Protein, cDNA, and Genomic DNA Sequences of the Towel Gourd Trypsin Inhibitor

A SQUASH FAMILY INHIBITOR*

Min-Hua Ling, Hai-yan Qi, and Cheng-wu Chi

From the Shanghai Institute of Biochemistry, Academia Sinica, Shanghai 200031 and the Department of Biochemistry, The Fourth Medical University, Xian 710032, China

Two trypsin inhibitor components of the squash family were isolated and purified from the juice of the towel gourd (Luffa cylindrica) using anhydrotrypsin affinity chromatography followed by high pressure liquid chromatography. The inhibitors were sequenced and found to consist of 28 and 29 amino acid residues. The determined sequences show high similarity to other inhibitors of the squash family, especially in the location of disulfide bonds and the reactive site and also in the COOH-terminal region. A cDNA library of towel gourd was constructed and used as a template for polymerase chain reaction amplification of two cDNA fragments of the inhibitor with an overlapping sequence. A full-length cDNA sequence coding for the inhibitor was then completed. The open reading frame codes for a prepro-inhibitor protein with the pre- and pro-peptides consisting of 21 and 13 residues, respectively. The deduced amino acid sequence of 29 residues for the inhibitor is consistent with that determined by primary structure analysis. The genomic sequence of the mature inhibitor was also ascertained using the total DNA of the towel gourd as a polymerase chain reaction template. The genomic sequence is completely identical with that of the cDNA, showing no intervening sequence.

* This work was supported by a state biological high technology research grant from the government of China. 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank/EMBL Data Bank with accession number(s) M98055.

§ To whom correspondence should be addressed: Shanghai Institute of Biochemistry, Academia Sinica, Shanghai 200031, China. Fax: 86-21-4336937.

Experimental procedures

Materials—Towel gourd was purchased from the market of a Shanghai suburb. Benzoyl-DL-arginine p-nitroanilide HCl was from Dongfeng Biochemical Reagent Factory; bovine trypsin was obtained from Sigma. Immobilized anhydrotrypsin was prepared according to Axen (22) and Knights (23). Restriction endonucleases, T4 ligase Taq DNA polymerase, isopropyl-1-thio-β-D-galactosidase, X-gal, and the cDNA library kit were purchased from Bethesda Research Laboratories. PCR reagents were from Perkin-Elmer Cetus Instruments. The Sequenase kit was a product of U. S. Biochemical Corp. ([α-32P]dATP and [α-32P]dCTP were purchased from Du Pont-New England Nuclear. All other reagents were of the finest commercial grade available.

Reverse-phase HPLC—HPLC was performed on a Waters system with two model 510 pumps and an ultraviolet detector (model 484). A 4.6 × 250-mm C18 column (Beckman Instruments) was employed. Elution was carried out with the trifluoroacetic acid-acetonitrile linear gradient system, namely eluting solution A was 0.1% trifluoroacetic acid in water and eluting solution B was 0.85% trifluoroacetic acid in 85% acetonitrile. The effluent was monitored by measuring the absorbance at 220 nm. The flow rate was 1 ml/min.

Trypsin Inhibitory Activity Assay—This was performed essentially according to the method described previously (8).

Amino Acid Sequence Analysis—The sequence of the inhibitor was determined by the Edman degradation method at the National Biological Information Service of Shanghai Institute of Biochemistry. The sequence of the inhibitor is consistent with that determined by primary structure analysis. The results are identical with that of the cDNA, showing no intervening sequence.
determined on an Applied Biosystems model 477 A/120 A protein sequenator and phenylthiohydantoin amino acid analyzer using the program provided by the manufacturer.

Isolation of Towel Gourd DNA and mRNA—Total DNA was isolated from fresh tender towel gourds according to a modified procedure, using NaClO4 extraction followed by RNase treatment (24). Total RNA was prepared by the guanidinium/phenol method as described (25). The RNA preparation was shown to be undegraded by formaldehyde-agarose gel electrophoresis and staining with ethidium bromide (26, 27). Poly(A)+ RNA was isolated from a total RNA extract after three cycles of affinity chromatography on oligo(dT)-cellulose.

Construction of a cDNA Library—The procedure was performed according to the recommendations of Bethesda Research Laboratories. Four micrograms of mRNA was used to synthesize double-stranded cDNA. The first strand was synthesized with reverse transcriptase using the oligo(dT) primer. The second strand was synthesized with DNA polymerase. The SalI adapter was then added to the double-stranded cDNA with T4 ligase. By SalI digestion, the transcribed cDNA thus had an SalI restriction site at one end and an NotI site at the other. The total cDNA was then cloned to plasmid pSPORT predigested with SalI and NotI and transformed into E. coli strain DH5, thus generating a library of more than 2 × 1010 ampicillin-resistant colonies.

PCR—PCR was used for amplification of the coding gene of the inhibitor from the cDNA library. The oligonucleotide primers were synthesized on an Applied Biosystems 380 A DNA synthesizer. The synthesized primers were purified on 8 M urea, 20% polyacrylamide gel. PCR was performed in a volume of 100 µl containing 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 2 mM MgCl2, 3 µg of cDNA or DNA template, 0.2 mM each dNTP, 2.5 units of Taq DNA polymerase, and 0.2 µM each oligonucleotide primer. The sample was overlaid with 100 µl of mineral oil to prevent condensation and subjected to 30 cycles of denaturation at 94 °C for 1 min, annealing at 55 or 65 °C for 1 min, and extension at 72 °C for 1 min with a final extension phase of 5 min, performed on a Perkin-Elmer Cetus Instruments thermocycler. Fifteen µl of the mixture was then applied to 7% PAGE. The PCR products were visualized with ethidium bromide staining.

Sequence Analysis of PCR Products—The PCR product of the appropriate size on PAGE was recovered by soaking the gel slice in a solution (0.5 M NH4 acetate, 10 mM MgAc2, 1 mM EDTA, pH 8.0, 0.1% SDS) with shaking at 37 °C for 4 h. The purified PCR products were then digested with two enzymes corresponding to the restriction sites at the 5’-ends of both primers and cloned into the same two sites of vector M13. The single-stranded DNA of recombinant clones containing PCR products was sequenced by Sequenase (version 2.0) using the Sanger dideoxynucleotide chain termination method (28) according to the supplier’s instructions.

RESULTS

Isolation and Purification of TGTI—Fresh towel gourd melons were peeled, homogenized into a greenish paste, and extracted with 50% ethanol at 4 °C overnight with stirring. The residue was removed by centrifugation. The supernatant was distilled under vacuum to remove ethanol and then adjusted to pH 8 and subjected to affinity chromatography in batches on Sepharose 4B-bound anhydrotrypsin previously equilibrated with 0.02 M Tris-HCl, pH 8, as described (8). The active fraction was eluted with 0.02 M HCl. The fraction collected was lyophilized and further chromatographed on a reversed-phase HPLC C18 column and fractionated by linear gradient elution. Two peaks (P1 and P2) with trypsin inhibitory activity were found as shown in Fig. 1 and designated as TGTI-I and TGTI-II. The trypsin inhibitory activity of TGTI-II was determined as shown in Fig. 2, which revealed equimolar inhibition. The same results were obtained with TGTI-I (data not shown). The total amount of inhibitors was about 1.6 mg/kg of towel gourd estimated by their inhibitory activity toward trypsin. The yield of both finally purified inhibitors was around 60%.

Amino Acid and Sequence Analyses—Both P1 and P2 components collected from HPLC as shown in Fig. 1 were taken to determine their amino acid compositions and sequences.

![Figure 1](image1.png)  
**FIG. 1.** Chromatography of the towel gourd trypsin inhibitor on an HPLC C18 column. The column (4.6 × 250 mm, 10 µm) was eluted as described under "Experimental Procedures." Peak fractions marked were pooled and used for amino acid analysis and sequencing.

![Figure 2](image2.png)  
**FIG. 2.** Inhibition curve of TGTI-II toward trypsin. Benzoyl- Dl-arginine p-nitroanilide was used as a substrate (0.8 × 10−5 M in 0.05 M Tris buffer, pH 8) with 30 µg of trypsin; reaction time, 10 min at 37 °C, measured at 405 nm.

| Table 1: Amino acid composition of the towel gourd trypsin inhibitor in mol/mol |
|---------------------------------|-----------------|-----------------|
| Amino acid | TGTI-I | TGTI-II |
| D/N | 2.8 (3) | 4.5 (5) |
| E/Q | 2.1 (2) |  |
| S | 2.7 (3) | 0.9 (1) |
| G | 1.9 (2) | 3.1 (3) |
| T |  | 1.2 (1) |
| A | 1.1 (1) |  |
| R | 0.8 (1) | 1.7 (2) |
| P | 1.9 (2) | 1.8 (2) |
| Y |  | 0.9 (1) |
| M | 0.7 (1) | 2.1 (2) |
| C | 6 (6) |  |
| I | 2.6 (3) | 1.7 (2) |
| L | 2.9 (3) | 2.9 (3) |
| F | 1.1 (1) |  |
| K | 0.6 (1) |  |
| Total | 28 | 29 |
Molecular Characterization of a Squash Family Inhibitor

The position of the conserved reactive site of the squash family of inhibitors is marked by an asterisk.

FIG. 3. Amino acid sequences of the towel gourd trypsin inhibitors and other squash family inhibitors. The results are shown in Table I and Fig. 3. TGTI-I and TGTI-II are composed of 28 and 29 amino acid residues, respectively, sharing 69% similarity and having common features of codon usage in other plant protease inhibitor genes, primer 1 was designed and synthesized corresponding to the COOH-terminal part of the inhibitor with minimal codon degeneracy. An EcoRI restriction site was added at the 5'-end of the primer (Fig. 3). Based on the determined sequence, primer 2 was designed and synthesized according to the SP6 and T7 promoter sequences in the pSPORT vector.

The results are shown in Table I and Fig. 3. TGTI-I and TGTI-II are composed of 28 and 29 amino acid residues, respectively, sharing 69% similarity and having common features of other squash family inhibitors. The reactive site of both inhibitors is Arg-Ile located near the NH$_2$ terminus.

Completion of the cDNA Sequence of TGTI-II—The cDNA sequence of the inhibitor was completed by PCR in two steps, using the total cDNA library of towel gourd constructed in pSPORT plasmid as a PCR template. The procedure is shown in Fig. 4a. Based on the determined TGTI sequences and the preference of codon usage in other plant protease inhibitor genes, primer 1 was designed and synthesized corresponding to a portion of the NH$_2$-terminal part of the inhibitor with minimal codon degeneracy. An EcoRI restriction site was added at the 5'-end of the primer (Fig. 4b). The SP6 polymerase promoter sequence present in the pSPORT vector was used as the reverse primer. The PCR-amplified product with a 350-bp fragment was separated on a 7% PAGE gel (Fig. 5) and cloned into M13 mp19 at the restriction sites of BamHI and EcoRI for sequence analysis.

The results revealed that the PCR fragment encompassed the NH$_2$-terminal part of TGTI-II (position 3–29) and a 185-bp 3'-flanking region upstream from the poly(A) tail. Based on the determined sequence, primer 2 was designed and synthesized corresponding to the COOH-terminal part of TGTI-II, i.e. Cys$^9$-Gly$^9$. A BamHI restriction site was added at the 5'-end of primer 2 (Fig. 4b). Using the same strategy, the T7 promoter sequence present in the pSPORT vector was exploited to pair with primer 2. Another PCR-amplified product with a size almost the same as that of the 350-bp product was obtained (Fig. 5). After separation on 7% PAGE and cloning into M13 mp18 with BamHI and EcoRI, sequence analysis showed that the second PCR product encoded a 5'-flanking sequence of 107 bp and the total coding sequence of TGTI-II. As a result, the full-length cDNA sequence of TGTI-II was 481 bp was completed by overlapping the sequences of the two cDNA segments as shown in Fig. 6. The open reading frame encoded a TGTI-II precursor with 63 amino acid residues. Apart from the mature inhibitor, there is a leading peptide with 34 residues upstream from the NH$_2$-terminal glycine to the initial codon methionine. The deduced amino acid sequence from the cDNA of TGTI-I was completely consistent with that determined by primary structure.
Molecular Characterization of a Squash Family Inhibitor

analysis. It is most likely that the signal peptide of the inhibitor consists of 21 residues with a hydrophobic core (from position -34 to -14) followed by a pro-peptide with 13 residues (from position -13 to -1).

**Genomic Sequence of Mature TGTI-II**—Compared with the determined cDNA sequence of TGTI-II, synthesized primer 1 is different only in four nucleotides, and its 3′-end is completely complementary to its target sequence. Thus, both primer 1 and primer 2 were again used for PCR amplification, and the total DNA of the towel gourd was used as a PCR template. An amplified PCR product with about 100 bp was obtained (Fig. 5). With the same strategy used for cDNA sequence analysis, the genomic DNA sequence of TGTI-II was determined. The result showed that the DNA sequence of the inhibitor is identical with its cDNA sequence. The finding implies that there is no intron in the gene structure of TGTI-II.

**DISCUSSION**

In this study, we describe two new members of the squash family of inhibitors from the towel gourd (Luffa cylindrica), a cucurbitaceous plant. Their structural features are quite similar to those of other squash family inhibitors. The residues located at the reactive site, the disulfide bonds, and the COOH terminus are completely conserved in all squash family inhibitors (Fig. 3). Except for the bitter gourd inhibitor (MCEI-II), the reactive site of all squash family inhibitors is either an arginine or a lysine. As a result, they are all trypsin inhibitors, whereas MCEI-II is an elastase inhibitor with a leucine reactive site. Recently, we have succeeded in converting the Trichosanthes trypsin inhibitor into a strong elastase inhibitor by protein engineering, in which the reactive site Arg-Ile was replaced by Ala-Ser (data not shown).

Until now, no cDNA sequence of any of the squash family inhibitors has been reported. To better understand the molecular characteristics of these inhibitors at the level of their genes, a cDNA library of the towel gourd was constructed. Instead of the conventional method of screening the cDNA clone by antibody or a DNA probe, which is quite time consuming and laborious, the PCR technique was used to directly amplify TGTI cDNA and TGTI DNA, respectively, from the total cDNA and DNA of the towel gourd without screening. Since the 3′-end of the primer is more important in the annealing and extension process in PCR amplification, primer 1 was designed and synthesized on the basis of the elucidated TGTI amino acid sequence and the preference in codon usage of other plant protease inhibitors. Primer 1, which corresponded to the NH₂-terminal part of TGTI, provided five nucleotides, ATGCC, of the 3′-end (Met codon + two nucleotides of the Pro codon) for exact pairing with the target template. The rest of the nucleotides of primer 1 conformed to the preferred codon usage of plant protease inhibitors. Compared with the elucidated TGTI-II cDNA sequence, there were mismatches of only 4 base pairs in primer 1 (Pro, CCC-CAA; Arg, CGC-AGA; Ile, ATT-ATC), which can be tolerated in PCR amplification.

The TGTI-II cDNA sequence comprises a 189-bp open reading frame flanked by 107 bp of the 5′-untranslated region and 185 bp of the 3′-untranslated region. As in other major plant proteins, the initiation site was found to be ATGCC (29). The 3′-untranslated region contains a polyadenylation sequence AATAAA and its poly(A) tail. However, the AATAAA sequence is found at 112 bp instead of the canonical polyadenylation site at around 20–30 bp upstream from the poly(A) tail (30). The deduced amino acid sequence of the coding region consists of 63 residues of the precursor protein (prepro-TGTI). The pre-peptide, a signal peptide, is presumed to consist of 21 amino acid residues (position -14 to -34) with a hydrophobic core necessary for entering the rough endoplasmic reticulum and cleaved at the small residue glycine (31). The pro-peptide comprises 13 residues (position -1 to -10), displaying hydrophilic features. While its functional significance remains to be clarified, it may not be too far fetched to speculate its involvement in directing the correct folding of the three disulfide bridges in TGTI, as the disulfide bridges in the squash family of inhibitors are folded in a very unique way (I-IV, II-V, and III-VI).

The PCR-amplified fragment was identified to be the TGTI-II cDNA, and no TGTI-I cDNA was found in any positive clone of the PCR fragment. It is uncertain whether this was due to the fact that TGTI-I mRNA does not appear until the towel gourd is quite ripe or the failure to detect the very low amount of TGTI-I mRNA as compared with that of TGTI-II mRNA. During purification of the two TGTI components, it was found that TGTI-II was indeed the major component.

The mature TGTI cDNA sequence is completely identical to its genomic structure, demonstrating that TGTI is encoded by a single exon without any intervening sequence, as often occurs in small molecular weight peptide genes.

**Acknowledgments**—We thank Lai-Geng Xu for amino acid sequence analysis, Jun Bian for the synthesis of primers, and Dr. Wen-Feng Xu for helpful advice.

**REFERENCES**

1. Laskowski, M., and Kato, I. (1980) *Ann. Rev. Biochem.* 49, 595–626
2. Lips, D., and Kato, I. (1979) *Int. Rev. Biochem.* 27, 97–122
3. Richardson, M. (1977) *Sci. Prog.* 61, 169–189
4. Gustafsson, G., and Ryan, C. A. (1976) *J. Biol. Chem.* 251, 7004–7010
5. Ryan, C. A. (1978) *Trends Biochem. Sci.* 5, 145–151
6. Hider, R., Gateshouse, A. M. R., Sheerman, S. E., Barker, R. F., and Boult, D. (1987) *Nature* 326, 160–163
7. Johnson, R. (1989) *Nature* 337, 39–505
8. Olewski, J., Whaley, H., Polanskowski, A., and Witosz, T. (1987) *Bioll. Chem. Hoppe-Seyler* 362, 1505–1507
9. Wielgosz, M., Olewski, J., Cook, J., Parks, K., Leiw, J., Willowska-Pelo, A., Polanskowski, A., and Witosz, T. (1989) *Biochem. Bioll. Res. Commun.* 162, 646–652
10. McWherter, C. A., Walkenhorst, W. F., Campbell, E. J., and Glover, C. J. (1989) *Biochem. Bioll. Res. Commun.* 162, 159–169
11. Huang, Q. C., Liu, S. P., Tang, Y. Q., and Qian, R. Q. (1992) *FEBS Lett.* 297, 141–146
12. Huang, Q. F., Wu, M. Z., and Chi, C. W. (1990) *Sci. China Ser. B* 33, 1129–1122

---

**Fig. 6. The nucleotide and deduced amino acid sequences of TGTI-II cDNA.** The first 21 amino acid residues correspond to a signal peptide followed by a 13-amino acid pro-peptide. The amino acid sequence of the mature inhibitor corresponds to TGTI-II. The putative polyadenylation signal AATAAA is underlined.
Molecular Characterization of a Squash Family Inhibitor

19. Kuprysiewski, G., Ragnarsson, U., Rolka, K., and Wilusz, T. (1986) Int. J. Pept. Protein Res. 27, 245–250
20. Le-Nguyen, D., Nalis, D., and Castro, B. (1989) Int. J. Pept. Protein Res. 34, 492–497
21. Chen, X. M., Qian, Y. W., Chi, C. W., Gan, K. D., Zhang, M. F., and Chen, C. Q. (1992) J. Biochem. (Tokyo) 112, 43–51
22. Axen, R., and Eruback, S. (1971) Eur. J. Biochem. 18, 351–360
23. Knights, R. J., and Fight, A. (1974) Arch. Biochem. Biophys. 160, 377–386
24. Bendick, A. L., and Belton, P. J. (1976) Plant Physiol. 52, 959–965
25. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
26. Gerard, G. F., and Miller, K. (1986) Focus (BRL) 8, 3
27. Matathias, A. S., and Komro, C. (1989) Focus (BRL) 11, 79
28. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
29. Lutcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F., and Scheele, G. A. (1987) EMBO J. 6, 43–48
30. Proudfoot, N. J., and Brownlee, G. G. (1970) Nature 263, 211–214
31. Heijne, V. (1983) Eur. J. Biochem. 133, 17–21