclic peptides and their acyclic counterparts in *Momordica* species will allow us to determine when, where, and how the genes for their biosynthesis evolved. In turn, this knowledge may serve as a basis for selecting *Momordica* species to further investigate how the cyclic peptides arise, for example, through comparative transcriptome analysis to identify gene sequences and enzymes essential for their processing. Furthermore, mapping the diversity of cyclic peptides in *Momordica* seeds will allow us to identify site-specific residues that are evolving under selection, that is, negative selection to maintain the existing structure whereas positive selection to adopt new function. This information may be particularly useful in the context of designing inhibitors of evolutionarily related serine proteases implicated in human diseases using the *Momordica* cyclic peptide scaffold.

In this study, we describe the distribution and diversity of cyclic peptides and their acyclic counterparts in the seeds of 24 *Momordica* species and an outgroup species *Siraitia grosvenorii* (fig. 2). We discover new *TIPTOP* (*Two Inhibitor Peptide TOPologies*) genes, which encode multiple cyclic peptide domains and terminate with an acyclic peptide domain (Mylne et al. 2012), and partial gene sequences, which we refer to as *TIPRE* (*Tandem Inhibitor Peptide REpeats*), that encode multiple acyclic peptide domains. We assemble transcripts

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

FIG. 2. *Momordica* seeds used in this study. The seeds are arranged from left to right, top to bottom, based on their phylogenetic relationship. *Siraitia grosvenorii* was included as a closely related outgroup species.

that encode a single acyclic peptide domain like the previously reported TGTI-II (towel gourd trypsin inhibitor-II) cDNA (Ling et al. 1993; Mylne et al. 2012). In addition, we identify diagnostic peptides that correspond to the cyclic peptides and their acyclic counterparts.

Despite having undergone long-distance dispersal events during the speciation of *Momordica* (Schaefer et al. 2009; Schaefer and Renner 2010), we found the sequence diversity of *Momordica* cyclic peptides to be low compared with other members of the cyclotide group (Kaas and Craik 2010). This conservation could be explained by the recentness of the event(s) that created the cyclic peptides or by the selection operating on the cyclic peptide domain repeats for maintaining their structure, and thus their documented function as potent trypsin inhibitors (Avrutina et al. 2005). As hydrogen bond networks are known to play an important role in protein recognition (Lu et al. 1997), constraining mutations that could compromise the functional fold conferred by the hydrogen bond networks is vital. Molecular dynamics of *Momordica* cyclic peptides in complex with trypsin reveal alterations in the intermolecular hydrogen bond network upon single amino acid substitutions, highlighting sites that have the potential to be engineered for selective target binding.

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

Results

In this study, we traced the occurrence of cyclic peptides in the seeds of 24 *Momordica* species and an outgroup species *S. grosvenorii*, mapped the residues under selection, and examined the effect of single amino acid substitutions to the intermolecular hydrogen bond network of selected naturally occurring *Momordica* cyclic peptides in complex with trypsin.

Precursors of *Momordica* Cyclic and Acyclic Peptides

Polymerase chain reaction (PCR) of *Momordica* genomic DNA using the primers that amplified *TIPTOP* genes from *Momordica cochinchinensis* and *M. sphaeroidea* (Mylne et al. 2012) resulted in new *TIPTOP* genes from two Asian *Momordica*, that is, one from *M. subangulata* (*TIPTOP4*) and two from *M. macrophylla* (*TIPTOP5* and *TIPTOP6*). *TIPTOP4–6*, respectively, encode six, four, and five cyclic peptides, each terminating with an acyclic peptide (the list of the encoded peptides is given in supplementary table S1, Supplementary Material online, and a representation of the precursors is given in fig. 3*A*). Two of the encoded cyclic peptides, that is, MCoTI-II and MCoTI-IV (hereafter, we remove the MCo prefix because some of the peptides are also present in other *Momordica* species and use an Arabic numeral for simplicity, e.g., MCoTI-II becomes TI-2), have been reported (Hernandez et al. 2000; Mylne et al. 2012) whereas the others are new but have similar sequences that share the Asp–Gly cyclization point. Similarly, two of the encoded acyclic peptides, that is, TI-5 and TI-6, have been reported (Mylne et al. 2012). The other acyclic peptide, that is, TI-19, differs from TI-5 in that it has an additional N-terminal Gln. An alignment of *Momordica* cyclic peptides and their acyclic counterparts is given in figure 3*B*.

A new set of primers for conserved sequences within the endoplasmic reticulum (ER) signal and the acyclic peptide domain was designed because the first primer set could not amplify *TIPTOP* genes from the remaining *Momordica* genomic DNA. PCR with this new set of primers resulted in five partial gene sequences that appear to have undergone expansion similar to *TIPTOP* and thus were named *TIPRE*, for *Tandem Inhibitor Peptide REpeats*. Four of the partial gene sequences were found in the African *M. anigosantha* (*TIPRE1–4*) and one was found in the African *M. friesiorum* (*TIPRE5*). The list of the encoded peptides is given in supplementary table S1, Supplementary Material online. The TIPRE peptides, that is, TI-24–27, have similar sequences to the TIPTOP acyclic peptides but with an additional four C-terminal residues like the TIPTOP cyclic peptides (fig. 3*B*). This finding suggests that the acyclic peptides acquired features for cyclization following extension of their tail, which provided the target residues for AEP to then perform transpeptidation that ligates their C-terminus to their N-terminus.

Analysis of a previously reported African *M. charantia* seed transcriptome (Yang et al. 2010) revealed five transcripts (the translation of the transcripts is given in supplementary table S2, Supplementary Material online), each encoding a single acyclic peptide domain. Two of the transcripts encode the acyclic peptides MCTI-I (*Momordica charantia* trypsin inhibitor-I) and MCTI-III (Hara et al. 1989; Hamato et al. 1995). One transcript encodes an acyclic peptide, which we refer to as TI-28, that is similar to MCTI-II (Hara et al. 1989) but with an extended N-terminus. Another transcript encodes an acyclic peptide, which we refer to as EI-1 (elastase inhibitor-1), that is similar to MCEI-IV (*Momordica charantia* elastase inhibitor-IV [Hamato et al. 1995]) but differs in one residue following the N-terminal Glu. The absence of a dedicated precursor for MCTI-II and MCEI-I–III (supplementary table S2, Supplementary Material online), which, respectively, are shorter than TI-28 and EI-1 in their N-terminus (fig. 3*B*), suggests that they are products of posttranslational N-terminal trimming, a process that has been proposed to give rise to the acyclic peptides hedyotide B2–4 from the Rubiaceae *Hedyotis biflora* (Nguyen et al. 2011). On the other hand, the other transcript potentially encodes a new peptide, which we refer to as TI-23, as judged by the sequence similarity to the other *Momordica* cyclic peptides. Given the lack of features for AEP processing in its precursor (supplementary table S2, Supplementary Material online), it would be interesting to confirm the presence of TI-23.

Identification of Peptides Using a Targeted Proteomics Approach

A targeted search for *Momordica* cyclic peptides and their acyclic counterparts was aided by the observation of tandem mass spectrometry (MS/MS) for diagnostic peptides, which results from the digestion of reduced and alkylated peptides. For cyclic peptides, the diagnostic peptides are chymotrypsin digests harboring sequence tags that extend over their cyclization points, that is, Cys_{29} to Leu_8 (for residue numbering, please refer to fig. 3*B*). For acyclic peptides, the diagnostic peptides result from trypsin, chymotrypsin, or endoproteinase Glu-C digestion. A list of the sequences of representative diagnostic peptides found is given in supplementary table S3, Supplementary Material online. The distribution of *Momordica* cyclic peptides and their acyclic counterparts is presented in figure 4.

MS/MS evidence for cyclic peptides was only found in the Asian *M. cochinchinensis*, *M. macrophylla*, *M. denticulata*, *M. subangulata*, and *M. clarkeana* and in a close relative African *M. gilgiana* (a representative MS/MS spectrum is given in supplementary fig. S1*A*, Supplementary Material online). No evidence was found for TI-23 in the African *M. charantia*. For the African *M. anigosantha* TIPRE peptides, evidence was only found to support acyclic peptides, as would be expected judging from the lack of a proto-N-terminal Gly. Observed MS/MS spectra were consistent with a chymotrypsin and an endoproteinase Glu-C digest product that correspond to acyclic peptides having an N-terminal pyrolated glutamine (supplementary fig. S1*B*, Supplementary Material online) and a four C-terminal residue extension (supplementary fig. S1*C*, Supplementary Material online). Evidence for acyclic trypsin inhibitors was found in all of the species analyzed (supplementary table S3, Supplementary Material online, representative MS/MS spectra are given in supplementary fig. S1*D* and *E*, Supplementary Material online). For acyclic

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

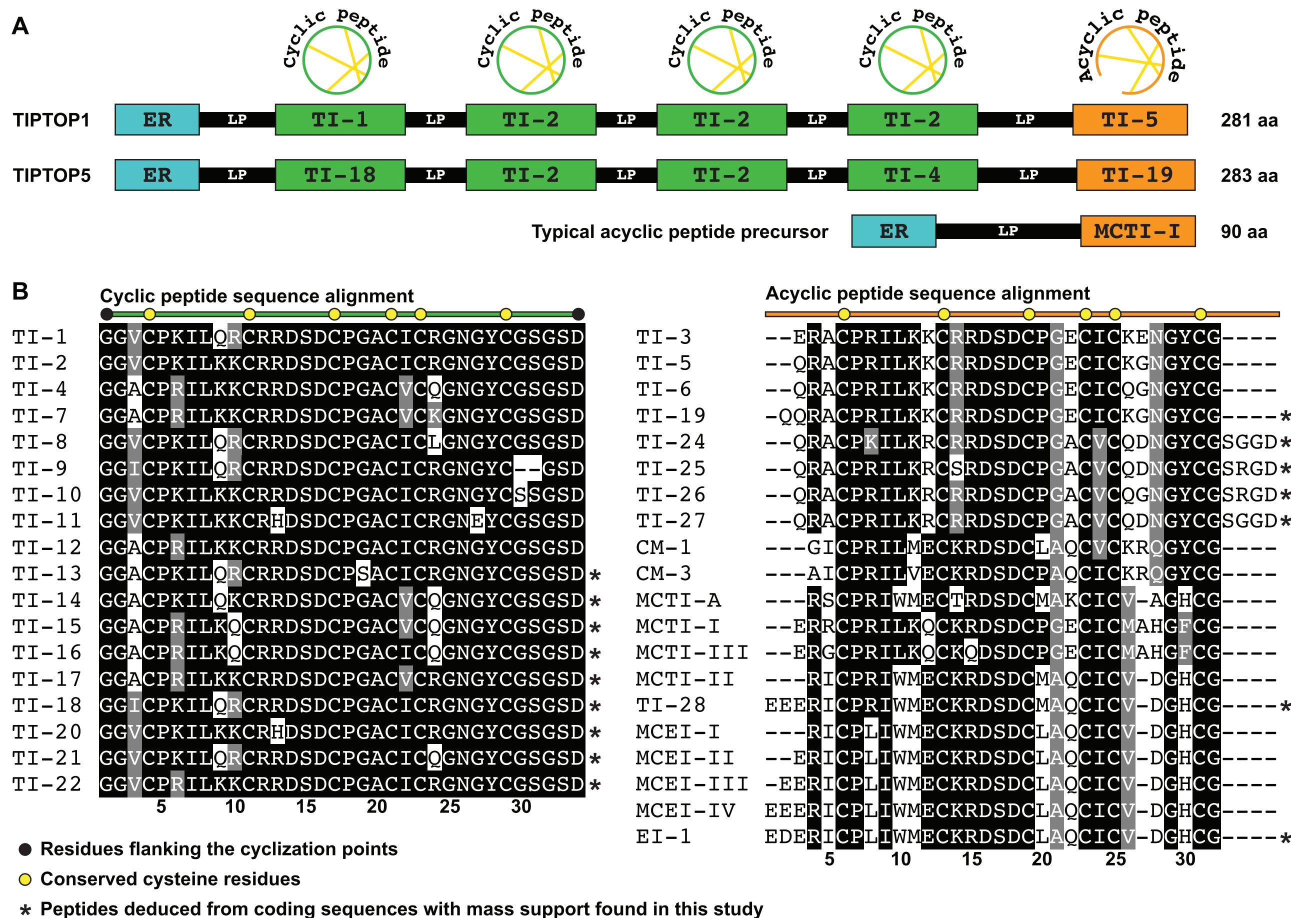

Fig. 3. Precursors and sequences of *Momordica* cyclic peptides and their acyclic counterparts. (*A*) Schematic representation of the precursors for TIPTOP1 and TIPTOP5 is shown along with a typical precursor of an acyclic peptide from *Momordica charantia*. ER, endoplasmic reticulum signal; LP, leader peptide domain—this naming follows the recommended nomenclature for ribosomally synthesized and posttranslationally modified peptides (Arnison et al. 2013); aa, amino acids. (*B*) Sequence alignment of *Momordica* cyclic peptides and their acyclic counterparts. Peptides deduced from coding sequences with mass support found in this study (asterisks) are aligned with previously reported peptides (Joubert 1984; Hara et al. 1989; Hamato et al. 1995; Hernandez et al. 2000; Mylne et al. 2012). Residues flanking the cyclization points are shown with black dots. The six conserved cysteine residues are shown with yellow dots. Residues are numbered from the N-terminus to the C-terminus. (For interpretation of the references to color, please refer to the web version of this article.).

elastase inhibitors, evidence was only found in the African *M. leiocarpa*, *M. foetida*, *M. balsamina*, and *M. charantia* (representative MS/MS spectra are given in supplementary fig. S1*F* and *G*, Supplementary Material online).

Mapping of Sites under Selection

To map the sites that are evolving under selection, we analyzed the number of synonymous substitution per site (dS) and the number of nonsynonymous substitution per site (d*N*) of the cyclic peptides. The value of the d*N* − dS indicates whether a particular site is evolving under negative (if the value is negative) or positive (if the value is positive) selection. The d*N*–dS analysis of the cyclic peptides (supplementary table S4, Supplementary Material online) revealed that 2 cysteines and 11 intercysteine residues are evolving under negative selection, which include four of five residues in Loop 2, one of three residues in Loop 3, one of five residues in Loop 5, and five of eight residues in Loop 6. On the other hand, eight intercysteine residues are evolving under positive selection, which include three of six residues in Loop 1, one of three residues in Loop 3, one of one residue in Loop 4, two of five residues in Loop 5, and one of eight residues in Loop 6. The remaining 13 residues (four cysteines, three residues in Loop 1, one residue in Loop 2, one residue in Loop 3, two residues in Loop 5, and two residues in Loop 6) are neutral. This finding highlights the types of selection operating on the sites of the *Momordica* cyclic peptide scaffold (fig. 5*A*).

Dynamics of the Hydrogen Bond Network of Selected Cyclic Peptides with Trypsin

Mutations provide an essential raw material for evolution and serve as a basis for acquiring new functions (Tokuriki and Tawfik 2009). As hydrogen bonds play a role in the recognition of inhibitors against their target proteins (Lu et al. 1997), examining the effect of mutations on the interaction of inhibitors against their target proteins is fundamental to understanding the mechanistic of their biological function. Here, we examined the effect of single amino acid substitutions to the intermolecular hydrogen bonds of the *Momordica* cyclic peptides TI-1 and TI-2 in complex with

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

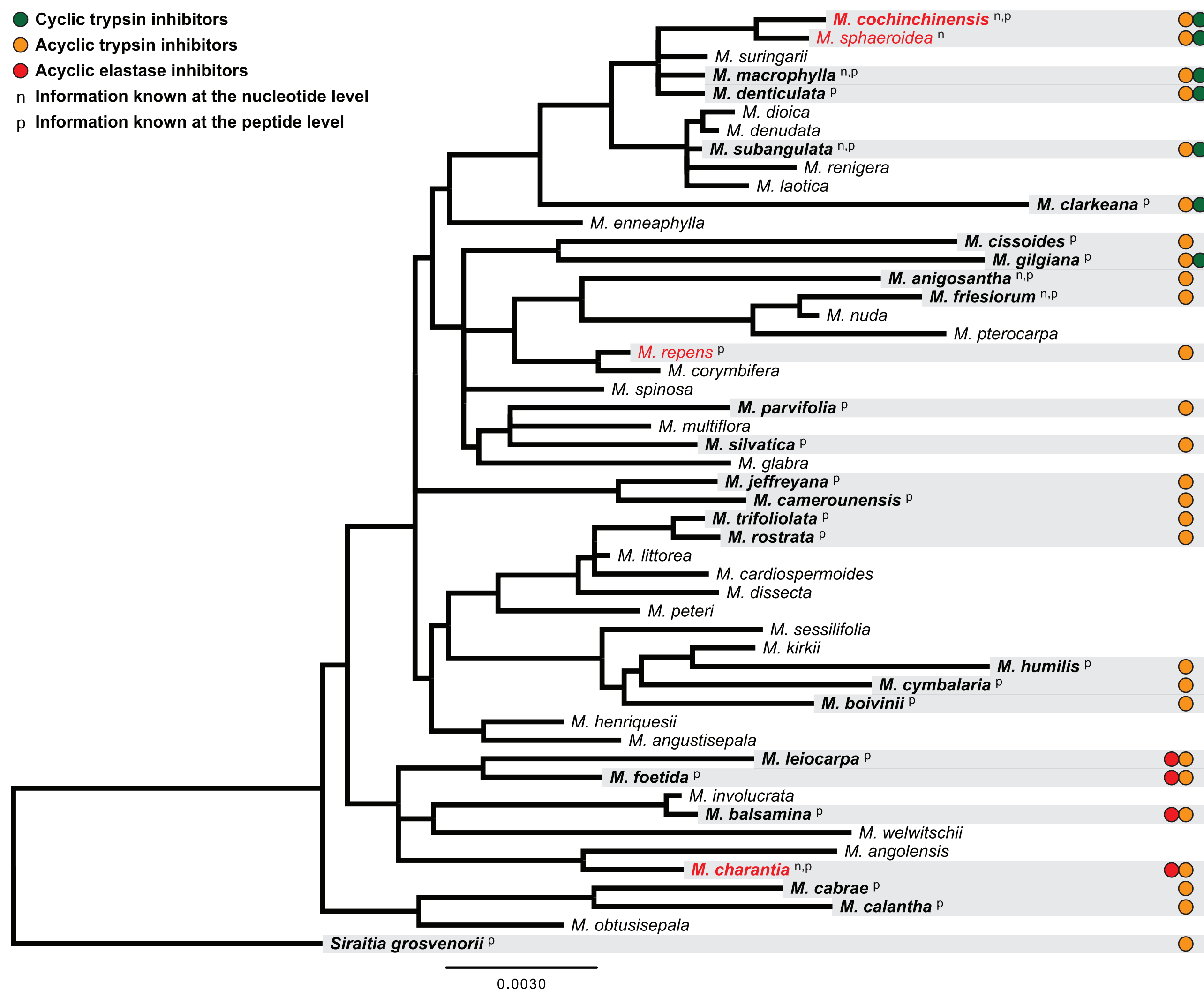

Fig. 4. Distribution of *Momordica* cyclic peptides and their acyclic counterparts mapped on a maximum-likelihood phylogeny estimate of the genus. Species analyzed in this study are shown with bold letters. Species previously studied are shown with red letters. Species containing cyclic peptides and their acyclic counterparts are shown with colored dots, that is, green for cyclic trypsin inhibitors, orange for acyclic trypsin inhibitors, and red for acyclic elastase inhibitors. Superscript letters following the species names denote currently known information at the nucleotide (n) or peptide (p) level. (For interpretation of the references to color, please refer to the web version of this article.).

trypsin using molecular dynamics. The mutations made to TI-1 were based on the sequence of TI-8, TI-18 and TI-21, whereas those in TI-2 were based on the sequence of TI-10, TI-20, and TI-22.

Analysis of the dynamics of the cyclic peptides in complex with trypsin revealed site-specific residues that play a role in forming the intermolecular hydrogen bond network (fig. 5*B*; list of donors, acceptors, and frequency of occupancy of the hydrogen bonds is given in supplementary table S5, Supplementary Material online). One of the prominent features of *Momordica* cyclic peptides is that their main chains, that is, of Loops 1 and 6, form the majority of the hydrogen bond network with trypsin. For TI-1, the residues involved in main chain hydrogen bonding with trypsin are $Gly_{1,2,32}$, $Cys_{4,29}$, Pro_5, Lys_6, Ile_7, Leu_8, and Asp_{34} whereas the residues involved in side chain hydrogen bonding are Lys_6, Gln_9, Arg_{24}, Asn_{26}, and Ser_{31}. For TI-2, the residues involved in main chain and side chain hydrogen bonding with trypsin are similar to TI-1, with the exception of Lys_{10} instead of Gly_{32} for the main chain and Tyr_{28} instead of Gln_9 for the side chain. Single amino acid substitutions in both cyclic peptides altered their hydrogen bond network with trypsin, notably are the introduction or abolishment of and the increase or decrease in frequency of occupancy of a number of the main chain and side chain hydrogen bonds (fig. 5*B*). This result highlights the dynamics of the hydrogen bond network of *Momordica* cyclic peptides in complex with trypsin, providing insights into the potential role of site-specific residues for target binding.

Discussion

Momordica Cyclic Peptides Occur in Species that Diverged from an African Ancestor around 19 Ma

TIPTOP genes were found following the report of the cyclic peptides MCoTI-I and MCoTI-II from *M. cochinchinensis* seeds (Hernandez et al. 2000). The unusual nature of how the genes are organized, that is, having multiple repeats of cyclic peptide domains that terminate with an acyclic peptide

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

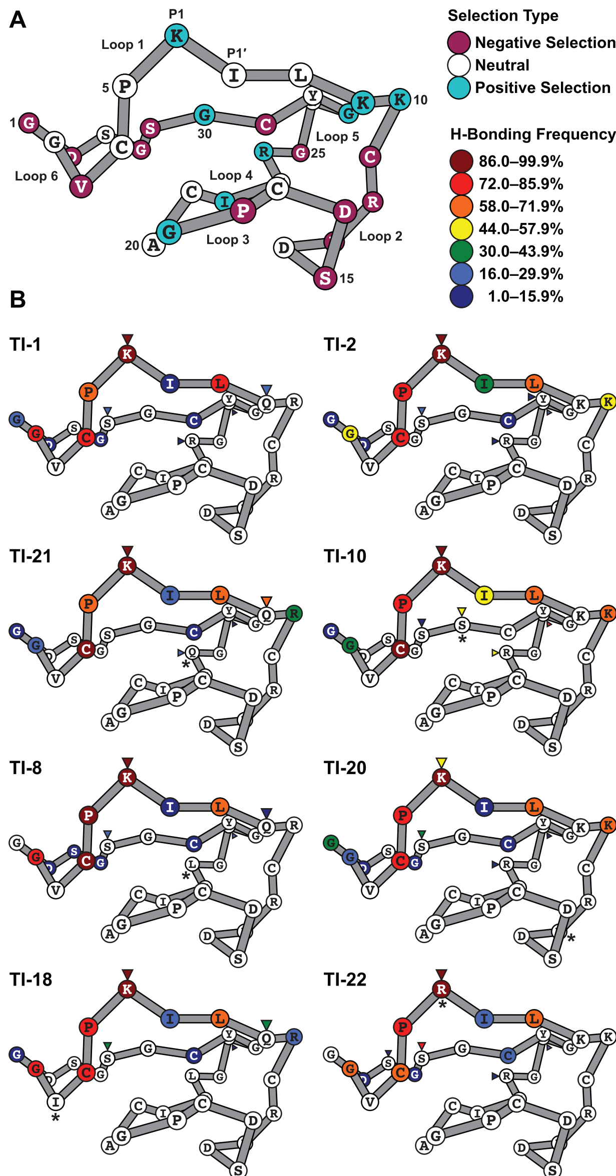

Fig. 5. Site selection and intermolecular hydrogen bond network mapping of *Momordica* cyclic peptides. (*A*) Sites under selection: Negative (purple), neutral (white), and positive (cyan). The sequence of TI-2 is used as a reference. The intercysteine residues that form the loops are labeled Loops 1–6. Sites are numbered from the proto-N-terminal Gly to the proto-C-terminal Asp. P1 and P1′, respectively, denote residues on the acyl and leaving group side of the peptide bond that would be hydrolyzed by trypsin. (*B*) Sites involved in hydrogen bonding with trypsin. TI-1 and TI-2 were used as references for single amino acid substitutions (shown with asterisks) based on the sequences of TI-8, TI-18 and TI-21 for TI-1; and TI-10, TI-20 and TI-22 for TI-2. Intermolecular main chain and side chain hydrogen bonds are shown with colored circles and triangles, respectively. The range of hydrogen bond frequency of occupancy is color-coded as per legend. (For interpretation of the references to color, please refer to the web version of this article.).

domain, led to the hypothesis that they might have expanded from an ancestral gene through internal duplication event(s) (Mylne et al. 2012). Tracing the distribution of TIPTOP cyclic peptides has provided insights into when and where they emerged.

A targeted search revealed the occurrence of cyclic peptides in the Asian *M. cochinchinensis*, *M. macrophylla*, *M. denticulata*, *M. subangulata*, and *M. clarkeana* and in a close relative African *M. gilgiana* (fig. 4). The lack of evidence for cyclic peptides in other representative African taxa suggests that the cyclic peptides have arisen within species that descended from a common ancestor to the Asian and a close relative African species. The Asian *Momordica* are a result of a long-distance dispersal of an African ancestor that came back to Asia around 19 Ma—an event that marks the divergence of the Asian species from their close relative African *M. gilgiana* (Schaefer and Renner 2010). Thus, the ancestral gene of *TIPTOP* presumably has been inherited from this African ancestor. Interestingly, the expansion that created *TIPTOP* genes appears to have occurred recently and independently, as suggested by the highly conserved signal peptides (fig. 6*A*), which are known to evolve rapidly (Li et al. 2009), and the distinct number of domain repeats and sequence of the encoded peptides. This scenario would require specific selective pressures operating on both the Asian species and the close relative African *M. gilgiana*.

Plausible Selective Advantage and Pressure Underlying the Expansion that Created *TIPTOP* Genes

The selective advantage conferred by the expansion that created *TIPTOP* genes might be related to tight regulatory control for expression of the encoded peptides. The repetitive nature of *TIPTOP* genes would allow the expression of multiple peptides from one transcript. On the other hand, the expansion would alter the RNA secondary structure, which is known to be one of the key determinants to posttranscriptional regulation in plants (Silverman et al. 2013). A recent study using the model plant *Arabidopsis thaliana* showed that *LOW MOLECULAR WEIGHT CYSTEINE-RICH*-encoding mRNA is among the highly structured mRNA that tends to be degraded more frequently than less structured mRNA (Li et al. 2012). Calculation of the folding energy of *TIPTOP* transcripts using Mfold (Zuker 2003) suggests that the expansion decreases the free energy for folding of *TIPTOP* transcripts into their secondary structures (fig. 6*B*). Taken together, the expansion might have allowed the encoded peptides to be produced efficiently but not excessively—a trait that fits well with a defense response and storage function.

Resistance to invaders is among the most ancient traits that evolved through discriminating self from nonself (Staal and Dixelius 2007). Having a biological activity as potent inhibitors of trypsin (Avrutina et al. 2005), one of the main digestive enzymes of invaders, TIPTOP peptides may be regarded as antinutritive agents. Given that many of the known seed-derived inhibitors are only active against digestive enzymes of insects but not against endogenous enzymes (Shewry and Casey 1999), it is tempting to speculate that the expansion that created *TIPTOP* genes might have been triggered by predatory cues. On the other hand, the high cysteine content of TIPTOP peptides suggests that they might also serve a dual function for storage purposes, providing sulfur along with nitrogen and carbon for germination

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

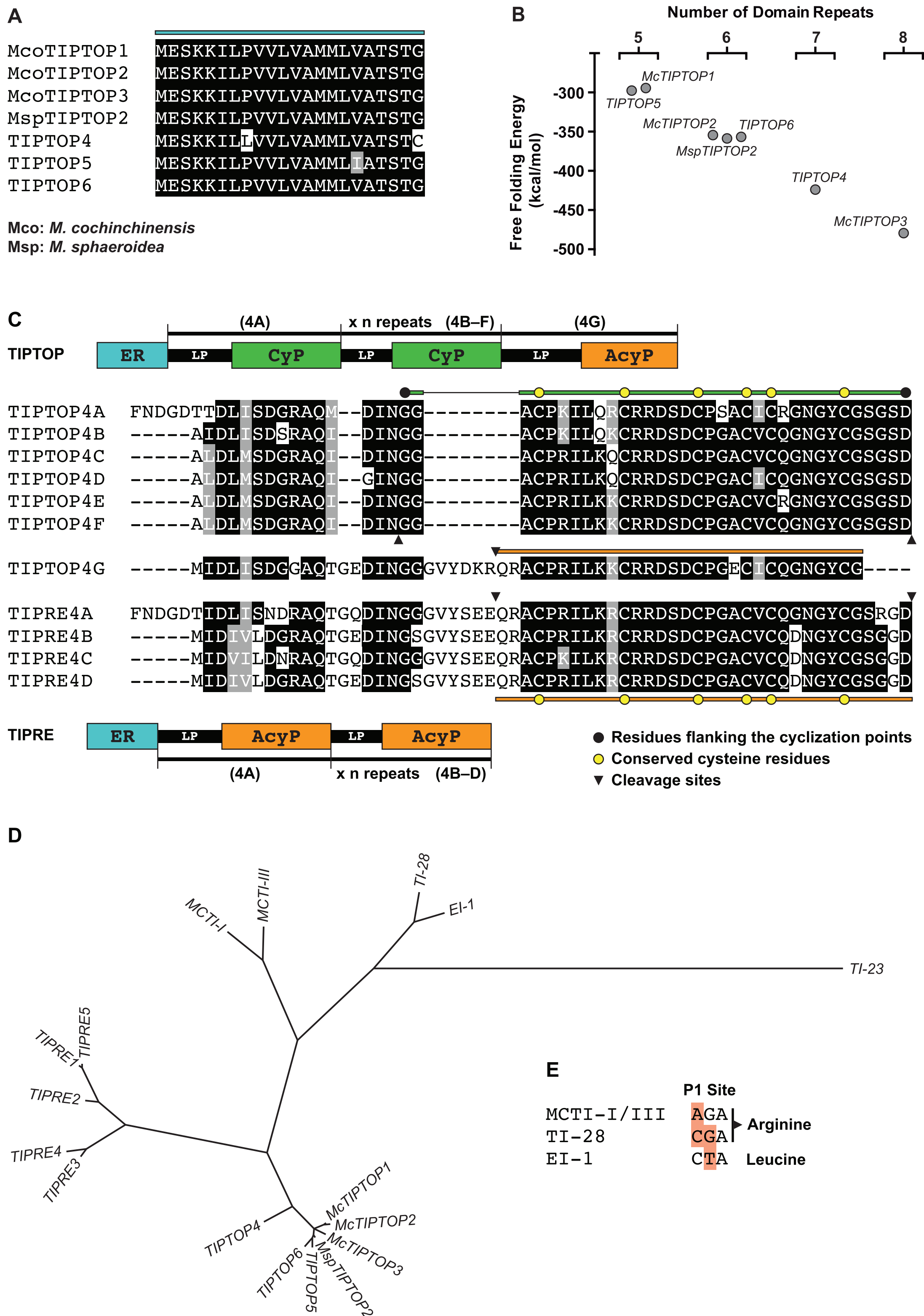

Fig. 6. Sequence analysis for *Momordica* peptides. (*A*) Alignment of the signal peptide sequence of TIPTOP precursors. (*B*) Free folding energy of *TIPTOP* transcripts calculated using Mfold (Zuker 2003). (*C*) Alignment of domain repeats. Used as an example here is the leader and mature peptide domains of TIPTOP4 from the Asian *Momordica subangulata* and TIPRE4 from the African *M. anigosantha*. The cyclic peptide domains (shown in green, TIPTOP4A–F) appear to be the result of insertion of a four-residue segment trailing the C-terminal and deletion of a segment within the N-terminal of the acyclic peptide domain (shown in orange, TIPTOP4G). TIPRE acyclic peptides acquired an extension of the C-terminal like the TIPTOP cyclic peptides, thus link the evolution of cyclic peptides from their acyclic counterparts. Residues flanking the cyclization points are shown with black dots. The six conserved cysteine residues are shown with yellow dots. Cleavage sites are shown with triangles. (*D*) Unrooted phylogram for *TIPTOP* genes, *TIPRE* sequences, and the transcripts from the African *M. charantia* seeds. (*E*) Mutations at P1 site of the African *M. charantia* peptides that led to the change from Arg to Leu. (For interpretation of the references to color, please refer to the web version of this article.).

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

and seedling growth. The exceptional stability of the cyclic cystine knot class of peptides (Colgrave and Craik 2004) might be related to their function as long-term storage proteins, supporting extended periods of dormancy as most Cucurbitaceae have orthodox seeds—they can tolerate considerable desiccation and thus have greater longevity compared with recalcitrant seeds (Ellis 1991). Further studies would be needed to test this dual function hypothesis.

Mutations in the Acyclic Peptide Precursors: The Link to Cyclization

The emergence of backbone cyclized peptides of Groups 1 and 2 (fig. 1) has been shown to be mediated by AEP (Saska et al. 2007; Gillon et al. 2008; Mylne et al. 2011, 2012), a vacuolar processing enzyme that cleaves Asn and Asp (Hara-Nishimura et al. 1991; Hiraiwa et al. 1999), which, respectively, precedes and ends the TIPTOP cyclic peptide domains (fig. 6C). The residues trailing the proto-C-terminal are usually a small residue at P1′ and Xle at P2′. Cyclization of the peptide backbone occurs through a transpeptidation that critically requires the presence of a proto-N-terminal Gly, which is thought to lack the steric hindrance of other side chains thus enabling transpeptidation to occur (Mylne et al. 2011). The absence of these features in the precursors of the acyclic peptides means that transpeptidation by AEP will not occur, thus AEP acts as the constraining evolutionary channel for cyclization (Mylne et al. 2012).

Alignment of the leader and mature peptide repeats of TIPTOP4 from the Asian *M. subangulata* and TIPRE4 from the African *M. anigosantha* (fig. 6C) suggests that a series of mutations following internal gene duplication provided the features for AEP to perform transpeptidation. We hypothesize that the acyclic peptides first acquired an extension of their C-terminal (by insertion of the SXXD segment) followed by deletion of their N-terminal region (GVYXXXQR segment) and mutation of their P1′ proto-C-terminal trailing residue (Met to a small residue, in this case Ala), which would then lead to the predisposition of the precursors to cyclization by AEP. With an N-terminal Gln like the TIPTOP acyclic peptides and an extended C-terminus like the TIPTOP cyclic peptides, the TIPRE peptides may be regarded as "intermediates" in the evolution of TIPTOP cyclic peptides from their acyclic counterparts. This hypothesis is consistent with the phylogenetic analysis that places the African *M. anigosantha* close to the species in which TIPTOP cyclic peptides occur (fig. 4; Schaefer and Renner 2010).

The absence of evidence for the putative cyclic peptide TI-23 in the African *M. charantia* might be due to the lack of common features for AEP processing in its precursor, which only harbors a single peptide domain and does not have residues trailing the proto-C-terminal Asn and Asx preceding the proto-N-terminal Gly (supplementary table S2, Supplementary Material online). Thus, internal gene duplication may be considered as a steppingstone to the acquisition of features for transpeptidation by AEP, which the *TIPTOP* genes then acquired through a series of mutations during their course of evolution. Molecular phylogenetic analysis reveals that the transcript encoding TI-23 is distantly related to the other coding sequences (fig. 6D), suggesting that it might be an ancestral vestige of a kindred evolutionary process that, in the Asian and a close relative African species, created the cyclic peptides.

Neofunctionalization of an Acyclic Peptide in the African *Momordica* Species

Gene duplication has been considered to be the main source of material for the emergence of new functions. The rate at which a gene duplication event occurs is considered high, with the duplicates being silenced within a few million years and the survivors being selected under strong negative pressure (Lynch and Conery 2000). The high sequence similarity of the transcripts encoding MCTI-I and MCTI-III and the transcripts encoding TI-28 and EI-1 (supplementary table S2, Supplementary Material online) suggests that they have arisen through gene duplication. In the case of the transcripts encoding TI-28 and EI-1, the acyclic peptides have different biochemical activities, that is, the former a trypsin inhibitor whereas the latter an elastase inhibitor. Because both of the acyclic peptides are expressed within the seeds of the African *M. charantia*, the duplicated gene may be considered to have undergone neofunctionalization, that is, it acquired a new function that is preserved by natural selection.

Sequence analysis reveals that the new function emerged from a point mutation at the second codon base of the P1 site—the site that interacts with the S1 pocket (or active site) of the target enzyme. The G of the CGA that encodes Arg, which is the preferred P1 residue for trypsin (Krieger et al. 1974), is replaced by T and thus changing it into Leu (fig. 6E), which is preferred for elastase (Hara et al. 1989), if assuming that elastase inhibitor is the new function. Interestingly, the acyclic elastase inhibitors were only found in the African *M. charantia* and three closely related African species, that is, *M. balsamina*, *M. leiocarpa*, and *M. foetida*, which diverged around 21 Ma (Schaefer and Renner 2010). This finding suggests that neofunctionalization is a rare evolutionary fate for this class of peptides.

The Role of Site-Specific Residues of *Momordica* Cyclic Peptides

A range of studies have shown remarkable success in the use of *Momordica* cyclic peptides as scaffolds for developing novel therapeutics (Poth et al. 2013). To better exploit *Momordica* cyclic peptides for biomedical applications, it is imperative that their evolvability, that is, the ability to acquire new function through structural changes (Tokuriki and Tawfik 2009), is understood. Given the function of *Momordica* cyclic peptides as potent trypsin inhibitors, knowledge of their evolvability is particularly useful for developing novel inhibitors of other evolutionarily related serine proteases—many of which play crucial roles in pathophysiological processes, such as inflammation and blood clotting (Bachovchin and Cravatt 2012).

One of the approaches that can be used to understand the evolvability of *Momordica* cyclic peptides is by mapping their

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

sequence diversity, which has been shaped by natural selection. Evolutionary analysis reveals the type of selection operating on the sites of *Momordica* cyclic peptides. As shown in figure 5*A*, the majority of the sites are either neutral (no substitution) or under negative selection (substitutions were synonymous). Residues that occupy these sites are presumably preserved for maintaining the cyclic cystine knot structure. Indeed, four of the six cysteines that form the cystine knot core are neutral and two are under negative selection. The N-terminal Gly and C-terminal Asp that have been shown to be vital for cyclization by AEP (Mylne et al. 2011) are also under negative selection. The significance of preserving site-specific residues in Loop 6 might be related to the effect that cyclization has on the folding pathway of the peptides, facilitating the formation of the correct cysteine connectivities, thus presumably reducing the entropic losses upon folding compared with their acyclic counterparts (Daly et al. 1999).

Residues that occupy sites under positive selection are presumably paving the way to adopting a new function. The P1 site that defines the selectivity of *Momordica* cyclic peptides is evolving under positive selection (fig. 5*A*) but the amino acid change has not introduced a new function, that is, Lys and Arg are both preferred in the S1 pocket of trypsin (Krieger et al. 1974). The neutral selection operating on P1′ site, which is occupied by Ile—a hydrophobic residue thus preferred by trypsin (Kurth et al. 1997), suggests that *Momordica* cyclic peptides are coevolving with trypsin, which has strictly conserved residues associated with its specificity in both prokaryotes and eukaryotes (Rypniewski et al. 1994). Because the three-dimensional structure of serine proteases is highly conserved and their active sites are virtually the same, that is, having the catalytic triad composed of Asp–His–Ser (Higaki et al. 1987), molecular dynamics of *Momordica* cyclic peptides with trypsin may serve as a model for studying the role of site-specific residues for target binding, particularly that imparted by hydrogen bonding.

As shown in figure 5*B*, the hydrogen bond network at the interface of *Momordica* cyclic peptides with trypsin is primarily formed by the main chain of the cyclic peptides, with P1, P2, P3, P5, and P2′ sites having a high frequency of occupancy of hydrogen bond. Although this hydrogen bond network is common in serine protease–inhibitor complexes where the inhibitory loops are locked in an extended antiparallel β-sheet conformation (Hedstrom 2002), this canonical conformation is not adopted by the inhibitory loop of the *Momordica* cyclic peptide scaffold (Daly et al. 2013). The occupancy of Pro at P2 and Cys at P3, which do not contribute to side chain interactions with trypsin, suggests that the selectivity of *Momordica* cyclic peptides is mediated by other sites, for example, their prime sites. This characteristic is unique because S2 and S3 are known to determine the specificity of serine proteases, for example, a hydrophobic residue at P2 is preferred by chymotrypsin (Brady and Abeles 1990) whereas at P3 by elastase (Thompson and Blout 1973). The P3′ site, which is under positive selection, might be important for target binding. Unlike Lys, the occupancy of Gln at P3′ introduces a side chain hydrogen bond and thus may explain the higher trypsin inhibitory activity of TI-1 compared with TI-2 (Avrutina et al. 2005).

As the basis of evolution, mutations play a major role in forming the hydrogen bond network of protease–inhibitor interaction sites. Mutations that introduce new hydrogen bonds at these sites or increase the frequency of occupancy of existing ones are highly desirable. However, mutations can also lead to the contrary and thus examining the effect of mutations to the hydrogen bond network at these interfaces is important. As shown in figure 5*B*, single amino acid substitutions altered the intermolecular main chain and side chain hydrogen bond network of *Momordica* cyclic peptides with trypsin. These alterations may serve as a basis for the identification of sites that have target binding potential, that is, sites 9 in Loop 1; 24, 26, 28 in Loop 5; and 30, 31, 33 in Loop 6. This prediction agrees with a study that showed that the aforementioned sites in Loops 1 and 5 play a role in binding with trypsin (Austin et al. 2009). In the context of drug design, this knowledge may translate into the design of site-specific libraries using the *Momordica* cyclic peptide scaffold. This design approach was successful for engineering kalata B1, the prototypic cyclotide found in Rubiaceae and Violaceae, into an antagonist of neuropilin-1 and -2, which are known to be regulators of vascular and lymphatic development (Getz et al. 2013).

In summary, this study presents evidence to suggest that *Momordica* cyclic peptides evolved in species that diverged from an African ancestor around 19 Ma. The findings provide a basis for selecting species to further investigate the biosynthetic origin of *Momordica* cyclic peptides. Knowledge of the genes that encode cyclic peptides and enzymes involved in their maturation could potentially be used for their production in suitable host plants. In addition, this study showcases an interesting biological example of how natural selection—as imparted by mutations—is presumably operating to acquire features essential for cyclization of the acyclic peptides by AEP and to fine-tune the selectivity of cyclic peptides while maintaining their structure. This knowledge may find useful application in medicine for designing inhibitors of evolutionarily related serine proteases implicated in human diseases using the *Momordica* cyclic peptide scaffold or in agriculture for designing improved pesticidal agents based on cyclotides (Poth et al. 2011). In the long run, knowledge of the biosynthesis and evolvability of *Momordica* cyclic peptides could translate into the production of "designer" peptide therapeutics in plant seeds.

Materials and Methods

Seed Material

Seeds of *M. anigosantha* Hook.f., *M. boivinii* Baill., *M. cabrae* (Cogn.) C. Jeffrey, *M. calantha* Gilg, *M. camerounensis* Keraudren, *M. cissoides* Planch. ex Benth., *M. clarkeana* King, *M. cymbalaria* Frenzl ex. Naudin, *M. denticulata* Miq., *M. foetida* Schumach., *M. friesiorum* (Harms) C. Jeffrey, *M. gilgiana* Cogn., *M. humilis* (Cogn.) C. Jeffrey, *M. jeffreyana* Keraudren, *M. leiocarpa* Gilg, *M. macrophylla* Gage,

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

M. parvifolia Cogn., *M. rostrata* Zimm., *M. silvatica* Jongkind, *M. subangulata* Blume, *M. trifoliolata* Hook.f., and *S. grosvenorii* (Swingle) C. Jeffrey ex. A.M. Lu & Zhi Y. Zhang were provided by Hanno Schaefer of the Technische Universität München. Seeds of *M. balsamina* L. (reference number: 406803), *M. charantia* L. (reference number: 51359), and *M. cochinchinensis* (Lour.) Spreng. (reference number: 69291) were purchased from B & T World Seeds sarl (Paguignan, Aigues Vives, France).

Genomic DNA Extraction, Gene Cloning, and Sequencing

Momordica seeds were dehusked and finely ground in liquid nitrogen using a mortar and pestle. Genomic DNA was extracted using Qiagen DNEasy Plant Mini Kit following the protocol suggested by the company. *TIPTOP* genes were amplified using primers JM482 (5′-CGT CTT GCT AGA GAA AGG GAG T-3′) and JM483 (5′-TCA GAA ACA GCA TAG CTT TCA C- 3′) (Mylne et al. 2012). *TIPRE* sequences were amplified using primers TM P1 (5′- GAA ATG GAG AGC AAG AAG ATT CT-3′) and TM P5 (5′-AAG ATT CTA GGA CAG GCT CTT TG-3′). PCR products were purified using QIAquick PCR Purification Kit (for single band) and QIAquick Gel Extraction Kit (for multiple bands). The purified PCR products were cloned into pGEM-T Easy (Promega) and sequenced at the Australian Genome Research Facility. A minimum of three independent clones was used to assemble the sequence using MacVector 12.7 software. SignalP 4.1 was used to predict the ER recognition site in the sequence (Petersen et al. 2011). Peptide cleavage sites were predicted based on the common features for processing and homology to previously reported peptides (Mylne et al. 2011, 2012).

Transcriptome Analysis

Transcriptome sequencing data of the African *M. charantia* seeds (Yang et al. 2010) were accessed through the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov, last accessed November 11, 2014). Short Read Archive (SRA) under the accession numbers SRX030203 (normalized sequence data) and SRX030204 (nonnormalized sequence data) was used to assemble transcripts containing plant-derived cystine knot peptide sequences (Gracy et al. 2008) with MIRA 3.4 (Chevreux 2005). A minimum of three overlapping contigs was used to assemble the transcripts.

Peptide Extraction and Fractionation

Peptides were extracted from *Momordica* seeds using a method based on acetonitrile/water/formic acid (25:24:1) as previously described (Mahatmanto et al. 2014). Targeted peptides were fractionated using a 3 cc cartridge, 200 mg sorbent, Waters Sep-Pak C18 55–105 µm. Peptides were eluted with 1.5 ml of solvent B (90% v/v acetonitrile, 0.1% v/v formic acid) in a 10% gradient. Fractions containing the majority of the targeted peptides (supplementary figs. S2 and S3, Supplementary Material online) were collected. Samples were lyophilized, redissolved in 1% v/v formic acid, and stored at 4 °C until further analysis.

Ultra High Performance Liquid Chromatography–MS/MS

Samples were reduced with dithiothreitol (final concentration 10 mM; incubated at 60 °C for 30 min under nitrogen), alkylated with iodoacetamide (final concentration 25 mM; incubated at 37 °C for 30 min in the dark), and split for overnight digestion with trypsin (approximately 1 µg per 100 µg lyophilized sample in 100 mM ammonium bicarbonate, pH 8.0; incubated at 37 °C), chymotrypsin (approximately 2 µg per 100 µg lyophilized sample in 100 mM Tris–HCl, 10 mM calcium chloride, pH 8.0; incubated at 30 °C), and endoproteinase Glu-C (approximately 3 µg per 100 µg lyophilized sample in 1) 100 mM ammonium bicarbonate, pH 8.0 and 2) 100 mM sodium phosphate, pH 7.8; both incubated at 37 °C). Following reduction, alkylation, and digestion, samples were analyzed on a Nexera Ultra High Performance Liquid Chromatography (Shimadzu) coupled to a TripleTOF 5600 mass spectrometer (AB SCIEX) equipped with a duo electrospray ion source. Data were processed using Analyst TF 1.6 software (AB SCIEX). MS/MS spectra were searched against a custom-built database of plant-derived cystine knot peptides using ProteinPilot 4.0 software (AB SCIEX).

Phylogenetic Analysis

The phylogeny of *Momordica* is based on a slightly simplified version of the sequence data set of Schaefer and Renner (2010) with addition of sequences for *S. grosvenorii* from Kocyan et al. (2007) (supplementary data S1, Supplementary Material online). For details on DNA sequencing, alignment, and phylogeny estimation, see Schaefer and Renner (2010).

The evolutionary relationship between *TIPTOP* genes, *TIPRE* sequences, and the transcripts from the African *M. charantia* seeds was inferred under Maximum Likelihood (ML) using the General Time Reversible model (Nei and Kumar 2000). The initial tree for heuristic search using Nearest-Neighbor-Interchange method was generated automatically by applying Neighbor-Join and BioNJ algorithms. The analysis was conducted in MEGA5 (Tamura et al. 2011). Nucleotide sequences used for this analysis are given in supplementary data S2, Supplementary Material online.

Mapping of Sites under Selection

Nucleotide sequences that encode the cyclic peptides were used to map sites that are under selection using MEGA5 (Tamura et al. 2011). The numbers of synonymous (*s*) and nonsynonymous (*n*) substitutions and the synonymous (*S*) and nonsynonymous (*N*) sites were estimated using joint ML reconstructions of ancestral states under the Muse–Gaut (Muse and Gaut 1994) and General Time Reversible (Nei and Kumar 2000) models. ML of codon undergoing selection was estimated through HyPhy (Kosakovsky Pond et al. 2005) using an automatically generated neighbor-joining tree. The probability of rejecting the null hypothesis of neutral

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

evolution (*P* value) was calculated as previously described (Suzuki and Gojobori 1999; Kosakovsky Pond and Frost 2005). Nucleotide sequences used for this analysis are given in supplementary data S3, Supplementary Material online.

Molecular Dynamics

To calculate the average structure of selected cyclic peptides against trypsin, the ordinate of the crystal structure of native MCoTI-II bound to trypsin (PDB#4GUX) was used as reference (Daly et al. 2013). Simulation of the complexes was performed as previously described (Swedberg et al. 2011).

RNA Secondary Structure Calculation

The folding energy of *TIPTOP* transcripts was calculated using Mfold (Zuker 2003) with default settings.

Accession Numbers

Sequence data from this work can be found in the GenBank database under the accession numbers KM408418 for *M. subangulata TIPTOP4*, KM408419 for *M. macrophylla TIPTOP5*, KM408420 for *M. macrophylla TIPTOP6*, KM408421 for *M. anigosantha TIPRE1*, KM408422 for *M. anigosantha TIPRE2*, KM408423 for *M. anigosantha TIPRE3*, KM408424 for *M. anigosantha TIPRE4*, and KM408425 for *M. friesiorum TIPRE5*.

Supplementary Material

Supplementary data S1–S3, tables S1–S5, and figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This research was supported by a grant from the National Health and Medical Research Council (APP1009267) and the Australian Research Council (LP130100550). D.J.C. is a National Health and Medical Research Council Professorial Research Fellow (APP1026501). J.S.M. is an Australian Research Council Future Fellow (FT120100013). J.E.S. is a National Health and Medical Research Council Early Career Fellow (APP1069819). T.M. is a recipient of an Endeavour Postgraduate Scholarship granted by the Australian Government. The authors thank E. Miles for comments on the manuscript, E.K. Gilding for helpful discussion, R. Widiatmojo for photo editing of *Momordica* seeds, and A. Jones of the Molecular and Cellular Proteomics Mass Spectrometry Facility at the Institute for Molecular Bioscience, The University of Queensland, for support and access to the facility.

References

Aboye TL, Ha H, Majumder S, Christ F, Debyser Z, Shekhtman A, Neamati N, Camarero JA. 2012. Design of a novel cyclotide-based CXCR4 antagonist with anti-human immunodeficiency virus (HIV)-1 activity. *J Med Chem.* 55:10729–10734.

Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, et al. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep.* 30:108–160.

Austin J, Wang W, Puttamadappa S, Shekhtman A, Camarero JA. 2009. Biosynthesis and biological screening of a genetically encoded library based on the cyclotide MCoTI-I. *Chembiochem* 10: 2663–2670.

Avrutina O, Schmoldt HU, Gabrijelcic-Geiger D, Le Nguyen D, Sommerhoff CP, Diederichsen U, Kolmar H. 2005. Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTI-II. *Biol Chem.* 386:1301–1306.

Bachovchin DA, Cravatt BF. 2012. The pharmacological landscape and therapeutic potential of serine hydrolases. *Nat Rev Drug Discov.* 11: 52–68.

Barber CJ, Pujara PT, Reed DW, Chiwocha S, Zhang H, Covello PS. 2013. The two-step biosynthesis of cyclic peptides from linear precursors in a member of the plant family Caryophyllaceae involves cyclization by a serine protease-like enzyme. *J Biol Chem.* 288: 12500–12510.

Brady K, Abeles RH. 1990. Inhibition of chymotrypsin by peptidyl trifluoromethyl ketones: determinants of slow-binding kinetics. *Biochemistry* 29:7608–7617.

Cascales L, Henriques ST, Kerr MC, Huang YH, Sweet MJ, Daly NL, Craik DJ. 2011. Identification and characterization of a new family of cell-penetrating peptides: cyclic cell-penetrating peptides. *J Biol Chem.* 286:36932–36943.

Chan LY, Gunasekera S, Henriques ST, Worth NF, Le SJ, Clark RJ, Campbell JH, Craik DJ, Daly NL. 2011. Engineering pro-angiogenic peptides using stable, disulfide-rich cyclic scaffolds. *Blood* 118: 6709–6717.

Chevreux B. 2005. MIRA: an automated genome and EST assembler [PhD thesis]. [Heidelberg (Germany)]: Ruprecht-Karls University.

Colgrave ML, Craik DJ. 2004. Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. *Biochemistry* 43:5965–5975.

Contreras J, Elnagar AYO, Hamm-Alvarez SF, Camarero JA. 2011. Cellular uptake of cyclotide MCoTI-I follows multiple endocytic pathways. *J Control Release.* 155:134–143.

Craik DJ. 2013. Joseph Rudinger memorial lecture: discovery and applications of cyclotides. *J Pept Sci.* 19:393–407.

Craik DJ, Daly NL, Bond T, Waine C. 1999. Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J Mol Biol.* 294:1327–1336.

Daly NL, Love S, Alewood PF, Craik DJ. 1999. Chemical synthesis and folding pathways of large cyclic polypeptide: studies of the cystine knot polypeptide kalata B1. *Biochemistry* 38: 10606–10614.

Daly NL, Thorstholm L, Greenwood KP, King GJ, Rosengren KJ, Heras B, Martin JL, Craik DJ. 2013. Structural insights into the role of the cyclic backbone in a squash trypsin inhibitor. *J Biol Chem.* 288: 36141–36148.

D'Souza C, Henriques ST, Wang CK, Craik DJ. Forthcoming 2014. Structural parameters modulating the cellular uptake of disulfide-rich cyclic cell-penetrating peptides: MCoTI-II and SFTI-1. *Eur J Med Chem.*

Elliott AG, Delay C, Liu H, Phua Z, Rosengren KJ, Benfield AH, Panero JL, Colgrave ML, Jayasena AS, Dunse KM. 2014. Evolutionary origins of a bioactive peptide buried within preproalbumin. *Plant Cell* 26: 981–995.

Ellis RH. 1991. The longevity of seeds. *HortScience* 26:1119–1125.

Getz JA, Cheneval O, Craik DJ, Daugherty PS. 2013. Design of a cyclotide antagonist of neuropilin-1 and-2 that potently inhibits endothelial cell migration. *ACS Chem Biol.* 8:1147–1154.

Gillon AD, Saska I, Jennings CV, Guarino RF, Craik DJ, Anderson MA. 2008. Biosynthesis of circular proteins in plants. *Plant J.* 53: 505–515.

Gracy J, Le-Nguyen D, Gelly JC, Kaas Q, Heitz A, Chiche L. 2008. KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007. *Nucleic Acids Res.* 36:D314–D319.

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

Greenwood KP, Daly NL, Brown DL, Stow JL, Craik DJ. 2007. The cyclic cystine knot miniprotein MCoTI-II is internalized into cells by macropinocytosis. *Int J Biochem Cell Biol.* 39:2252–2264.
Hamato N, Koshiba T, Pham T-N, Tatsumi Y, Nakamura D, Takano R, Hayashi K, Hong Y-M, Hara S. 1995. Trypsin and elastase inhibitors from bitter gourd (*Momordica charantia* LINN.) seeds: purification, amino acid sequences, and inhibitory activities of four new inhibitors. *J Biochem.* 117:432–437.
Hara S, Makino J, Ikenaka T. 1989. Amino acid sequences and disulfide bridges of serine proteinase inhibitors from bitter gourd (*Momordica charantia* LINN.) seeds. *J Biochem.* 105:88–92.
Hara-Nishimura I, Inoue K, Nishimura M. 1991. A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. *FEBS Lett.* 294:89–93.
Hedstrom L. 2002. Serine protease mechanism and specificity. *Chem Rev.* 102:4501–4524.
Hernandez JF, Gagnon J, Chiche L, Nguyen TM, Andrieu JP, Heitz A, Hong TT, Pham TTC, Nguyen DL. 2000. Squash trypsin inhibitors from *Momordica cochinchinensis* exhibit an atypical macrocyclic structure. *Biochemistry* 39:5722–5730.
Higaki JN, Gibson BW, Craik CS. 1987. Evolution of catalysis in the serine proteases. *Cold Spring Harb Symp Quant Biol.* 52:615–621.
Hiraiwa N, Nishimura M, Hara-Nishimura I. 1999. Vacuolar processing enzyme is self-catalytically activated by sequential removal of the C-terminal and N-terminal propeptides. *FEBS Lett.* 447:213–216.
Ji Y, Majumder S, Millard M, Borra R, Bi T, Elnagar AY, Neamati N, Shekhtman A, Camarero J. 2013. *In vivo* activation of the p53 tumor suppressor pathway by an engineered cyclotide. *J Am Chem Soc.* 135:11623–11633.
Joubert FJ. 1984. Trypsin isoinhibitors from *Momordica repens* seeds. *Phytochemistry* 23:1401–1406.
Kaas Q, Craik DJ. 2010. Analysis and classification of circular proteins in Cybase. *Pept Sci.* 94:584–591.
Kocyan A, Zhang L-B, Schaefer H, Renner SS. 2007. A multi-locus chloroplast phylogeny for the Cucurbitaceae and its implications for character evolution and classification. *Mol Phylogenet Evol.* 44: 553–577.
Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 22:1208–1222.
Kosakovsky Pond SL, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.
Krieger M, Kay LM, Stroud R. 1974. Structure and specific binding of trypsin: comparison of inhibited derivatives and a model for substrate binding. *J Mol Biol.* 83:209–230.
Kurth T, Ullmann D, Jakubke H-D, Hedstrom L. 1997. Converting trypsin to chymotrypsin: structural determinants of S1′ specificity. *Biochemistry* 36:10098–10104.
Li F, Zheng Q, Vandivier LE, Willmann MR, Chen Y, Gregory BD. 2012. Regulatory impact of RNA secondary structure across the *Arabidopsis* transcriptome. *Plant Cell* 24:4346–4359.
Li Y-D, Xie Z-Y, Du Y-L, Zhou Z, Mao X-M, Lv L-X, Li Y-Q. 2009. The rapid evolution of signal peptides is mainly caused by relaxed selection on non-synonymous and synonymous sites. *Gene* 436: 8–11.
Ling M-H, Qi H-Y, Chi C-W. 1993. Protein, cDNA, and genomic DNA sequences of the towel gourd trypsin inhibitor. A squash family inhibitor. *J Biol Chem.* 268:810–814.
Lu W, Qasim M, Laskowski M, Kent SB. 1997. Probing intermolecular main chain hydrogen bonding in serine proteinase-protein inhibitor complexes: chemical synthesis of backbone-engineered turkey ovomucoid third domain. *Biochemistry* 36: 673–679.
Luckett S, Garcia RS, Barker J, Konarev AV, Shewry P, Clarke A, Brady R. 1999. High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *J Mol Biol.* 290:525–533.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
Mahatmanto T, Poth AG, Mylne JS, Craik DJ. 2014. A comparative study of extraction methods reveals preferred solvents for cystine knot peptide isolation from *Momordica cochinchinensis* seeds. *Fitoterapia* 95:22–33.
Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol.* 11: 715–724.
Mylne JS, Chan LY, Chanson AH, Daly NL, Schaefer H, Bailey TL, Nguyencong P, Cascales L, Craik DJ. 2012. Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis. *Plant Cell* 24:2765–2778.
Mylne JS, Colgrave ML, Daly NL, Chanson AH, Elliott AG, McCallum EJ, Jones A, Craik DJ. 2011. Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nat Chem Biol.* 7: 257–259.
Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. New York: Oxford University Press.
Nguyen GKT, Zhang S, Wang W, Wong CTT, Nguyen NTK, Tam JP. 2011. Discovery of a linear cyclotide from the bracelet subfamily and its disulfide mapping by top-down mass spectrometry. *J Biol Chem.* 286:44833–44844.
Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8:785–786.
Poth AG, Chan LY, Craik DJ. 2013. Cyclotides as grafting frameworks for protein engineering and drug design applications. *Pept Sci.* 100: 480–491.
Poth AG, Colgrave ML, Lyons RE, Daly NL, Craik DJ. 2011. Discovery of an unusual biosynthetic origin for circular proteins in legumes. *Proc Natl Acad Sci U S A.* 108:10127–10132.
Rypniewski WR, Perrakis A, Vorgias CE, Wilson KS. 1994. Evolutionary divergence and conservation of trypsin. *Protein Eng.* 7:57–64.
Saska I, Gillon AD, Hatsugai N, Dietzgen RG, Hara-Nishimura I, Anderson MA, Craik DJ. 2007. An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclization. *J Biol Chem.* 282: 29721–29728.
Schaefer H, Heibl C, Renner SS. 2009. Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proc R Soc Lond B Biol Sci.* 276:843–851.
Schaefer H, Renner SS. 2010. A three-genome phylogeny of *Momordica* (Cucurbitaceae) suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia. *Mol Phylogenet Evol.* 54: 553–560.
Shewry PR, Casey R. 1999. Seed proteins. Dordrecht (The Netherlands): Springer.
Silverman IM, Li F, Gregory BD. 2013. Genomic era analyses of RNA secondary structure and RNA-binding proteins reveal their significance to post-transcriptional regulation in plants. *Plant Sci.* 205: 55–62.
Sommerhoff CP, Avrutina O, Schmoldt HU, Gabrijelcic-Geiger D, Diederichsen U, Kolmar H. 2010. Engineered cystine knot miniproteins as potent inhibitors of human mast cell tryptase β. *J Mol Biol.* 395:167–175.
Staal J, Dixelius C. 2007. Tracing the ancient origins of plant innate immunity. *Trends Plant Sci.* 12:334–342.
Suzuki Y, Gojobori T. 1999. A method for detecting positive selection at single amino acid sites. *Mol Biol Evol.* 16:1315–1328.
Swedberg JE, de Veer SJ, Sit KC, Reboul CF, Buckle AM, Harris JM. 2011. Mastering the canonical loop of serine protease inhibitors: enhancing potency by optimising the internal hydrogen bond network. *PLoS One* 6:e19302.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
Tan N-H, Zhou J. 2006. Plant cyclopeptides. *Chem Rev.* 106:840–895.

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016

Thompson RC, Blout ER. 1973. Dependence of the kinetic parameters for elastase-catalyzed amide hydrolysis on the length of peptide substrates. *Biochemistry* 12:57–65.

Thongyoo P, Bonomelli C, Leatherbarrow RJ, Tate EW. 2009. Potent inhibitors of β-tryptase and human leukocyte elastase based on the MCoTI-II scaffold. *J Med Chem.* 52:6197–6200.

Tokuriki N, Tawfik DS. 2009. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol.* 19:596–604.

Trabi M, Craik DJ. 2002. Circular proteins—no end in sight. *Trends Biochem Sci.* 27:132–138.

Yang P, Li X, Shipp MJ, Shockey JM, Cahoon EB. 2010. Mining the bitter melon (*Momordica charantia* L.) seed transcriptome by 454 analysis of non-normalized and normalized cDNA populations for conjugated fatty acid metabolism-related genes. *BMC Plant Biol.* 10:250.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406–3415.

Downloaded from http://mbe.oxfordjournals.org/ at UB der TU Muenchen on October 14, 2016