Circular Proteins in Plants

SOLUTION STRUCTURE OF A NOVEL MACROCYCLIC TRYPsin INHIBITOR FROM MOMORDICA COCHINCHINENSIS* [S]

Received for publication, February 22, 2001, and in revised form, April 2, 2001
Published, JBC Papers in Press, April 5, 2001, DOI 10.1074/jbc.M101666200

Maria Elena Felizmenio-Quimio, Norelle L. Daly, and David J. Craik‡
From the Institute for Molecular Bioscience, University of Queensland, Brisbane, 4072 Queensland, Australia

Much interest has been generated by recent reports on the discovery of circular (i.e. head-to-tail cyclized) proteins in plants. Here we report the three-dimensional structure of one of the newest such circular proteins, MCoTI-II, a novel trypsin inhibitor from Momordica cochinensis, a member of the Cucurbitaceae plant family. The structure consists of a small β-sheet, several turns, and a cystine knot arrangement of the three disulfide bonds. Interestingly, the molecular topology is similar to that of the plant cyclotides (Craik, D. J., Daly, N. L., Bond, T., and Waine, C. (1999) J. Mol. Biol. 294, 1327–1336), which derive from the Rubiaceae and Violaceae plant families, have antimicrobial activities, and exemplify the cystic cyclic knot structural motif as part of their circular backbone. The sequence, biological activity, and plant family of MCoTI-II are all different from known cyclotides. However, given the structural similarity, cyclic backbone, and plant origin of MCoTI-II, we propose that MCoTI-II can be classified as a new member of the cyclotide class of proteins. The expansion of the cyclotides to include trypsin inhibitory activity and a new plant family highlights the importance and functional variability of circular proteins and the fact that they are more common than has previously been believed. Insights into the possible roles of backbone cyclization have been gained by a comparison of the structure of MCoTI-II with the homologous acyclic trypsin inhibitors CMTI-I and EETI-II from the Cucurbitaceae plant family.

Over the last few years a number of small disulfide-rich proteins that have a head-to-tail cyclized peptide backbone have been discovered in plants. These circular proteins include kalata B1 (1), the circulins (2, 3), and cyclopsychotride A (4) from the Rubiaceae family, and various peptides from plants in the Violaceae family (3, 5–8). They contain ~30 amino acids, including six highly conserved cysteine residues that form three structurally important disulfide bonds. The latter contribute to the molecules having well defined and stable three-dimensional structures (8). Because of their well defined structures and potent biological activities, the molecules may be regarded as miniproteins. We recently proposed that the known circular proteins most likely form part of a large family of proteins that we refer to as the plant cyclotides (8). They contain the conserved cysteine spacing CX₃–CX₄–CX₄–7–CX₁–CX₄–5–CX₅–7 within their circular backbone.

The plant cyclotides display a diverse range of biological activities ranging from uterotonic action (kalata B1 (1)), to anti-HIV activity (circulins (2)), to hemolytic activity (5, 9, 10), and inhibition of neurotensin binding (cyclopsychotride A (4)). Although their precise role in plants has not yet been reported, it appears that they are most likely present as defense molecules. All of their biological activities in mammalian systems seem to be related in one way or another to membrane interactions, a common feature of plant defense molecules. Recently the cyclotides have also been reported to display a wide range of anti-microbial activities against Gram-positive and Gram-negative bacteria, and against certain fungi (9), further enhancing the likelihood that they are indeed plant defense molecules.

The three-dimensional structures of three members of the plant cyclotide family have now been determined, including the prototypic member kalata B1, the first macrocyclic peptide for which a full three-dimensional structure was determined (1), circulin A (11), and cycloviolacin O1 (8). The structures are all highly conserved and contain a distorted triple stranded β-sheet, perhaps better described as a well defined β-hairpin with a third strand forming a β-bulge (8). The three disulfide bonds are arranged in a so called cystine knot, in which an embedded ring in the structure formed by two disulfide bonds and their connecting backbone segments forms a ring that is penetrated by the third disulfide bond (12–15). Although a seemingly unlikely structural motif, the cystine knot has now been seen in a variety of small disulfide-rich proteins from both the plant and animal kingdoms (15), and seems to be associated with exceptionally high stability and resistance to proteolytic cleavage of molecules containing it. This is particularly so for the cyclotides, which contain a cyclic cystine knot motif (8).

Recently Hernandez et al. (16) reported two new macrocyclic peptides derived from Momordica cochinensis, a vine plant from the Cucurbitaceae plant family. The two molecules, named MCoTI-I and MCoTI-II are trypsin inhibitors and comprise 34 amino acids, a cyclized peptide backbone, and three disulfide bonds. They are homologous to a large range of open-chain trypsin inhibitors from plants (16), but are the first reported macrocyclic trypsin inhibitors (although note that a smaller, single disulfide-containing cyclic trypsin inhibitor was recently reported from sunflower seeds (17)). This exciting discovery supports our view that macrocyclic peptides are much simpler coordinates and structure factors (code 1HB) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

The on-line version of this article (available at http://www.jbc.org) contains Tables I–3 and Fig. 1.

‡ Australian Research Council Senior Fellow. To whom correspondence should be addressed: Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD 4072, Australia. Tel.: 61-7-3365-4945; Fax: 61-7-3365-2487; E-mail: d.craik@imb.uq.edu.au.

*This work was supported in part by a grant from the Australian Research Council (to D. J. C.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

© 2001 by The American Society for Biochemistry and Molecular Biology, Inc.
The numbering scheme used for MCoTI-II throughout the text is based on the numbering scheme previously used for kalata B1 (8). The numbering scheme used for MCoTI-II throughout the text is based on the numbering scheme previously used for kalata B1 (8). The numbering scheme used for MCoTI-II throughout the text is based on the numbering scheme previously used for kalata B1 (8). The residues are numbered by the three-dimensional structure determined using NMR spectroscopy and similarity to the plant cyclotides and the six backbone loops that make up the three-dimensional structure of the new macrocyclic peptides is a template in pharmaceutical design applications is thus very attractive (19). The three-disulfide knotted framework as a template in pharmaceutical design applications is thus very attractive (19). The three-disulfide knotted framework as a template in pharmaceutical design applications is thus very attractive (19). The three-dimensional structure of MCoTI-II has been modeled (16) based on its sequence similarity to open chain trypsin inhibitors, however, no experimental determination of the three-dimensional structure of the new macrocyclic peptides is available. We were interested to see whether this class was more common in plants than has previously been realized, and that many more are likely to be discovered over coming years. The sequences of the two new trypsin inhibitors, MCoTI-I and MCoTI-II, show little homology with the known plant cyclotides apart from the six-conserved cysteine residues and the macrocyclic peptide backbone, as seen by a comparison of MCoTI-II and kalata B1 in Fig. 1. Note that because the molecules are cyclic the choice of numbering of residue 1 is arbitrary but throughout the text we have adopted the numbering scheme used by Hernandez et al. (16). In addition, we adopted our earlier nomenclature for defining the six cysteine residues (labeled I-VI) and the six backbone loops that make up the sequences between the cysteine residues (loops 1–6). This numbering scheme is defined in Fig. 1.

The three-dimensional structure of MCoTI-II has been modeled (16) based on its sequence similarity to open chain trypsin inhibitors, however, no experimental determination of the three-dimensional structure of the new macrocyclic peptides is available. We were interested to see whether this class was similar to the plant cyclotides and contained the cystine knot motif. We therefore isolated and purified a sample of MCoTI-II from M. cochinchinensis and herewith report the three-dimensional structure determined using NMR spectroscopy and simulated annealing calculations. One of the motivating factors for examining structures of this class of molecules is that they are known to be highly stable and, indeed, in the case of kalata B1, are apparently orally bioavailable, a most unusual feature for peptide-based molecules. This bioavailability is demonstrated in the native medicinal applications of Oldenlandia affinis, the plant from which kalata B1 is derived. Women of certain tribes in the Congo region ingest a tea made from the plant to accelerate labor (18). The possibility of using the cyclic cystine knot macrocyclic disulfide knotted framework as a template in pharmaceutical design applications is thus very attractive (19). The three-dimensional structure reported here provides a basis for these applications, as well as demonstrating that the new trypsin inhibitors form part of the cyclotide family.

EXPERIMENTAL PROCEDURES

Isolation and Purification of MCoTIs—Ripe frozen fruits of M. cochinchinensis were purchased from a Vietnamese store in Melbourne, Australia. The dormant seeds were homogenized using a blender and extracted with 20 mM sodium acetate. The fine debris was removed using a cotton plug. 40% Acetone, H₂O (v/v) solution was added to the mixture in a ratio of ~1:1 (v/v). Acetone was reduced using a rotary evaporator after which the mixture was cooled down to 4 °C and centrifuged for 30 min. The supernatant was then passed through a filter before purification using preparative RP-HPLC on a Vydac C18 column. Gradients of 0.1% aqueous trifluoroacetic acid and 99% acetonitrile, 0.09% trifluoroacetic acid were employed with a flow rate of 8 ml/min, and the eluant monitored at 230 nm. Further purification was performed using semipreparative RP-HPLC on a Vydac C18 column. The final purity was examined with analytical RP-HPLC and electrospray mass spectrometry.

NMR Spectroscopy—Samples for 1H NMR measurements contained 1 mM peptide in either 99.99% D₂O or 90% H₂O, 10% D₂O (v/v) at pH 3.5. Spectra were recorded at 288, 293, and 298 K on a Bruker AVANCE 750 MHz spectrometer. The two-dimensional NMR experiments employed were similar to those used for the related cyclic peptide circulin A (11) and included DQF-COSY, E-COSY, and TOCSY using a MLEV-17 spin lock sequence with a mixing time of 80 ms, and NOESY with mixing times of 150, 200, and 300 ms. Slowly exchanging NH protons were detected by acquiring a series of one-dimensional and TOCSY spectra of the fully protonated peptide immediately following dissolution in D₂O. JNAT-HT coupling constants were measured from E-COSY spectra and JHT-HT coupling constants were measured from DQF-COSY spectra, transformed over the fingerprint region to 8 × 1 K. Spectra were processed according to methods given in Daly et al. (11). Chemical shifts were referenced to sodium 2,2-dimethyl-2-silapentane-5-sulfonate at 0.00 ppm and the spectra were analyzed and assigned using XEASY (20).

Structural Restraints—Peak volumes were integrated using XEASY and distance restraints obtained from CALIBA with appropriate pseudoatom corrections. Backbone dihedral restraints were inferred from JNAT-HT coupling constants, with φ restrained to 120 ± 30° or a JNAT-HT greater than 9 Hz, and 65 ± 30° for a JNAT-HT less than 6 Hz. Stereoepecific assignments and φ, angle restraints were obtained for 7 residues using coupling constants measured in an ECOSY spectrum in conjunction with NOE intensities. Both proline residues were in the trans-conformation based on characteristic NOEs. The β-methylene protons of the proline residues were stereospecifically assigned based on the peak intensities in the NOESY spectra. Six hydrogen bonds were determined based on slow exchange data and preliminary structure calculations. Hydrogen bond restraints of 1.7–2.2 and 2.7–3.2 Å were used for the HN–O and NH–O distances, respectively.

Structure Calculations—Three-dimensional structures were initially calculated using DYANA (21) to check NOE restraints for violations and resolve ambiguous cross-peaks. Dihedral angle and hydrogen bond restraints were subsequently included in these structure calculations. The final 50 structures were generated using a torsion angle dynamics protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24). Refinement of these structures was achieved using the conjugate gradient Powell protocol (22, 23) within the program X-PLOR 3.851 (24).

RESULTS

The seeds of M. cochinchinensis were extracted with sodium acetate as described previously (16) and the extract was purified using RP-HPLC. Various peptides were isolated and their molecular weights were analyzed by mass spectrometry. Molecular weights consistent with the previously isolated macrocyclic peptides, MCoTI-I (3480) and MCoTI-II (3453), were present in addition to isomers corresponding to the β-aspartyl (3453) and succinimide (3435.5) forms of MCoTI-II (16). In the previous study (16) the dipeptide bond between Asp-4 and Gly-5 was shown to be susceptible to succinimide and β-aspartyl formation and this was confirmed in the current study. Sufficient quantities of MCoTI-II and its β-aspartyl isomer were isolated to allow characterization with NMR spectroscopy. Amino acid sequencing was performed essentially using NMR spectroscopy, as the spin systems observed in TOCSY spectra and the sequential connectivities observed in NOESY spectra were completely consistent with those of the published

The abbreviations used are: RP-HPLC, reversed-phase high performance liquid chromatography; MCo, Momordica cochinchinensis; TOCSY, two-dimensional total correlation spectroscopy; NOE, nuclear Overhauser effect; NOESY, two-dimensional NOE spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; RMDSD, root mean square deviation.
sequence (16). Assignments were made using established techniques (29) and the 1H chemical shifts are supplied as supplementary material. The fingerprint regions of a TOCSY and NOESY spectra of MCoTI-II are given in Fig. 2. The amide region is very well dispersed and this facilitated the assignment process. The excellent dispersion and large number of cross-peaks in the NOESY spectrum also provided the first indication of a well defined three-dimensional structure.

As the β-aspartyl homologue of MCoTI-II has an identical molecular weight to the normal (α-aspartyl) isomer, NMR spectroscopy was used to discriminate between these molecules. Only one form (i.e. that eluting later on HPLC) displayed the expected sequential NOE connectivities (Fig. 3a) between residues 4 and 5, which strongly suggests this form contains the “normal” peptide backbone. Comparison of the 1H chemical shifts of both isomers (Fig. 3b) suggests that the molecules have very similar structures, with the exception of the region surrounding residues 4 and 5. Having established the similarity of the two forms, in the current study we have determined the three-dimensional structure of the isomer of MCoTI-II with the α-aspartyl configuration.

Analysis of the NOE data, coupling constants and slow exchange amide protons (Fig. 4a) allowed the secondary structure to be elucidated as shown in Fig. 4b. A β-hairpin is present within residues 26–34 and there are β-sheet interactions between this hairpin and residues 13–15. An analysis of secondary structure and indicated the presence of β-strands but no helical regions.

The three-dimensional structure of MCoTI-II was determined using torsion angle dynamics coded within the program XPLOR (24). The disulfide bonding pattern had not been determined chemically and so it was important to define this using NMR methods. For small disulfide-rich molecules in which the disulfide bonds form a compact core such an approach is often the only one available due to the difficulty of obtaining suitable cleavage products. Preliminary structure calculations without any disulfide bonds included as restraints were consistent with the disulfide connectivities being identical to those in linear squash trypsin inhibitors (31) previously characterized and those in the cyclotide peptides (8) (i.e. Cys I-Cys IV, Cys II-Cys V, and Cys III-Cys VI). This conclusion was based on an analysis of distances between all possible pairs of S atoms in the preliminary structures (Supplementary Material). This clearly shows that the optimal S-S pairing corresponds to the Cys I-Cys IV, Cys II-Cys V, and Cys III-Cys VI connectivity. Subsequent structure calculations were performed with the disulfide bonds formally connected in this way and these yielded structures of low energy and excellent covariant geometry. Combined with the consistency with the homologous squash inhibitors and the plant cyclotides this was taken as strong evidence for the proposed disulfide connectivity. Hydrogen bonds were also inferred from the preliminary structure calculations in conjunction with the slow exchange data, and all were consistent with the β-sheet shown in Fig. 4b.

The family of structures for MCoTI-II is well defined over most of the molecule (residues 8–34), however, there are few medium or long range NOEs for the first seven residues and this results in a lack of definition in this region (Fig. 5a). This is reflected in the angular order parameters (Supplementary Material) that are generally >0.9 for the φ and ψ angles for residues 8–34, but vary between 0.2 and 0.9 for residues 1–7. Geometric and energetic statistics that define the 20 lowest energy structures are given as supplementary data. The RMSD’s for residues 8–34 are 0.70 ± 0.17 and 1.86 ± 0.38 Å for

![Fig. 2. NMR spectra of MCoTI-II. Regions of the 50-ms TOCSY spectrum (top) and 200 ms NOESY spectrum of MCoTI-II in 90% H2O, 10% D2O at 293 K. Spin systems are shown in the TOCSY spectrum and the sequential connectivities in the aH-NH region of the NOESY spectrum. The one-letter code for amino acids combined with the residue number is used to denote assignment. The excellent signal dispersion is indicative of a well defined three-dimensional structure.](image1)

![Fig. 3. Comparison of the α and β aspartyl isomers of MCoTI-II. a, structure of α and β aspartic acid residues and sequential NOEs expected to be observed in the α-aspartyl form. b, comparison of the α-proton chemical shifts of the two isomers of MCoTI-II. The fact that the differences occur only near residue Asp-5 suggests that isomerization of this residue (α/β aspartic acid) is responsible for the differences between the two isolated peptides.](image2)
the backbone atoms and all heavy atoms, respectively. The corresponding values are 0.45 ± 0.11 and 1.56 ± 0.22 Å for superimposition over residues 12–34.

The major features of the three-dimensional structure are a β-sheet region and several turns. Analysis of the secondary structure elements in the final family of calculated structures with PROMOTIF (27) reveals a γ turn between residues 18 and 20 and β-turns between residues 15–18 and 22–25. A β-hairpin is formally recognized by PROMOTIF between residues 26 and 34, as expected from the preliminary secondary structure analysis, and a type I β-turn between residues 28 and 31 connects the strands. The β-sheet interactions observed between the hairpin and residues 13–15 (Fig. 4b) are not formally recognized by PROMOTIF as part of a triple-stranded β-sheet. However, this is also found for other related small disulfide-rich peptides and is hence not surprising. In the plant cyclotides (19) the third β-strand is only loosely connected to the well defined β-hairpin. Thus, despite the presence of NOEs and slowly exchanging amide protons characteristic of β-sheets, the orientations of the strands in the structure calculations are such that they are not formally recognized as part of a triple-stranded β-sheet in all of the structures.

The disulfide bonds of MCoTI-II form a cystine knotted structure similar to that in the plant cyclotides. In this arrangement the embedded ring formed by Cys I-Cys IV and Cys II-Cys V and their connecting backbone segments is threaded by the third disulfide bond, Cys III-Cys VI. A schematic representation of the secondary structure and the cystine knot of MCoTI-II is given in Fig. 5b.

Since the peptide backbone of MCoTI-II is cyclic, a convenient way to describe the structural features is in terms of the six segments, or loops, that are flanked by pairs of successive Cys residues. Fig. 6 shows the backbone considered in this way and highlights several features of the derived structures, including the location of the three β-strands and the various turns. The disulfide bond connectivity is also shown. Loop 1, between Cys I and Cys II contains a highly positively charged (+3) hexapeptide segment which, based on homology with cor-
responding acyclic trypsin inhibitors, comprises the putative active site residues Lys/Ile. Loop 2 contains two positive and two negative charges and includes a β-turn between residues 15–18 and a γ turn between residues 18 and 20. Loop 3 contains no charged residues but incorporates a β-turn within its PGA sequence. Loop 4 contains a single amino acid (I) and is incorporated within one of the β-strands that make up the β-hairpin. It is followed by loop 5, which contains a type I β-turn and leads into loop 6, which incorporates the second strand of the β-hairpin. In the final folded structures the disulfide bonds “pinch” the circular backbone and loops 2 and 5 are drawn very close to one another.

Loop 6 protrudes from the structure more than any of the other loops and is also the most disordered. Fig. 5 shows that individual structures vary widely in their placement of this loop relative to the core of the molecule. This may reflect intrinsic flexibility or simply a lack of NOEs defining this region. Loop 6 contains a high Gly content and hence the number of possible protons, and hence NOEs, is reduced relative to larger amino acids. To assess whether it was lack of NOEs, or flexibility that results in the disorder for loop 6, the chemical shift separation of the β-protons of various amino acid side chains was examined. In general, rigid conformations are associated with widely separated β-proton shifts while, in contrast, degenerate β-proton shifts can be indicative of conformational averaging. For example, in an analysis of acyclic permутants of kalata B1, those permutants with decreased stability, based on amide exchange rates, displayed a decreased shift dispersion for cysteine β-protons (32).

Analysis of the β-proton chemical shift separations for the six cysteine residues of MCoTI-II supports the proposal of flexibility in loop 6. The disulfide bond connecting residues 8 to 25 is the closest disulfide bond to the disordered region and the β-protons are essentially degenerate for both of these residues, suggesting conformational averaging in this region of the molecule. By contrast, the β-shifts for the remaining four cysteine residues are separated by 0.17 to 0.34 ppm and correspond to well defined regions of the molecule. The flexibility in this putative linker region was predicted in the model determined by Hernandez et al. (16). The model also suggested contacts between Gly-6 and Ala-24, and indeed NOEs are observed between these residues in our study. However, the possible hydrogen bonding predicted between Ser-1 and Arg-28 was observed in very few of the NMR structures.

A comparison of MCoTI-II with the trypsin inhibitor CMTI-I (31) and the plant cyclotide, kalata B1, is given in Fig. 7. It is apparent the proteins possess similar topologies. For example, the backbone atoms of the cysteine residues of MCoTI-II and CMTI-I superimpose with an RMSD of 0.98 Å. The equivalent
RMSD for kalata B1 and MCoTI-II is much higher (6.02 Å) because of the differences in spacings between the cysteine residues. However, when superimposed over the backbone atoms of the four cysteine residues that make up the ring of the cystine knot the RMSD value is 1.83 Å. Another trypsin inhibitor, EETI-II (33), also has a similar structure to MCoTI-II and superimposes with an RMSD of 1.37 Å over the backbone atoms of the cysteine residues.

**DISCUSSION**

In the current study we have determined the three-dimensional structure of MCoTI-II, a 34-amino acid trypsin inhibitor from *M. cochinchinensis*. It has the unusual feature of a head-to-tail cyclized peptide backbone, and thus joins the ranks of an increasing number of small circular proteins discovered over the last five years. The main element of secondary structure is a β-hairpin and associated cystine knot framework. A third strand of β-sheet is loosely connected with this hairpin and the molecule also contains several turns. The overall fold is very similar to that of the squash trypsin inhibitors EETI-II and CMTI-I. Both have high sequence homology with MCoTI-II but are “conventional” proteins in that they lack the head-to-tail cyclization of MCoTI-II.

The similarity of the overall fold of the cyclic inhibitor to those of linear homologues raises the question of what is the advantage/biological role of cyclization? Several possibilities could be contemplated, including factors relating to conformational or dynamic stabilization, changes in physicochemical properties, addition of new functionality, or protection from proteolytic cleavage. We examine these possibilities in turn and then compare this newly discovered cyclic plant protein with the previously reported cyclotide family (8) of macrocyclic knotted peptides.

Possible conformational effects of cyclization could include stabilization of the active site residues into a preferred conformation for binding, or entropic stabilization associated with the active site conformation. It has been suggested that the mechanism of inhibition of the MCo peptides is identical to that for other squash trypsin inhibitors, based on the sequence conservation in the putative inhibitory loop (16). The local conformations of the residues presumed to be at the active site are summarized in Table I for MCoTI-II and a range of linear homologues, both in solution and complexed to trypsin. Comparison of a linear squash inhibitor, CMTI-I (31), with the cyclic MCoTI-II reveals similar backbone angles in solution, with the most significant difference being in the ψ angle of the P1 residue (Lys-10) and to a lesser extent in the φ angle of the P1′ residue. The solution structures of MCoTI-II and CMTI-I are also very similar to the bound structures listed in Table I with the exception of the ψ angle of the P1 residue. The lack of a major conformational change between MCoTI-II and linear homologues, either in solution or bound states, suggests that cyclization does not play a significant role in determining the active site conformation.

The final inhibitor listed in Table I, sunflower trypsin inhibitor I differs significantly from the others in that it is much smaller (14 residues), contains a single disulfide bond and unlike all except MCoTI-II, is cyclic (17). It is a particularly potent trypsin inhibitor despite this small size and hence provides an opportunity to assess the essential residues for activity. Its three-dimensional structure in complex with bovine β-trypsin has been determined using x-ray crystallography (17) and the active site residues are in similar positions to those observed for the other squash inhibitors (Table I). The fact that potent trypsin inhibition can be obtained with such a small inhibitor is consistent with the fact that peripheral regions of the MCoTI-II inhibitor are not essential for binding activity and function, and hence play other roles. Conformational or entropic roles for the loop 6 linker region have been discounted and hence factors relating to additional functionality or modulation of biophysical properties can be considered.

Another possible role for cyclization is that of protection from proteolytic cleavage. Indeed, based on the discussion above, which suggests that conformational or entropic roles are unlikely, this seems the most likely function of cyclization in MCoTI-II. Cyclization of small peptides is a well established method used in the pharmaceutical industry to protect small bioactive peptides against cleavage by *in vivo* exopeptidases. Being a seed component, the MCoTI-II peptide most likely has to survive intact for long periods prior to germination and its function would be compromised if broken down by exopro-
Solution Structure of MCoTI-II

Table I

Backbone angles (°) of MCoTI-II and selected trypsin inhibitors

| Inhibitor | P2  | P2  | P1  | P1  | P2' | P2' |
|-----------|-----|-----|-----|-----|-----|-----|
| Solution state |     |     |     |     |     |     |
| MCoTI-II   | −127| 138 | −71 | 153 | −80 | 96  |
| MCTI-I     | 67  | 123 | −74 | 137 | −59 | −37 |
| Complexed* |     |     |     |     |     |     |
| MCTI-A     | −137| 113 | −55 | 148 | −91 | 35  |
| CPTI-II    | −122| 110 | −57 | 149 | −90 | 30  |
| SFTI-I     | −126| 108 | −57 | 149 | −94 | 39  |
| LDIT       | −141| 153 | −70 | 155 | −91 | 38  |
|            | −143| 144 | −74 | 166 | −104| 33  |

* The complexed structures were determined using x-ray crystallography with inhibitors complexed to trypsin. The Protein Data Bank accession codes are CMTI-I (solution), 2cti (31); MCTI-A, 1met (40); CMTI-I (complexed), 1ppe (41); CPTI-II, 2btc (42); SFTI-I, 1sfi (17); LDIT, 1anl (43).

Fig. 8. The three-dimensional structure of MCoTI-II shown in CPK format. The negatively charged residues are in red, positively charged in dark blue, hydrophobic residues in green, polar residues in light blue, and cysteine residues in yellow. Surface exposed hydrophobic residues appear mainly on one face and the other face contains most of the positively and negatively charged residues. The views are rotated 180° about the vertical axis with respect to each other. The diagram was generated using MOLMOL (39).

The previously reported cyclotides can be regarded as part of the same family of proteins. Thus the three-dimensional structure and cyclic backbone suggests they have been known for many years, however, so far there has been no evidence for linear forms of the cyclotides. Linear squash inhibitors have been known for many years, however, so far there has been no indication of linear counterparts for the cyclotides. Linear squash inhibitors have been known for many years, however, so far there has been no evidence for linear forms of the cyclotides. This may be a consequence of the isolation techniques that have been employed to date, or may reflect highly efficient processing of the cyclotides from putative linear precursors. Despite these differences between MCoTI-II and the plant cyclotides the similar three-dimensional structure and cyclic backbone suggests they can be regarded as part of the same family of proteins. Thus the cyclotide family can be expanded to include the new cyclic trypsin inhibitors MCoTI-I and MCoTI-II.

An interesting difference between the previously reported cyclotides and the circular squash peptides is the absence of linear counterparts for the cyclotides. Linear squash inhibitors have been known for many years, however, so far there has been no evidence for linear forms of the cyclotides. This may be a consequence of the isolation techniques that have been employed to date, or may reflect highly efficient processing of the cyclotides from putative linear precursors. Despite these differences between MCoTI-II and the plant cyclotides the similar three-dimensional structure and cyclic backbone suggests they can be regarded as part of the same family of proteins. Thus the cyclotide family can be expanded to include the new cyclic trypsin inhibitors MCoTI-I and MCoTI-II.
the use of intein-based methods for the artificial production of circular proteins (35–38).

Acknowledgments—We thank the Institute for Molecular Bioscience, a Special Research Center of the Australian Research Council.

REFERENCES

1. Saether, O., Craik, D. J., Campbell, I. D., Sletten, K., Juul, J., and Norman, D. G. (1995) *Biochemistry* **34**, 4147–4158.

2. Gustafson, K. R., Sowder, R. C. II, Henderson, L. E., Parsons, I. C., Kashman, Y., Cardellina II, J. H., McMahon, J. B., Buckheit, R. W., Jr., Pannell, L. K., and Boyd, M. R. (1994) *J. Am. Chem. Soc.* **116**, 9337–9338.

3. Gustafson, K. R., Walton, L. K., Sowder, R. C. I., Johnson, D. G., Pannell, L. K., Cardellina, J. H. I., and Boyd, M. R. (2000) *J. Nat. Prod.* **63**, 176–178.

4. Witherup, K. M., Bogusky, M. J., Anderson, P. S., Ramjit, H., Ransom, R. W., Wood, T., and Sardana, M. (1994) *J. Nat. Prod.* **57**, 1619–1625.

5. Schöpke, T., Hasan Agha, M. I., Kraft, R., Otto, A., and Hiller, K. (1993) *Sci. Pharm.* **61**, 145–153.

6. Hallock, Y. F., Sowder, R. C. I., Pannell, L. K., Hughes, C. B., Johnson, D. G., Hallock, Y. F., Sowder, R. C. I., Pannell, L. K., Hughes, C. B., Johnson, D. G., and Hendrickson, W. A. (1993) *Cell* **73**, 265–268.

7. Go¨ransson, U., Luijendijk, T., Johansson, S., Bohlin, L., and Claeson, P. (1999) *Biochemistry* **38**, 10806–10814.

8. Daly, N. L., Love, S., Alewood, P. F., and Craik, D. J. (1999) *Biochemistry* **38**, 10606–10614.

9. Tam, J. P., Lu, Y. A., Yang, J. L., and Chiu, K. W. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8913–8918.

10. Daly, N. L., Love, S., Alewood, P. F., and Craik, D. J. (1999) *Biochemistry* **38**, 10606–10614.

11. Daly, N. L., Koltay, A., Gustafson, K. R., Walton, L. K., Alexander, P. S., Sowder, R. C. I., Johnson, D. G., Pannell, L. K., and Hendrickson, W. A. (1999) *J. Mol. Biol.* **294**, 1327–1336.

12. McDonald, N. Q., and Hendrickson, W. A. (1993) *J. Mol. Biol.* **239**, 285–292.

13. Isaacs, N. W. (1995) *Biochemistry* **34**, 283–286.

14. Pallaghy, P. K., Nielsen, K. J., Craik, D. J., and Norton, R. S. (1994) *Protein Sci.* **3**, 1833–1839.

15. Craik, D. J., Daly, N. L., and Waine, C. (2000) *Biochemistry* **39**, 391–395.

16. Pauli, G. M., Bru¨nger, A. T., Karplus, M., and Gronenborn, A. M. (1986) *Acta Crystallogr. D Biol.* **42**, 139–148.

17. Heitz, A., Chiche, L., Le-Nguyen, D., and Castro, B. (1999) *J. Mol. Biol.* **294**, 1327–1336.

18. Holak, T. A., Gundel, D., Otlevski, J., and Wilusz, T. (1989) *J. Biol. Chem.* **264**, 635–648.

19. Daly, N. L., and Craik, D. J. (2000) *Biopolymers* **56**, 190–19075.

20. Heitz, A., Chiche, L., Le-Nguyen, D., and Castro, B. (1989) *Biochemistry* **28**, 2392–2398.

21. Guntert, P., Mumenthaler, C., and Wu¨ thrich, K. (1997) *FEBS Lett.* **435**, 283–292.

22. Heitz, A., Chiche, L., Le-Nguyen, D., and Castro, B. (1989) *Biochemistry* **28**, 2392–2398.

23. Ling, M. H., Qi, H. Y., and Chi, C. W. (1993) *J. Mol. Biol.* **239**, 285–292.

24. Wilusz, T., and Wachtler, J. (1989) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.

25. Wood, T., and Sardana, M. (1994) *J. Mol. Biol.* **239**, 285–292.

26. Heitz, A., Chiche, L., Le-Nguyen, D., and Castro, B. (1989) *Biochemistry* **28**, 2392–2398.

27. Scott, C. P., Abel-Santos, E., Wall, M., Wahn, D. C., and Benkovic, S. J. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13638–13643.

28. Rice, L. M., and Bru¨nger, A. T. (1994) *Acta Crystallogr. D Biol.* **51**, 51–51.

29. Heitz, A., Chiche, L., Le-Nguyen, D., and Castro, B. (1999) *Biochemistry* **28**, 2392–2398.

30. Evans, T. C., Jr., Benner, J., and Xu, M. Q. (1999) *J. Biol. Chem.* **274**, 18359–18363.

31. Iwai, H., and Pluckthun, A. (1999) *FEBS Lett.* **459**, 166–172.

32. Wilusz, T., and Wachtler, J. (1989) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.

33. Wilusz, T., and Wachtler, J. (1989) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.

34. Evans, T. C., Jr., Benner, J., and Xu, M. Q. (1999) *J. Biol. Chem.* **274**, 18359–18363.