Squash Trypsin Inhibitors from *Momordica cochinchinensis* Exhibit an Atypical Macrocyclic Structure†

Jean-François Hernandez,*,‡ Jean Gagnon,‡ Laurent Chiche,§ Tuyet Mai Nguyen,|| Jean-Pierre Andrieu,‡ Annie Heitz,§ Thai Trinh Hong,|| T. Trân Châu Pham,|| and Dung Le Nguyen⊥

Institut de Biologie Structurale Jean-Pierre Ebel (CEA-CNRS), 41, rue Jules Horowitz, 38027 Grenoble Cedex 1, France, Centre de Biochimie Structurale, Faculté de Pharmacie, UMR5048 CNRS-INSERM-Université Montpellier I, 15, avenue Charles Flahaut, 34060 Montpellier, France, Centre de Biotechnologie, Université Nationale du Viet-Nam, 90, Nguyen Trai Street, Hanoï, Viet-Nam, and INSERM U376, CHU Arnaud-de-Villeneuve, 371, rue du doyen Gaston Giraud, 34295 Montpellier, France

Received December 28, 1999; Revised Manuscript Received March 10, 2000

ABSTRACT: Three trypsin inhibitors (TIs), from the seeds of the squash *Momordica cochinchinensis* (MCo), have been isolated and purified using gel filtration, ion exchange chromatography, and reverse-phase HPLC. Their sequences could be determined only after proteolytic cleavages. In the case of MCoTI-I and -II, it was shown that their polypeptide backbones are cyclic, a structure that has never been described in squash TIs. They contain 34 amino acid residues with 3 disulfide bridges and measured molecular masses of 3453.0 and 3480.7, respectively. They are the largest known macrocyclic peptides containing disulfide bridges. Their sequences show strong homology to other squash TIs, suggesting a similar three-dimensional structure and an analogous mechanism of action. A model of MCoTI-II was constructed by analogy to the crystal structure of the complex between bovine trypsin and CMTI-I, indicating that the linker connecting the two termini is flexible and does not impose significant geometrical constraints. This flexibility allows an Asp−Gly peptide bond rearrangement to occur in this region, giving rise to two isoforms of MCoTI-II. Although the importance of cyclization is not clear, it might confer increased stability and resistance to proteolysis. A minor species, MCoTI-III, was also characterized as containing 30 amino acid residues with a molecular mass of 3379.6. This component possesses a linear backbone with a blocked N-terminus. MCoTIs represent interesting candidates for drug design, either by changing their specificity of inhibition or by using their structure as natural scaffolds bearing new binding activities.

Serine proteinase inhibitors from various species belonging to both animal and plant kingdoms have been thoroughly studied and classified into several families according to their homology in primary structures, location of the reactive site, and disulfide bridge connectivities (*1*). The squash family (*2*) consists of almost 30 inhibitors identified from seeds of various Cucurbitaceae species (*2*−*5*). They are characterized by an open chain of about 30 amino acids and display high sequence identities (Figure 1A). A major common feature is the presence of six well-conserved Cys residues. The disulfide bridge pattern was established for EETI-II[1] from *Ecballium elaterium* (*6, 7*) and CMTI-I from *Curcubita maxima* (*8*), respectively, by 2-D NMR and X-ray crystallography, and shown to be 1−4, 2−5, 3−6. This pattern is thought to be shared by all members of the family and has been shown to assemble as a cystine knot (*9*). Such a structural motif has also been observed in other toxic or inhibitory peptides including ω-conotoxin GVIA and macrocyclic peptides such as kalata B1 and circulin A (*10*−*12*). Most squash inhibitors target trypsin with high efficiency ($K_a = 10^{10}-10^{12}$ M^{-1}), and their reactive site is located at an Arg (Lys)−Ile bond close to the amino terminus. One exception is the elastase inhibitor MCEI-I from *Momordica charantia* which possesses a Leu−Ile bond at its reactive site (*13*).

Due to their small size and unique structure, these inhibitors have been studied extensively. Peptide synthesis (*14, 15*) and synthetic gene expression (*16, 17*) are two convenient routes to their preparation, and have allowed detailed structure−function studies (*18, 19*). All of these features, and also their high stability and rigidity, make these natural compounds very interesting models for the design of inhibitors for other proteases, such as elastase and chymotrypsin. Indeed, changing amino acid residues which interact with the protease can modify the specificity of

† This work was supported in part by the Commissariat à l'Energie Atomique and the Centre National de la Recherche Scientifique.

* To whom correspondence should be addressed. Phone: +33 (0)4 76 88 50 79. Fax: +33 (0)4 76 88 54 94. E-mail: hernande@ibs.fr.
‡ Institut de Biologie Structurale Jean-Pierre Ebel.
§ UMR5048 CNRS-INSERM-Université Montpellier I.
|| Centre de Biotechnologie, Université Nationale du Viet-Nam.
⊥ INSERM U376, CHU Arnaud-de-Villeneuve.

[1] Abbreviations: CM-Cys, (carboxymethyl)cysteine; CMTI, *Curcubita maxima* trypsin inhibitor; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; EETI, *Ecballium elaterium* trypsin inhibitor; endo-Asp-N, endoproteinase Asp-N; endo-Lys-C, endoproteinase Lys-C; ES-MS, electrospray-mass spectrometry; HPLC, high-pressure liquid chromatography; IU, inhibitory unit(s); MCoTI, *Momordica cochinchinensis* trypsin inhibitor; rt, retention time; TFA, trifluoroacetic acid; TI, trypsin inhibitor; TIA, trypsin inhibitory activity; Tris, tris(hydroxymethyl)aminomethane.

10.1021/bi9929756 CCC: $19.00
Published on Web 04/21/2000

inhibition (*9*, *20*). These small proteins may also be used as structural scaffolds for the presentation of new binding activities (*21*, *22*).

Here, we describe new members of the squash family inhibitors, which have been isolated from the seeds of *Momordica cochinchinensis*, a common Cucurbitaceae in Vietnam. These seeds are used in cooking rice and in traditional Chinese medicine (*23*). The amino acid sequences of the major component, MCoTI-II, and of two minor species, MCoTI-I and MCoTI-III, have been determined. It is also shown that MCoTI-I and MCoTI-II are cyclic polypeptides. These inhibitors are the first members of the squash family shown to exhibit such a structural feature.

EXPERIMENTAL PROCEDURES

Materials. Dormant MCo seeds were obtained from ripe MCo. Chymotrypsin, endoproteinase Lys-C (endo-Lys-C), endoproteinase Asp-N (endo-Asp-N), and pyroglutamyl aminopeptidase were obtained from Boehringer (Mannheim, Germany). Iodoacetic acid was obtained from Sigma (St. Quentin Fallavier, France). Acetonitrile was obtained from Acros (Noisy-Le-Grand, France). All other chemicals were of analytical grade.

Isolation and Purification of MCoTIs. Dormant seeds were crushed using a mixer, and extracted with 20 mM sodium acetate, pH 4.5. After centrifugation at 4 °C, the supernatant was loaded onto a Sephadex G75 column (6 × 90 cm) (Pharmacia, Uppsala, Sweden) and eluted with 200 mM sodium acetate, pH 5.0, at a flow rate of 70 mL/h. Trypsin inhibitory activity (TIA) was measured for each fraction (15 mL) using the method described by Hanspal et al. (*24*), as modified by Pham et al. (*25*). Fractions containing TIA were pooled and fractionated on a Mono-S column HR515 (Pharmacia) equilibrated with 5 mM sodium acetate, pH 3.6. Stepwise elution by increasing concentrations of NaCl was performed as indicated in Figure 2, at a flow rate of 0.5 mL/min. Several peaks containing TIA were separated, and called A−F, in the order of elution.

MCoTIs were further purified by semipreparative reverse-phase HPLC on a Waters DeltaPrep 4000 apparatus using two PrepPak cartridges Delta-Pak C18 (22 × 125 mm) (Waters). Elution was carried out with a linear gradient of 12−30% acetonitrile in 0.1% aqueous TFA in 90 min, at a flow rate of 10 mL/min. A further step was necessary to achieve purification of some fractions using the analytical HPLC system described below.

Analytical Reverse-Phase HPLC. Chromatographic analysis of MCoTIs, either intact or after enzymic digestion, was performed using the Beckman Gold system, including a diode-array detector (detection at 215 and 280 nm), on a Vydac (Hesperia, USA) C18 column (0.46 × 25 cm), by means of a linear gradient (indicated in the appropriate figure legend) of acetonitrile in 0.1% aqueous TFA over 30 min (flow rate: 1 mL/min).

Mass Spectrometry Analyses. Electrospray ionization mass spectra were obtained on an API III triple-quadrupole mass spectrometer (PE/Sciex), equipped with a nebulizer-assisted electrospray (ionspray) source, as described previously (*26*).

Amino Acid Analysis. Inhibitor samples were hydrolyzed for 24 h under reduced pressure at 110 °C in constant-boiling 6 N HCl containing 1% (w/v) phenol. Analyses were performed with a Beckman 7300 amino acid analyzer.

N-Terminal Sequence Analysis. N-terminal sequence analyses were performed using an Applied Biosystems model 477A protein sequencer, and amino acid phenylthiohydantoin derivatives were identified and quantitated on-line with a model 120A HPLC system, as recommended by the manufacturer.

Proteolytic Cleavages of MCoTIs. The inhibitors (2 nmol of each) were first treated with DTT (150-fold excess), at pH 8.3 for 3 h at 37 °C. The reduced cysteines were then alkylated with iodoacetic acid (3-fold excess over DTT) for 1 h at 4 °C. The reduced and alkylated peptides were isolated by reverse-phase HPLC and characterized by ES-MS. They were then digested with endo-Lys-C at an enzyme/substrate ratio of 1/10 (w/w) in 0.1 M Tris-HCl, with 1 mM EDTA (pH 8.5), for 4 h at 37 °C. The resulting fragments were separated by reverse-phase HPLC on a C18 Vydac column as described, and characterized by ES-MS and N-terminal sequence analyses. The largest fragments were further digested with bovine chymotrypsin at an enzyme/substrate ratio of 1/50 (w/w), in 50 mM sodium phosphate buffer, pH 8.0, for 3 h at 37 °C. The subfragments were separated and analyzed as described above. The reduced and alkylated MCoTI-II was also submitted to digestion with endo-Asp-N at an enzyme/substrate ratio of 1/50 (w/w) in 50 mM sodium phosphate, pH 8.0, for 5 h at 37 °C. In the case of MCoTI-III, the reduced and alkylated inhibitor was treated with pyroglutamyl aminopeptidase as described by Podell and Abraham (*27*).

Three-Dimensional Modeling. The complex between trypsin and MCoTI-II was homology-modeled from the crystal structure of the complex formed between bovine trypsin and CMTI-I, an inhibitor from the seeds of the squash *Cucurbita maxima* (*8*) (PDB accession no. 1ppe). Homology modeling was performed with the MODELLER program release 4 (*28*). To build the macrocyclic inhibitor, the default patching of residues at the N- and C-termini was turned off, and typical peptide bond restraints were applied to the inhibitor ends to ensure proper cyclization. The N- and C-termini of trypsin were explicitly patched. To generate models in which the P1 residue of MCoTI-II (Lys) fits adequately within the specificity pocket of trypsin, distance constraints had to be imposed between this residue and Asp^{189} at the bottom of the trypsin specificity pocket. Ten models were computed with slight randomization of the starting coordinates. Small deviations from the X-ray coordinates were invariably observed that seemed to result from the optimization procedure rather than from true structural problems arising from the mutations and the cyclization. To reduce these apparent artifacts, the models were energy-minimized with the program AMBER 5 (*29*). These energy minimizations were done in 6 Å shells of water. A short molecular dynamics was first performed to best fit the solvent molecules around the complex. Then several runs of minimization were applied in which positional constraints that pull the molecule toward X-ray coordinates for trypsin and conserved residues of the inhibitor were gradually lowered. At the end of the refinement process, only very small deviations from the X-ray coordinates were obtained for conserved atoms, as well as large negative energies, indicating that the MCoTI-II

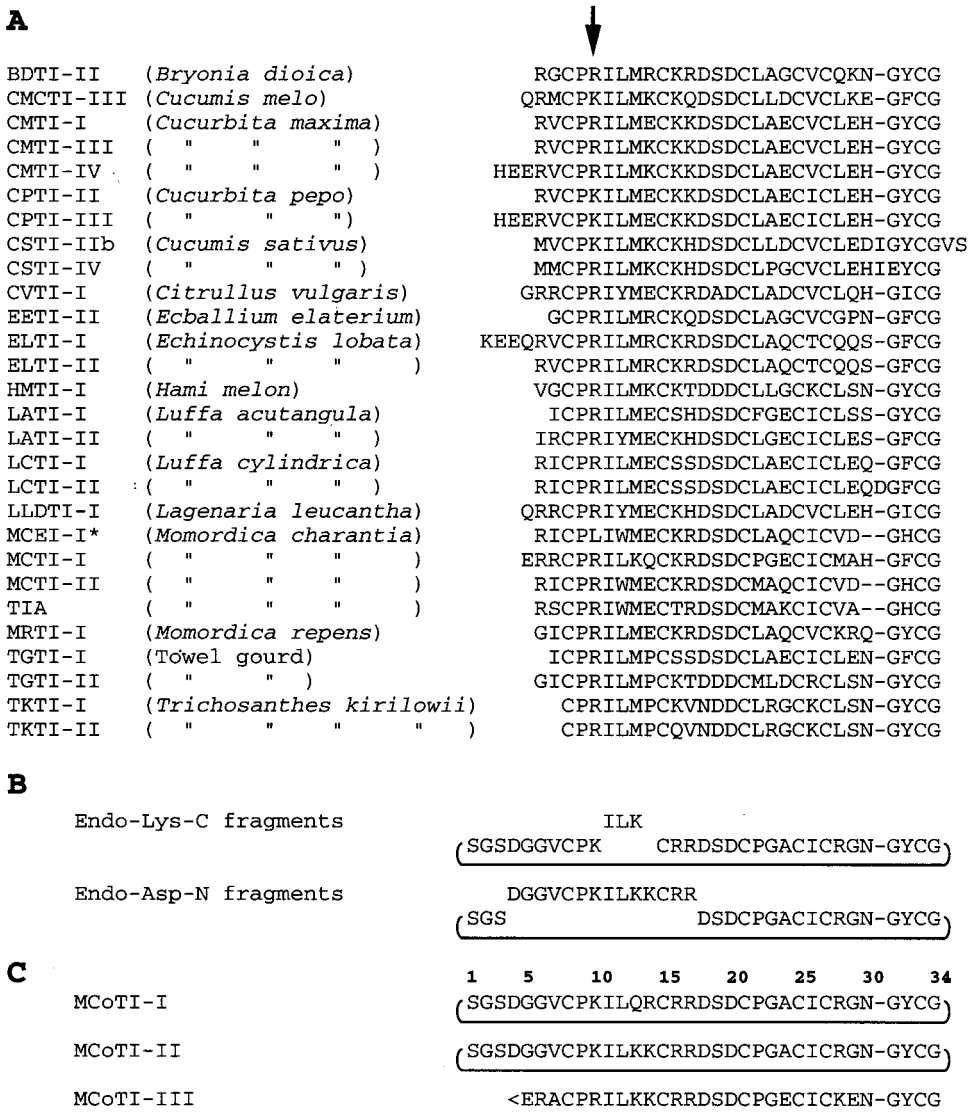

FIGURE 1: Comparison of MCoTI sequences with known squash family inhibitors. (A) Inhibitor sequences were taken from Swiss-Prot and TrEMBL protein sequence databases, excepted ELTI-I and -II (*54*), LATI-I and -II (*55*), and HMTI-I (*36*). (B) Amino acid sequences of peptides generated by digestion of MCoTI-II with endoproteinases Lys-C or Asp-N. The sequences are aligned against the squash TIs presented in (A). Cys was detected as its *S*-(carboxymethyl)cysteine derivative. (C) Amino acid sequences of MCoTI-I, MCoTI-II, and MCoTI-III.

sequence and cyclic feature were totally compatible with the known structure of CMTI-I in complex with trypsin.

RESULTS

Trypsin inhibitory activities (TIA) were extracted from dormant seeds of *Momordica cochinchinensis*, as described under Experimental Procedures. The separation profile obtained by cation-exchange chromatography is shown in Figure 2 with the superimposed profile of TIA. Several components were isolated from peaks A−F, except for the poorly defined peak C, with a purity suitable for sequence determination, as shown by reverse-phase HPLC and mass spectrometry. MCoTI-I, -II, and -III, named according to their order of elution (*2*), were found to be the major components of peaks B, E, and F, respectively. Isoforms of MCoTI-II were present in peaks D and F. Taken together, MCoTI-II was the major inhibitor present in seeds.

The purified MCoTI-II species eluted as a single homogeneous peak on reverse-phase HPLC (Figure 3A), and its UV spectrum indicated that it contains at least one Tyr but no Trp. The molecular weight of this inhibitor was determined by ES-MS to be 3453.0 ± 0.2. To check that the peptide contained three disulfide bridges as shown for other TIs of the squash family, and prior to determination of its sequence, it was reduced and alkylated as described under Experimental Procedures. As shown by mass spectrometry, this treatment yielded a single derivative with a molecular weight of 3807.0 ± 0.3 (net increase of 354), indicating incorporation of six carboxymethyl groups. This result suggested that, as expected, the inhibitor contained six Cys residues involved in three disulfide bridges (calculated value for a three-disulfide-containing starting material: 3807.24). Amino acid analysis indicated that MCoTI-II was composed of 34 amino acid residues, including a high content of Gly (Table 1). However, the calculated molecular mass derived from this analysis was between 15 Da (4 Asp) and 19 Da (4 Asn) greater than the molecular weight of MCoTI-II, as determined by ES-MS. Furthermore, attempts to directly sequence the carboxymethylated peptide by Edman degradation were unsuccessful, suggesting a blocked N-terminus.

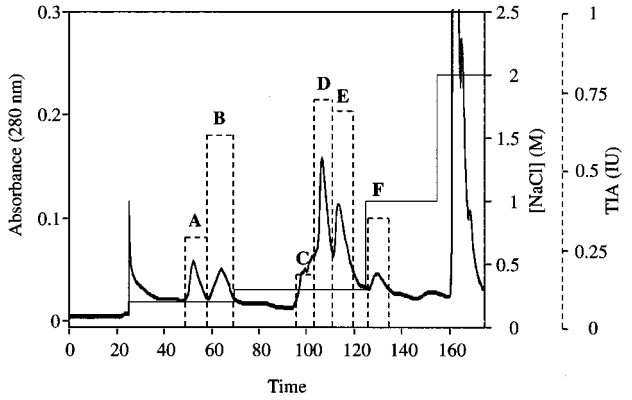

FIGURE 2: Ion-exchange chromatography of compounds with TIA eluted from Sephadex G75. Collected fractions with TIA from Sephadex G75 chromatography were loaded on a mono-S column, and elution was performed as described under Experimental Procedures. Detection was made by absorbance at 280 nm. The NaCl gradient was achieved by stages as indicated on the figure. Several peaks of TIA indicated A−F were collected; their respective TIAs [expressed as inhibitory units (IU)] are shown (1 IU = amount of inhibitor which reduces the activity of 2 mg of trypsin by 50%).

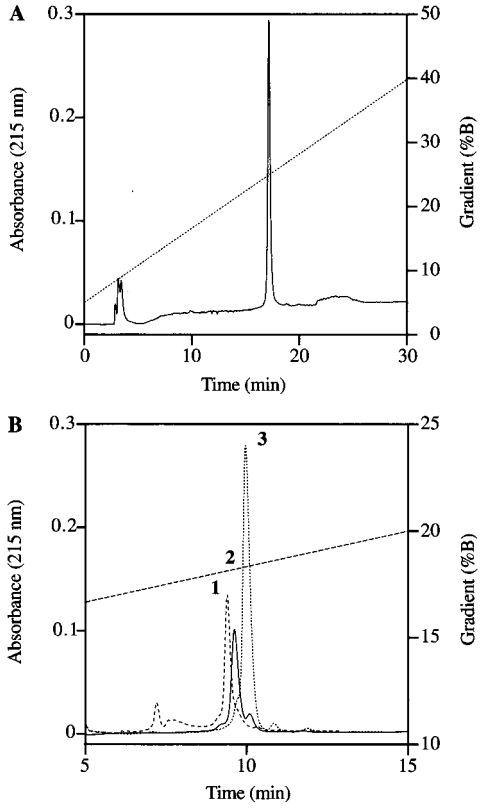

FIGURE 3: (A) Analytical reverse-phase HPLC elution profile of purified MCoTI-II. Vydac C18 column, 0.46 × 25 cm; flow rate, 1 mL/min; solvent A, 0.1% TFA; solvent B, 0.09% TFA in acetonitrile. The gradient was 5−40% solvent B in solvent A in 30 min. (B) Relative reverse-phase HPLC retention times (rt) of MCoTI-II isoforms. Peak 1 (rt = 9.33 min), 3452.8 species. Peak 2 (rt = 9.55 min), 3435.5 species. Peak 3 (rt = 9.91 min), MCoTI-II. The relative peak heights of the superimposed selected chromatographic profiles are not representative of the real proportion of each species. Same conditions as in (A), excepted for the gradient which was 15−25% solvent B in solvent A in 30 min.

The presence of a pyroglutamate residue was precluded as no Glu residue was significantly detected by amino acid analysis. The alkylated derivative of MCoTI-II was digested with endoproteinase Lys-C, yielding two fragments amenable to amino acid sequencing. The sequence of the small fragment was found to be ILK, corresponding to the measured mass of 372.2 ± 0.2. The sequence of the large fragment (measured mass = 3342.1 ± 0.7) was obtained by N-terminal sequence analysis coupled with further digestion using chymotrypsin.

Table 1: Amino Acid Analysis of MCoTI-II

amino acid	residues from amino acid analysis[a]	residues from sequencing
Asp + Asn	4.0	4 (3D, 1N)
Thr	0.0	0
Ser	2.7	3
Glu + Gln	0.2	0
Pro	2.0	2
Gly	7.4	7
Ala	1.1	1
Cys	5.3[b]	6[c]
Val	1.2	1
Met	0.0	0
Ile	1.8	2
Leu	1.0	1
Tyr	0.9	1
Phe	0.0	0
His	0.0	0
Lys	2.9	3
Arg	3.0	3
Trp	0.0	0

[a] The amounts determined for Ala, Val, Ile, Leu, Lys, and Arg were used to calculate the integer unit. [b] Cys was analyzed as cystine. [c] Cys was detected as its *S*-(carboxymethyl)cysteine derivative.

The fragments were aligned as shown in Figure 1B, indicating that the sequence of the largest endo-Lys-C peptide displays strong similarities with the known TIs from the squash family. However, it is remarkable that the sequence of the first 20 residues is homologous to the C-terminal portion of the aligned TIs, whereas, conversely, the sequence of the C-terminal part, ending by the putative reactive site, matches the N-terminal portion of the TIs. This strongly suggests that MCoTI-II possesses a macrocyclic structure in addition to its three disulfide bridges.

Addition of the molecular weights of the two fragments produced by endo-Lys-C showed that a small portion with a mass of 110.5 was missing. Considering the restricted specificity of the enzyme for Lys residues and the amino acid analysis of MCoTI-II, it was hypothesized that the missing link was a Lys residue. If the peptide was linear, the sum of the masses of the three pieces would yield a value of 3825.0, i.e., an extra mass of 18 compared to the measured molecular weight of the alkylated MCoTI-II. The position of the extra Lys residue was deduced from the alignment with the known TIs. The resulting sequence of MCoTI-II is shown in Figure 1C. This sequence was fully confirmed by digestion of the reduced and alkylated peptide with endo-Asp-N. The produced fragments identified by ES-MS and N-terminal sequencing are aligned in Figure 1B. The amino acid analysis (Table 1) was therefore fully consistent with the amino acid content derived from sequence analyses. These combined data clearly indicated that MCoTI-II has a cyclic structure, in agreement with its observed resistance to Edman degradation.

Other species were also separated and identified using a similar approach. Two of these species, isolated from peaks D and F (Figure 2), were derived from MCoTI-II. Figure

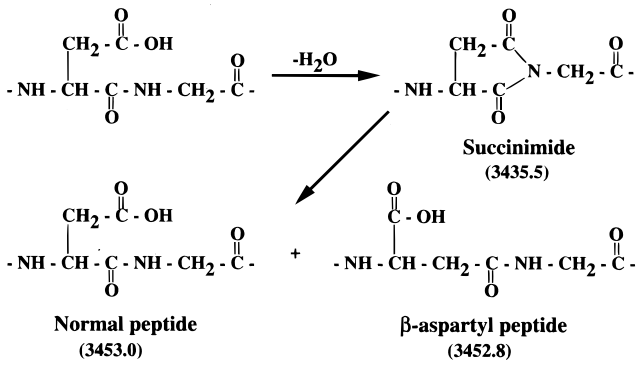

FIGURE 4: Formation of succinimide at the Asp[4]−Gly[5] bond of MCoTI-II. The succinimide can then re-open to give MCoTI-II and a species with identical mass possessing a β-aspartyl bond. Measured masses are in parentheses.

3B illustrates the relative chromatographic retention of these isoforms. They displayed molecular masses of 3452.8 ± 0.3 and 3435.5 ± 0.7, i.e., identical to and about 18 mass units less than that of MCoTI-II, respectively. Endo-Lys-C digestion of the carboxymethylated 3452.8 species yielded a short fragment, ILK, previously observed in the case of MCoTI-II, and a large fragment with the same mass but a slightly different HPLC retention time compared to that obtained for MCoTI-II. The residues identified for the large fragment are also identical, starting at CM-Cys[15] [numbering of MCoTI-II (see Figure 1C)]. However, its sequencing stopped at Ser[3], suggesting that the next two residues (Asp-Gly in MCoTI-II) do not form a regular peptide bond. In the case of the 3435.5 species, digestion with endo-Lys-C also produced the short ILK fragment, and, surprisingly, two isoforms of the large fragment with a mass identical to that obtained with MCoTI-II, instead of a single one with an expected default mass of 18 units. In addition, one isoform had an HPLC retention time identical to that of the corresponding fragment of MCoTI-II, whereas the other one coeluted with the large fragment obtained for the 3452.8 species. The sequence of the former fragment was found to be identical to that obtained for MCoTI-II, while sequencing of the latter stopped at Ser[3]. Taken together, these results strongly suggested that the Asp[4]−Gly[5] bond of MCoTI-II can cyclize and form a succinimide (3435.5 species, peak F), which is then able to reopen, as observed during digestion of the 3435.5 species, yielding two isoforms, one corresponding to MCoTI-II (peak E), the other containing an unsequenceable β-Asp−Gly bond (3452.8 species, peak D) (Figure 4).

A third species, with a measured mass of 3480.7 ± 0.3, MCoTI-I, was purified from peak B and shown to possess a sequence different from that of MCoTI-II. Endo-Lys-C digestion of its carboxymethylated form yielded a single fragment, with a mass (3852.3 ± 0.4) approximately 18 units greater than that of the starting material, showing that MCoTI-I also possesses a macrocyclic structure. Its sequence was determined (Figure 1C) and found to differ from that of MCoTI-II at two positions: Gln[13]-Arg[14] instead of Lys[13]-Lys[14]. Peak A (Figure 2) contained a species with a molecular mass identical to that of MCoTI-I. It was thought to be derived from the same Asp−Gly bond rearrangement as observed for MCoTI-II. It has not been further characterized. The last species, MCoTI-III (peak F), was not cyclic, although it had a blocked N-terminus. Its sequence (Figure 1C) could be determined after treatment with pyroglutamyl aminopeptidase. Taking into account the N-terminal pGlu residue, the molecular mass calculated from the sequence (3379.9) corresponded to the experimental value (3379.6 ± 0.5).

Two different macrocyclic TIs were identified from the seeds of *Momordica cochinchinensis*. To evaluate the possible impact of cyclization on the structure and the function of MCoTI-II, its complex with trypsin was homology-modeled from the crystal structure of the complex formed between bovine trypsin and CMTI-I (*8*). The resulting model is displayed in Figure 5. Although it has not been experimentally established, the disulfide pattern of MCoTI-II was assumed to be the same as that derived from the three-dimensional structure of the homologous TIs EETI-II and CMTI-I (*6*−*8*).

DISCUSSION

MCoTI-I and MCoTI-II represent the first examples of naturally occurring squash TIs that exhibit cyclization of the entire amino acid backbone. To our knowledge, they are not only the largest known squash TIs with 34 residues, but also the largest of the known macrocyclic peptides containing disulfide bridges (*10*, *30*−*35*). The extra residues form a probably flexible linker (see Figure 1C), which is essentially hydrophilic and composed of small amino acid residues. It is likely that such features are necessary to accommodate the junction of the two "termini". Analysis of the main chain conformation of the computed model indicates that the linker does not impose significant geometrical constraints (Figure 5A). On the other hand, the three-dimensional model shows that Gly[6] is in close contact with Ala[24]. Therefore, beside the flexibility afforded by the glycines, Gly[6] might also have been favored due to spatial requirements. The only specific interaction between the linker and the remainder of the inhibitor, as suggested by the model, is a possible H-bonding between the side chains of Ser[1] and Arg[29]. The central residues of the linker, Ser[3], Asp[4], and Gly[5], are largely solvent-exposed. Cyclization to succinimide of an Asp−X bond, as strongly suggested in MCoTI-II (Asp[4]−Gly[5] bond), is a frequent phenomenon arising in proteins, and is mainly observed when X is small. The most favorable case is when X = Gly, as in MCoTI-II. Cyclization to succinimide in this region reinforces the idea that the N-to-C linker in MCoTI-II is flexible.

It is worth noting that an N-to-C cyclization of a different type has been previously observed in squash TIs since the N-terminal arginine in CMTI-I is engaged in a salt bridge with the C-terminal carboxylate (*8*). This arginine is conserved in 60% of the TIs shown in Figure 1A. Therefore, at present, three different cases have been recognized regarding the N-to-C interaction in the squash inhibitors as shown in Figure 5B: (i) no N-to-C interaction (i.e., in EETI-II); (ii) an electrostatic cyclization (i.e., in CMTI-I); (iii) a full peptide cyclization, as shown in this paper.

Comparison of the sequences of MCoTI-I and -II with the sequences of known squash TIs revealed 48−70% sequence identity (Figure 1). The homology is particularly high for the sequence of the trypsin inhibitory loop (CPKILQRC and CPKILKKC for MCoTI-I and -II, respectively, and CPRILKQC for the closest known TI, MCTI-I). This strongly

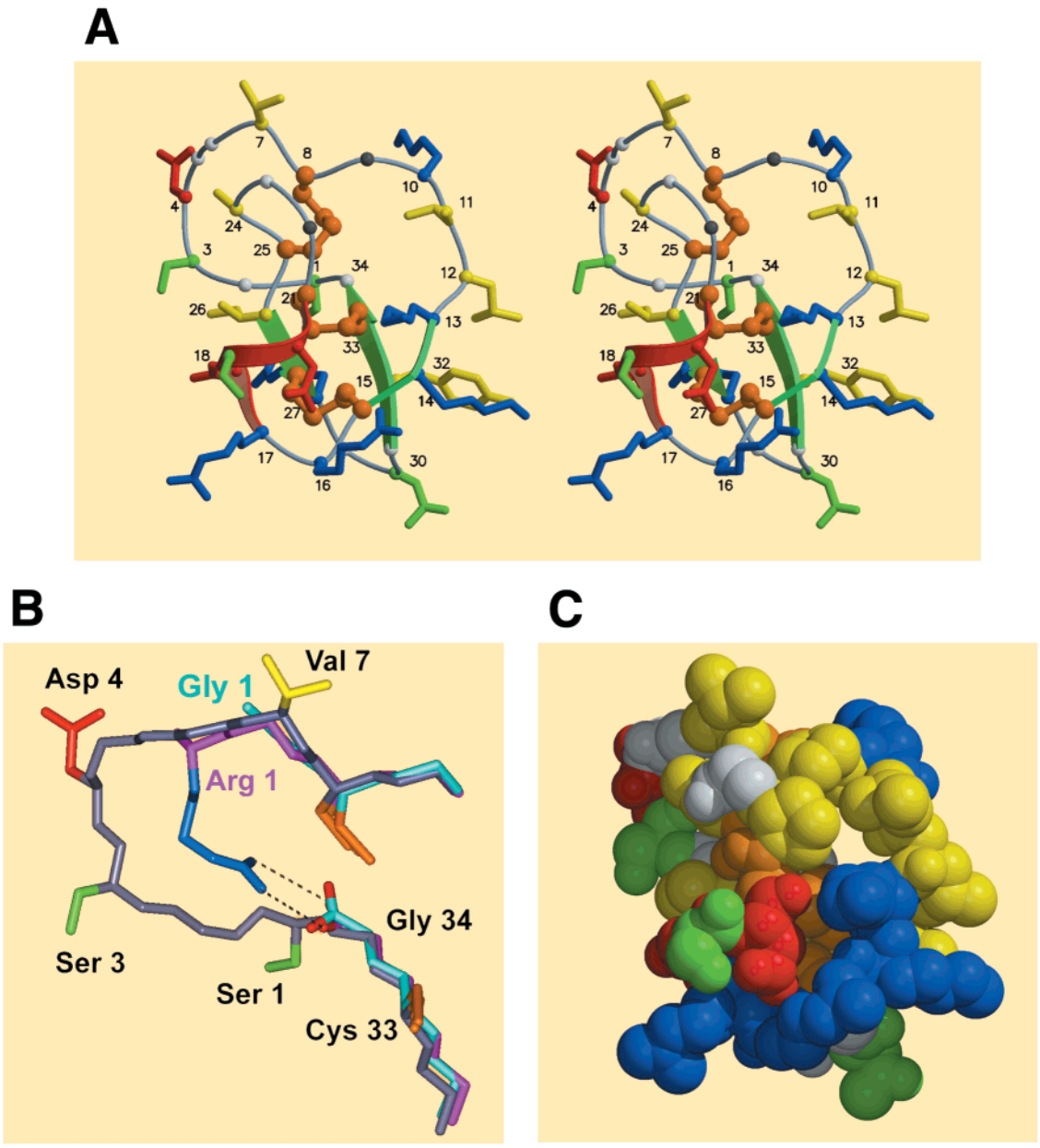

FIGURE 5: Schematic representations of the three-dimensional model of MCoTI-II. Side chains are colored blue, red, green, yellow, and orange for basic, acidic, polar, apolar residues, and cysteines, respectively. (A) Stereoview of MCoTI-II. Most of the residues are numbered. Cysteines are displayed as balls and sticks. All C-α atoms are displayed as spheres (colored white and black for Gly and Pro residues, respectively). The triple β-strand is displayed as green arrows and the 3_{10} helix as a red ribbon. Note that the first strand (residues 13−15) of the sheet is viewed from the thin edge. (B) Comparison of the N-to-C linker in MCoTI-II (gray backbone and black labels), CMTI-I (purple backbone and label), and EETI-II (light blue backbone and label). The salt-bridge between Arg^1 and the C-terminus in CMTI-I is shown as black dashed lines. The orientation of MCoTI-II is approximately the same as in panel A. (C) Corey−Pauling−Kortun representation of MCoTI-II in exactly the same orientation as in panel A. Cationic residues are in blue and hydrophobic residues in yellow.

suggests that MCoTI-I and -II share the same mechanism of trypsin inhibition as other squash inhibitors and that their cyclic structure originates from a different processing. The posttranslational modification leading to the cyclization of the peptide backbone probably involves the Gly residue "34" which corresponds to the C-terminus of most known TIs and was shown to be the last gene-coded residue of TGTI-II (*36*). Processing of TI precursors in their N-terminal region is variable as already pointed out by Wieczorek et al. (*2*) (Figure 1A). In the case of MCoTI-I and -II, seven residues are present. It is noteworthy that this heptapeptide (SGSDGGV) shares high homology with the corresponding part of the precursor of TGTI-II (SGRHGGI) (*36*). Since TGTI-II has a linear structure, this homology suggests that the linker sequence may not be involved in the cyclization process. The reasons why MCoTI-I and -II are the only characterized TIs of the squash family to be cyclic remain to be determined. It might depend on the presence of a specific and yet unknown transpeptidase. One may also argue that all squash TIs undergo cyclization first, and then reopen under the action of specific endo-peptidases, except in the case of MCoTI-I and -II. However, such a process would seem complicated and wasteful. In addition, we show here that *M. cochinchinensis* is also able to produce a linear TI (MCoTI-III). Information regarding the structural requirements for cyclization would be obtained from the sequence determination of the precursors of the macrocyclic peptides isolated from plants (see below).

It is not known whether cyclization occurs before or after disulfide bridging. It is interesting to note that the formation of disulfide bridges during folding of synthetic kalata B1, a peptide also containing a cystine knot and a macrocyclic structure (see below), has been shown to be facilitated when the N-to-C cyclization is first performed (*37*). However, it is very likely that, in vivo, cyclization occurs after oxidative folding which would bring the N- and C-termini of the linear precursor in close proximity. Recently, a cyclic antimicrobial 18-residue peptide also containing 3 disulfide bridges, but with a different pairing, was isolated from primate leukocytes (*38*). The search for its precursor showed that the peptide originates from a double head-to-tail ligation of two homologous nine-residue peptides, each containing three cysteines. The ligation requires that the peptide termini are in close proximity, this being most probably achieved by specific interaction between the two precursors and/or formation of one interchain disulfide bridge.

Very recently, the isolation of a TI from *M. cochinchinensis* was reported by Huang et al. (*39*). This TI, called MCCTI-1, is derived from cleavage at the level of the reactive site. Only the first 13 N-terminal residues were determined, and these were found to be identical to the corresponding part of MCoTI-II. However, its reported molecular mass (3480 ± 2, see Figure 4 within reference *39*) is different than expected for cleaved MCoTI-II (3471). As a matter of fact, this mass is closer to that of intact MCoTI-I (3481). One might therefore hypothesize that MCCTI-1 is identical to MCoTI-I, but that only a contaminant corresponding to a cleaved form of MCoTI-II could be directly sequenced. This would explain that only partial data concerning the sequence of MCCTI-1 have been reported, and that no firm evidence pointing to a possible macrocyclic structure could be reliably deduced.

The physiological role of the cyclization observed in MCoTI-I and -II remains to be determined. As linear squash inhibitors already display both high stability and strong affinity for trypsin, cyclization might confer resistance to exopeptidases and provide additional interaction sites with trypsin. Indeed, examination of the three-dimensional models of the complex formed between MCoTI-II and trypsin suggests that Asp^4 in the N-to-C linker could possibly form a salt bridge with Lys^{224} of trypsin. Confirmation of the effectiveness of such an interaction and its possible role in trypsin inhibition will await further experimental information.

MCoTI-I and -II share structural features with a series of macrocyclic peptides isolated from the Rubiaceae and Violaceae plant families. These include circulins A and B (*Chassalia parvifolia*) which exhibit anti-HIV activity (*30*), cyclopsychotride A (*Psychotria longipes*) which inhibits neurotensin binding to cell membranes (*31*), kalata B1 (*Oldenlandia affinis*) characterized for its uterotonic activity (*11, 40*), and a number of homologous peptides isolated from *Viola arvensis*, *V. hederaceae*, and *V. odorata* (*32*−*34*). These small peptides (29−31 amino acids) contain 6 half-cystine residues with disulfide pairings 1−4, 2−5, 3−6 (*11, 41*) as in TIs of the squash family. Structural studies performed on members of this family (kalata B1, circulin A, and cycloviolacin O1) showed that they contain a triple-stranded, antiparallel β-sheet and a cystine knot, which superimpose reasonably well onto the corresponding parts of the squash inhibitors (Figure 6) (*7, 10*−*12, 34*). These

```
                    →          →   →
Circulin A   -RNG-IPC---GESCVWIP-CISAALGCSCKN-KVCY
Kalata B1    -RNGLPVC---GETCVGGT-CNTP--GCTCSW-PVCT
MCoTI-II     SGSDGGVCPKILKKCRRDSDC---PGACICRGNGYCG
```

FIGURE 6: Sequence alignment of the cyclic peptides MCoTI-II, kalata B1, and circulin A on the basis of half-cystine positions and β-strand (arrows) hydrogen-bonding patterns. β-Strands for MCoTI-II are predicted from the 3-D structure of EETI-II and CMTI-I (*6*−*8*). The half-cystines are highlighted in bold. (Adapted from *10* and *45*.)

structural similarities confirm that squash TIs could, a priori, easily accommodate such a cyclic configuration. However, although the members of the squash inhibitor family, on the one hand, and of the Rubiaceae and Violaceae families, on the other hand, share a similar three-dimensional structure, they show very little sequence similarity apart from the six Cys residues, as exemplified by MCoTI-II in Figure 6. Moreover, folding studies of EETI-II (*42, 43*) as well as careful three-dimensional structural comparisons (*44*) have suggested that the structural similarity between families is limited to the elementary Cystine-Stabilized Beta-Sheet (CSB) submotif that includes the triple-stranded β-sheet but only two out of three disulfide bridges (*45*). In particular, the N-terminal segment in the squash TIs, which bears the inhibitory loop, has little in common with the corresponding part of other cystine-knot peptides. Therefore, the N-terminus of each family is clearly differentiated, and attempts to make structural comparisons of the N-to-C linkers between the two families would probably be irrelevant.

The biological roles of the macrocyclic peptides from the Rubiaceae and Violaceae families in these plants are not known, but, as for the squash TIs, they might be present as a defense mechanism. Indeed, antimicrobial activity was recently reported for some of these peptides (*46*). Interestingly, such properties were also observed for Momosertatin, a TI preparation isolated from *M. cochinchinensis* seeds and containing mainly MCoTI-II (Pham, T. T. C., et al., personal communication). It is generally considered that antimicrobial peptides are characterized by clusters of hydrophobic and cationic amino acids exposed on their surface (*47*). Similarly, the surface of MCoTI-II displays both well-defined cationic and hydrophobic clusters (Figure 5C), which could be related to antimicrobial properties.

One very interesting feature of macrocyclic peptides is their high stability. Kalata B1 was shown to be resistant to proteases, including thermolysin, trypsin, and pepsin, and to boiling (*11, 40*). Circulins were also not significantly cleaved by proteases (*41*). Similarly, MCoTI-II is resistant to cleavage by thermolysin for more than 48 h at 50 °C, and to heat treatment of the seeds (unpublished data). Furthermore, the lack of N- and C-termini confers resistance to exopeptidases. It is also noteworthy that these compounds are probably orally active, as is the case of kalata B1, the active component of extracts used in traditional medicine (*40*). These plant macrocyclic peptides thus represent interesting structures for drug design. Considering squash TIs, it has already been shown that it is possible to modify their specificity of inhibition. For instance, changing the P1 Arg residue in EETI-II to Ala resulted in a powerful elastase inhibitor (*9*). Grafting an anti-carboxypeptidase motif onto the EETI-II structure yielded a double-headed inhibitor efficient against both trypsin and carboxypeptidase A (*21*). It would be interesting to use the very stable MCoTI-II as a

starting point for a similar approach, either to change its primary specificity or to target other Arg-dependent serine proteases such as proteases of the coagulation and complement cascades. Another approach would be to transfer specific sites to this structure used as a natural scaffold in order to create new binding activities, as already described for animal toxins (*48*, *49*) and other disulfide-constrained peptides (*50*, *51*). The linking loop of MCoTI-II might be a good candidate for such a modification as it is absent in most squash TIs and, therefore, probably not involved in peptide folding. Some TIs have already been synthesized by the solid-phase method, in high yield as in the case of EETI-II (*15*). Although the cyclic nature of MCoTI-II represents a challenge, methods for chemical synthesis of macrocyclic peptides with three disulfide bridges have recently been proposed (*46*, *52*, *53*).

ACKNOWLEDGMENT

We are grateful to Yves Pétillot and Christine Saint-Pierre for ES-MS measurements, and to Christopher White and Gérard Arlaud for their stylistic revision of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Three tables, S1−S3, complete ES-MS and N-terminal sequencing data for MCoTI-II and isoforms, MCoTI-I and MCoTI-III (3 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Laskowski, M., Jr., and Kato, I. (1980) *Annu. Rev. Biochem. 49*, 593−626.
2. Wieczorek, M., Otlewski, J., Cook, J., Parks, K., Leluk, J., Wilimowska-Pelc, A., Polanowski, A., Wilusz, T., and Laskowski, M., Jr. (1985) *Biochem. Biophys. Res. Commun. 126*, 646−652.
3. Polanowski, A., Wilusz, T., Nienartowicz, B., Cieslar, E., Slominska, A., and Nowak, K. (1980) *Acta Biochim. Pol. 27*, 371−381.
4. Favel, A., Mattras, H., Coletti-Previero, M.-A., Zwilling, R., Robinson, E. A., and Castro, B. (1989) *Int. J. Pept. Protein Res. 33*, 202−208.
5. Otlewski, J., and Krowarsch, D. (1996) *Acta Biochim. Pol. 43*, 431−444.
6. Chiche, L., Gaboriaud, C., Heitz, A., Mornon, J.-P., Castro, B., and Kollman, P. A. (1989) *Proteins: Struct., Funct., Genet. 6*, 405−417.
7. Heitz, A., Chiche, L., Le-Nguyen, D., and Castro, B. (1989) *Biochemistry 28*, 2392−2398.
8. Bode, W., Greyling, H. J., Huber, R., Otlewski, O., and Wilusz, T. (1989) *FEBS Lett. 242*, 285−292.
9. Le-Nguyen, D., Heitz, A., Chiche, L., Castro, B., Boigegrain, R.-A., Favel, A., and Coletti-Previero, M. A. (1990) *Biochimie 72*, 431−435.
10. Pallaghy, P. K., Nielsen, K. J., Craik, D. J., and Norton, R. S. (1994) *Protein Sci. 3*, 1833−1839.
11. Saether, O., Craik, D. J., Campbell, I. D., Sletten, K., Juul, J., and Norman, D. G. (1995) *Biochemistry 34*, 4147−4158.
12. Daly, N. L., Koltay, A., Gustafson, K. R., Boyd, M. R., Casas-Finet, J. R., and Craik, D. J. (1998) *J. Mol. Biol. 285*, 333−345.
13. Hara, S., Makino, J., and Ikenaka, T. (1989) *J. Biochem. (Tokyo) 105*, 88−92.
14. Kupryszewski, G., Ragnarsson, U., Rolka, K., and Wilusz, T. (1986) *Int. J. Pept. Protein Res. 27*, 245−250.
15. Le-Nguyen, D., Nalis, D., and Castro, B. (1989) *Int. J. Pept. Protein Res. 34*, 492−497.
16. Chen, X. M., Qian, Y. W., Chi, C. W., Gan, K. D., Zhang, M. F., and Chen, C. Q. (1992) *J. Biochem. (Tokyo) 112*, 45−51.
17. Topczewska, J., Rempola, B., and Fikus, M. (1996) *Acta Biochim. Pol. 43*, 255−264.
18. Jaskiewicz, A., Lis, K., Rozycki, J., Kupryszewski, G., Rolka, K., Ragnarsson, U., Zbyryt, T., and Wilusz, T. (1998) *FEBS Lett. 436*, 174−178.
19. Kojima, S., Miyoshi, K., and Miura, K. (1996) *Protein Eng. 9*, 1241−1246.
20. Favel, A., Le-Nguyen, D., Coletti-Previero, M. A., and Castro, B. (1989) *Biochem. Biophys. Res. Commun. 162*, 79−82.
21. Le-Nguyen, D., Mattras, H., Coletti-Previero, M.-A., and Castro, B. (1989) *Biochem. Biophys. Res. Commun. 162*, 1425−1430.
22. Christmann, A., Walter, K., Wentzel, A., Kratzner, R., and Kolmar, H. (1999) *Protein Eng. 12*, 797−806.
23. Cheung, S. C., and Li, N. H. (1985) in *Chinese Medicinal Herbs of Hong Kong*, Vol. 4, p 146, Commercial Press, Hong Kong.
24. Hanspal, J. S., Bushell, G. R., and Ghosh, P. (1983) *Anal. Biochem. 132*, 288−293.
25. Pham, T.-C., Konopska-Waliszkiewicz, L., and Leluk, J. (1986) *Biochem. Physiol. Pflanz. 181*, 565−569.
26. Hernandez, J.-F., Bersch, B., Pétillot, Y., Gagnon, J., and Arlaud, G. J. (1997) *J. Pept. Res. 49*, 221−231.
27. Podell, D. N., and Abraham, G. N. (1978) *Biochem. Biophys. Res. Commun. 81*, 176−185.
28. Sali, A., and Blundell, T. L. (1993) *J. Mol. Biol. 234*, 779−815.
29. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, J. J., Vincent, M., Crowley, M., Ferguson, D., Radmer, R. J., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1997) AMBER 5, University of California, San Francisco.
30. Gustafson, K. R., Sowder, R. C., II, Henderson, L. E., Parsons, I. C., Kashman, Y., Cardellina, J. H., II, McMahon, J. B., Buckheit, R. W., Jr., Pannell, L. K., and Boyd, M. R. (1994) *J. Am. Chem. Soc. 116*, 9337−9338.
31. Witherup, K. M., Boguski, M. J., Anderson, P. S., Ramjit, H., Ransom, R. W., Wood, T., and Sardana, M. (1994) *J. Nat. Prod. 57*, 1619−1625.
32. Claeson, P., Goransson, U., Johansson, S., Luijendijk, T., and Bohlin, L. (1998) *J. Nat. Prod. 61*, 77−81.
33. Goransson, U., Luijendijk, T., Johansson, S., Bohlin, L., and Claeson, P. (1999) *J. Nat. Prod. 62*, 283−286.
34. Craik, D. J., Dally, N. L., Bond, T., and Waine, C. (1999) *J. Mol. Biol. 294*, 1327−1336.
35. Luckett, S., Garcia, R. S., Barker, J. J., Konarev, A. V., Shewry, P. R., Clarke, A. R., and Brady, R. L. (1999) *J. Mol. Biol. 290*, 525−533.
36. Ling, M.-H., Qi, H.-Y., and Chi, C.-W. (1993) *J. Biol. Chem. 268*, 810−814.
37. Daly, N. L., Love, S., Alewood, P. F., and Craik, D. J. (1999) *Biochemistry 38*, 10606−10614.
38. Tang, Y.-Q., Yuan, J., Ösapay, G., Ösapay, K., Tran, D., Miller, C. J., Ouellette, A. J., and Selsted, M. E. (1999) *Science 286*, 498−502.
39. Huang, B., Ng, T. B., Fong, W. P., Wan, C. C., and Yeung, H. W. (1999) *Int. J. Biochem. Cell Biol. 31*, 707−715.
40. Gran, L. (1973) *Lloydia 36*, 174−178 and 207−208.
41. Derua, R., Gustafson, K. R., and Pannell, L. K. (1996) *Biochem. Biophys. Res. Commun. 228*, 632−638.
42. Heitz, A., Chiche, L., Le-Nguyen, D., and Castro, B. (1995) *Eur. J. Biochem. 233*, 837−846.
43. Le-Nguyen, D., Heitz, A., Chiche, L., el Hajji, M., and Castro, B. (1993) *Protein Sci. 2*, 165−174.
44. Chiche, L., Heitz, A., Padilla, A., Le-Nguyen, D., and Castro, B. (1993) *Protein Eng. 6*, 675−682.
45. Heitz, A., Le-Nguyen, D., and Chiche, L. (1999) *Biochemistry 38*, 10615−10625.
46. Tam, J. P., Lu, Y.-A., Yang, J.-L., and Chiu, K.-W. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 8913−8918.

47. Hancock, R. E., Falla, T., and Brown, M. (1995) *Adv. Microb. Physiol. 37*, 135−175.
48. Vita, C., Roumestand, C., Toma, F., and Mènez, A. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6404−6408.
49. Vita, C. (1997) *Curr. Opin. Biotechnol. 8*, 429−434.
50. Smith, J. W., Tachias, K., and Madison, E. L. (1995) *J. Biol. Chem. 270*, 30486−30490.
51. Vella, F., Hernandez, J.-F., Molla, A., Block, M. R., and Arlaud, G. J. (1999) *J. Pept. Res. 54*, 415−426.
52. Tam, J. P., and Lu, Y.-A. (1998) *Protein Sci. 7*, 1583−1592.
53. Tam, J. P., Lu, Y.-A., and Yu, Q. (1999) *J. Am. Chem. Soc. 121*, 4316−4324.
54. Stachowiak, D., Polanowski, A., Bieniarz, A., and Wilusz, T. (1996*) Acta Biochim. Pol. 43*, 507−513.
55. Haldar, U. C., Saha, S. K., Beavis, R. C., and Sinha, N. K. (1996) *J. Protein Chem. 15*, 177−184.

BI9929756