Pergamon

0031-9422(95)00734-2

*Phytochemistry*, Vol. 41, No. 5, pp. 1265–1274, 1996
Copyright © 1996 Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0031-9422/96 $15.00 + 0.00

# PRIMARY STRUCTURE OF TRYPSIN INHIBITORS FROM *SICYOS AUSTRALIS*

RANDOLPH I. MAR, JOHN A. CARVER,* MARGARET M. SHEIL, JACQUELINE BOSCHENOK, SHANLIN FU and DENIS C. SHAW†

Department of Chemistry, The University of Wollongong, Northfields Ave., Wollongong, N.S.W. 2522, Australia; †Protein Biochemistry Group, Division of Biochemistry and Molecular Biology, John Curtin School of Medical Research, Australian National University, Canberra, A.C.T. 2601, Australia

(*Received in revised form* 13 *September* 1995)

**Key Word Index**—*Sicyos australis*; Cucurbitaceae; protein purification; trypsin inhibitor; amino acid sequence; electrospray ionization mass spectrometry.

**Abstract**—Three trypsin inhibitors from *Sicyos australis*, have been isolated, purified and sequenced. Following protein extraction with ammonium sulphate, the mixture of inhibitors was separated from other proteins by trypsin-affinity chromatography. Subsequent purification of the individual inhibitors was accomplished by reversed-phase HPLC. The primary structures of each inhibitor were elucidated by a combination of protein sequencing and electrospray ionization mass spectrometry (ESI-MS) and tandem mass spectrometry (MS-MS) on both the untreated and the reduced and *S*-carboxymethylated inhibitors. All three inhibitors show extensive sequence similarity with inhibitors from cultivated Cucurbitaceae species, although there are a number of novel residues present. One of the inhibitors has a blocked *N*-terminus (pyroglutamic acid) and the use of MS-MS was crucial to the elucidation of its primary structure. ESI-MS was further used to characterize the non-covalent complex between one of the inhibitors and trypsin.

## INTRODUCTION

Serine proteinase inhibitors in plants can be grouped into six distinct families [1], of which the most recently categorized is that from the Cucurbitaceae [2]. This family includes the well-known commercially cultivated plants such as squash (*Cucurbita pepo*), pumpkin (*Cucurbita maxima*), cucumber (*Cucumis sativus*), watermelon (*Citrullus vulgaras*), bitter gourd (*Momordica charantia*) and choko (*Sechium edule*). In the wild, Cucurbitaceae are found mainly in tropical and subtropical regions and the family is represented in Australia by 18 genera and 44 species [3].

All of the Cucurbitaceae, or squash family, serine proteinase inhibitors studied to date have been purified from the cultivated species [2,4–9]. The inhibitors exist as a number of isoforms of 29–32 amino acids with extensive sequence similarity and invariance in the position of the three disulphide bonds [7, 10]. They are the smallest known protein inhibitors of serine proteinases, are very resistant to heat and acid denaturation, and strongly inhibit trypsin but are inactive towards chymotrypsin. The reactive site peptide bond is between Arg5 or Lys5 and Ile6 [11, 12].

There is considerable pharmaceutical interest in proteinase inhibitors and some, e.g. basic pancreatic trypsin inhibitor or aprotinin, marketed as Trasylol™ [13], are already in clinical use. Inhibitors from Cucurbitaceae species were identified as potential pharmaceuticals even before it became evident that these inhibitors formed a new family of serine proteinase inhibitors. In 1982, Hojima *et al.* [14] purified from pumpkin a potent inhibitor of Hageman factor (factor XIIa), a serine proteinase involved in blood clotting; this inhibitor was later identified as CMTI-III from *Cucurbita maxima* [15]. Subsequently, squash family inhibitors have attracted much attention, particularly because their small size makes them readily amenable to chemical synthesis [16, 17].

In the present study, a variety of chemical and instrumental methods (affinity chromatography, HPLC, protein sequencing and mass spectrometry) have been used to characterize the trypsin inhibitors from an Australian native Cucurbitaceae plant, *Sicyos australis*, an annual climber that grows in areas of disturbed subtropical rainforest. Three trypsin inhibitors have been purified and sequenced. Their amino acid sequences show extensive sequence similarity with the inhibitors from cultivated species.

*Author to whom correspondence should be addressed.

## RESULTS

### *Purification of inhibitors and molecular weight determination*

Initial purification was undertaken via trypsin-affinity chromatography. All tryptic inhibitory activity was contained in the peak eluting from the trypsin-affinity column with 0.01 M HCl, i.e. the application of acid decoupled the inhibitors from trypsin. While bound to trypsin, the majority of the native inhibitors were cleaved at the reactive site peptide bond between residues 5 (arginine or lysine) and 6 (isoleucine), with subsequent hydrolysis [4, 18]. Following cleavage, however, the polypeptide chain remains held together by a disulphide bond between Cys3 and Cys20. The species modified at the reactive site inhibited trypsin.

The HPLC profile of the inhibitor mixture obtained from the affinity column is shown in Fig. 1. Six peaks were collected and assayed for trypsin and chymotrypsin inhibition. The isolates from peaks 1, 3 and 5 did not inhibit either of these enzymes and so were not studied further. The isolates of peaks 2, 4 and 6 inhibited trypsin but not chymotrypsin and were subjected to further extensive characterization by mass spectrometry and protein sequencing.

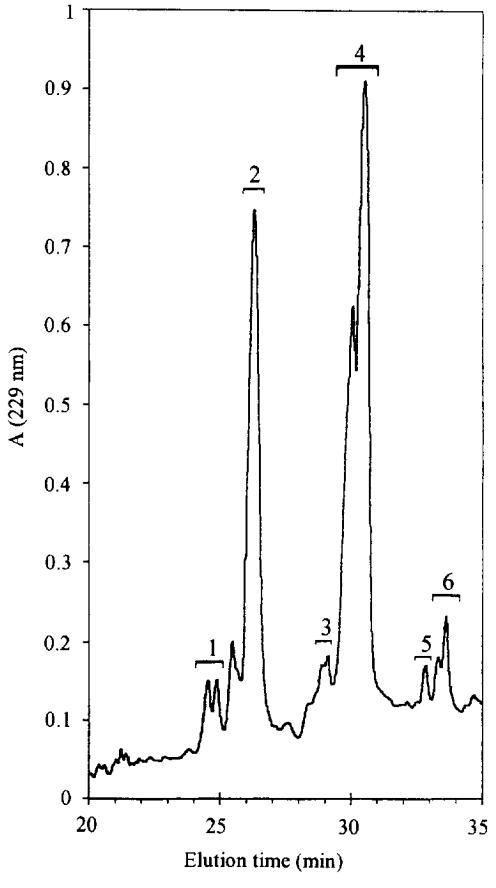


Fig. 1. HPLC profile of the mixture of inhibitors eluted from trypsin-affinity column; see text for elution details

The results of electrospray ionization mass spectrometry (ESI-MS) analyses of the active peaks are summarized in Table 1. In all, six different peptides were identified: a single inhibitor of $M_r$ 3236.8 in peak 2, three inhibitors of $M_r$ 3218.8, 3235.8 and 3503.1 in peak 4 and two inhibitors of $M_r$ 3217.8 and 3485.1 in peak 6. Of these, three peptides are expected to arise when the reactive-site peptide bond is cleaved upon binding to the trypsin-affinity column. Thus, the reactive site modified and intact (native) inhibitor pairs can be readily identified by a difference of 18 ($H_2O$) in their respective $M_r$ values (Table 1). The three native inhibitors were termed SATI (*Sicyos Australis Trypsin Inhibitor*) according to the convention of Wieczorek *et al.* [2] and designated SATI-I, SATI-II and SATI-III, with the suffix indicating the elution order from reversed-phase HPLC.

Although the HPLC separation of peak 4 by the gradient of 21–28% B over 45 min allowed the three inhibitor species to be identified by ESI-mass spectrometry, all attempts to separate these components to the purity required for protein sequencing using a range of different HPLC methods were unsuccessful. The mixture of native SATI-II and SATI-III (peak 6), however, could be further separated into its two components (peaks 6.1 followed by peak 6.2) by HPLC purification using a gradient of 19–27% B over 160 min. These two peptides and reactive-site-modified SATI-I (SATI-I-mod) (peak 2) were isolated and subjected to protein sequencing.

### *Amino acid sequences of SATI-I, SATI-II and SATI-III*

A combination of liquid-phase sequencing and mass spectrometry was used to determine the sequences of SATI-I, II and III, the results of which are summarized in Fig. 2.

(*i*) *Liquid-phase sequencing methods.* SATI-III (peak 6.2) gave no sequence information presumably because its *N*-terminus was blocked (see (*ii*)). The remaining two peptides gave almost complete sequence of information. In each case, however, there were gaps in the amino acid sequence at positions 3, 10, 16, 20, 22 and 29 which corresponded to the disulphide linkages expected from the similarity with other squash family inhibitors (Fig. 3). The first, second and fourth automated sequencing cycles for SATI-I-mod (peak 2) yielded two amino acids due to the cleavage of the reactive site peptide bond producing two *N*-termini during the isolation procedure. Only one

Table 1. $M_r$s of HPLC-purified SATIs determined by ESI-MS

| Inhibitor | | Peak no. (Fig. 1) | $M_r$ |
|---|---|---|---|
| SATI-I | Modified | 2 | 3236.8 |
| SATI-I | Native | 4 | 3218.8 |
| SATI-II | Modified | 4 | 3235.8 |
| SATI-III | Modified | 4 | 3503.1 |
| SATI-II | Native | 6 (6.1) | 3217.8 |
| SATI-III | Native | 6 (6.2) | 3485.1 |

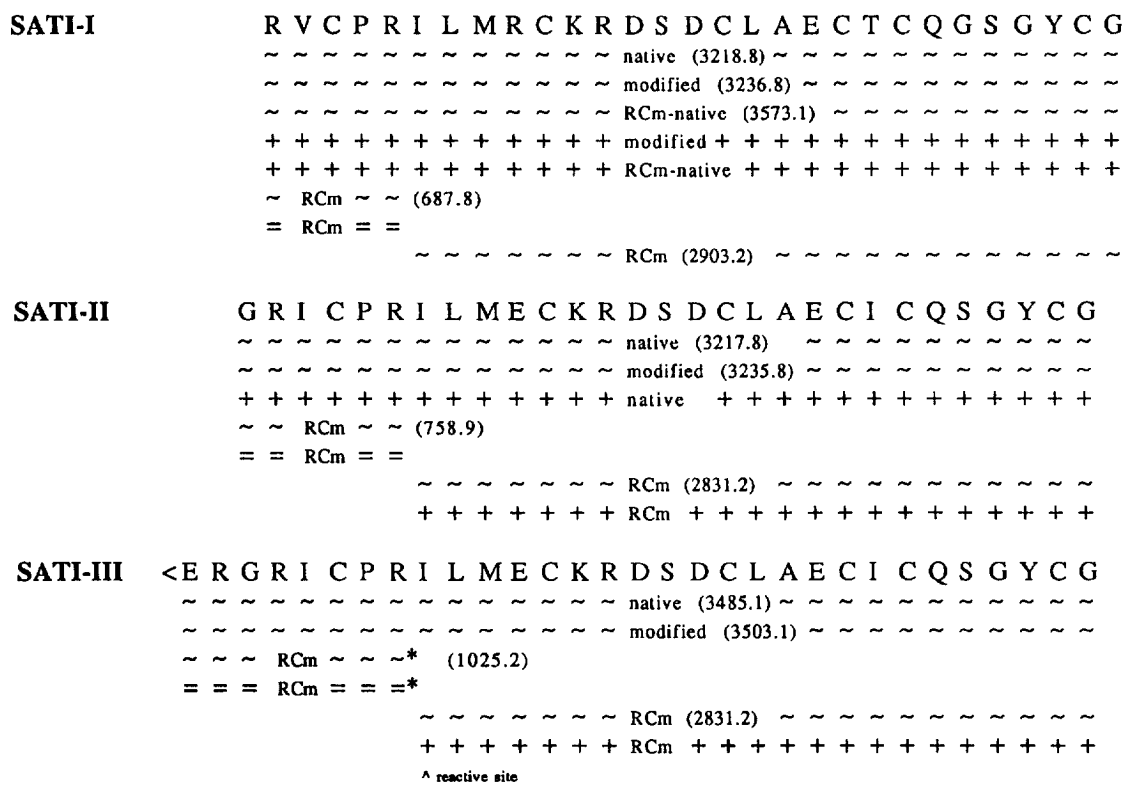


Fig. 2. Amino acid sequences of three trypsin inhibitors from *Sicyos australis* and a summary of methods used to elucidate sequences, ~, ESI-MS (bracketed values are $M_r$); +, protein sequencing; =, MS-MS; *, amino acid analysis; < E, pyroglutamic acid.

residue was detected from the third and fifth cycles, owing to retention of (PTH)Cys through disulphide linkages to the rest of the peptide chain. The sequences were aligned by comparison with other squash family inhibitors (Fig. 3) to maximize positional identity with the positions of the six conserved cysteine residues.

(*ii*) *Mass spectrometric methods.* In order to confirm the position of the cysteine residues, SATI-I-mod (peak 2) and the mixture of peptides in peak 4 were reduced and *S*-carboxymethylated (RCm). The cleavage of the Cys3–Cys20 disulphide bond in the reactive-site-modified inhibitors produced two separate chains which could then be separated by HPLC. The HPLC profiles for the RCm reaction mixtures of peaks 2 and 4 are shown in Fig. 4 and Fig. 5, respectively. Table 2 summarizes the results of ESI-mass spectral analyses of these peaks which allowed identification of each peptide in the respective mixtures. In the reaction mixture of peak 4, however, only four peptide-containing peaks were observed, one less than expected (i.e. one from native SATI-I as it is not cleaved at the reactive site, and two each from SATI-II-mod and SATI-III-mod). This suggested that the sequences of SATI-II and SATI-III were identical from the reactive site to the *C*-terminus which was confirmed by protein sequencing and ESI-mass spectrometry. Furthermore, protein sequencing of native RCm-SATI-I confirmed the sequence previously obtained from unreduced SATI-I-mod.

The smaller RCm-*N*-terminal peptide fragments (RCm-2A from SATI-I-mod, RCm-4A from SATI-II-mod, and RCm-4B from SATI-III-mod) were also analysed by tandem mass spectrometry (MS-MS) [19]. For the first two peptides, these experiments confirmed the sequences of the peptide fragments from the *N*-terminus to the reactive site that were obtained from protein-sequencing of the untreated SATI-I-mod and native SATI-II. As RCm-4B was *N*-terminally blocked, however, the use of MS-MS was crucial to the determination of its sequence. The tandem mass spectra of the doubly charged $[M + 2H]^{2+}$ ions of these very basic peptides generally showed more extensive fragmentation than the corresponding singly-charged $[M + 2H]^{+}$ species. Table 3 lists the major peaks in the tandem mass spectrum of the $[M + 2H]^{2+}$ of RCm-4B from SATI-III-mod. The nomenclature given in Table 3 is a variation of that proposed by Roepstorff and Fohlman [20]. The tandem mass spectrum of RCm-4B was dominated by fragment ions relating to the basic arginine residues; however, there were sufficient sequence ions (i.e. *N*-terminal $B_n$ ions and *C*-terminal $Y_n$ ions) in the spectrum to enable the sequence of this peptide to be deduced in combination with the results of the amino acid analysis. The *N*-terminal $B_1$, $B_2$, $A_1$, $A_2$ and $A_3$ ions indicated that the three *N*-terminal residues were < Glu–Arg–Gly, where < Glu refers to pyroglutamic acid which confirmed that the *N*-terminus was blocked. The *C*-terminal

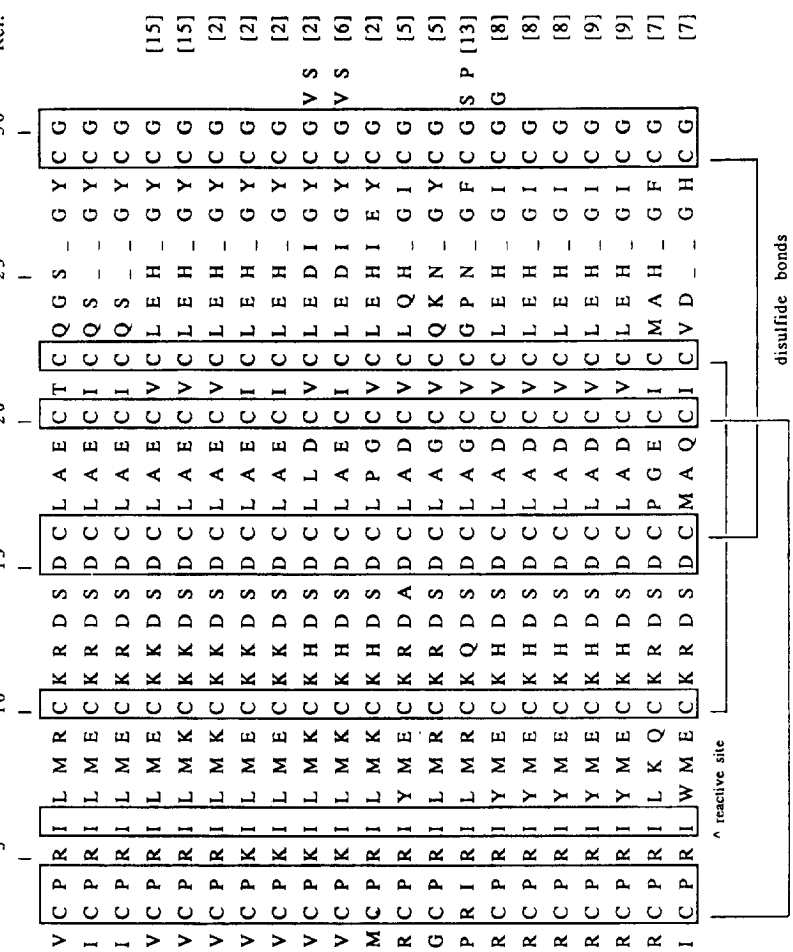


Fig. 3. Amino acid sequence similarity between SATI-I, SATI-II and SATI-III and some other known squash family trypsin inhibitors. < E, pyroglutamic acid.

fragments, $Y''_1$ and $Y''_2$, placed Pro–Arg at the *C*-terminus and the remaining two residues, known to be isoleucine and carboxymethylcysteine (CmC) from the amino-acid analysis, were aligned based on similarity with SATI-II. Thus, the sequence deduced from the combined data was < Glu–Arg–Gly–Arg–Ile–CmC-Pro-Arg which is identical to the *N*-terminal peptide of SATI-II, albeit with an additional two residues, < Glu–Arg, being attached to the *N*-terminus. Therefore, as their *C*-terminal fragments are identical, SATI-III is simply SATI-II with these additional two *N*-terminal residues.

By combining all these data it was possible to determine the complete sequences of all three inhibitors. These are given in Fig. 2, which also summarizes the various experiments used to determine the sequences. The similarities in sequences within the SATIs and between the

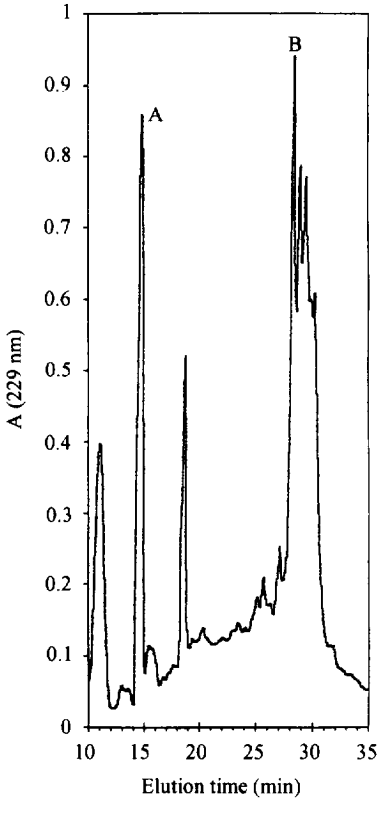


Fig. 4. HPLC profile of reduced and *S*-carboxymethylated peak 2 (from Fig. 1); see text for elution details. Peaks A and B correspond to peptides listed in Table 2.

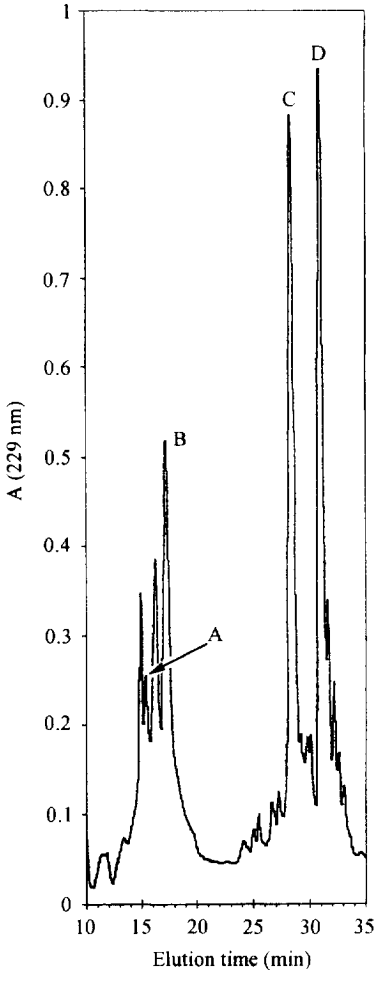


Fig. 5. HPLC profile of reduced and *S*-carboxymethylated peak 4 (from Fig. 1); see text for elution details. Peaks A–D correspond to peptides listed in Table 2.

Table 2. $M_r$s of HPLC-purified reduced and *S*-carboxymethylated (RCm) peptides determined by ESI-MS

| Fraction from HPLC separation | Sequence | SATI parent | Peak label | $M_r^*$ |
|---|---|---|---|---|
| Peak 2 | RVCmCPR | I | A (Fig. 4) | 687.8 |
| (SATI-I-mod) | ILMRCmCKRDSDCmCL-AECmCTCmCQGSYCmCG | I | B (Fig. 4) | 2903.2 |
| Peak 4 | GRICmCPR | II | A (Fig. 5) | 758.9 |
| (SATI-I, SATI-II-mod | < ERGRICmCPR | III | B (Fig. 5) | 1025.2 |
| and SATI-III-mod) | RCm-SATI-I-native | I | C (Fig. 5) | 3573.1 |
| | ILMRCmCKRDSDCmCL-AECmCICmCQSGYCmCG | II and III | D (Fig. 5) | 2831.2 |

*Reduction and *S*-carboxymethylation increases the $M_r$ of the peptide by 58 per cysteine.
CMC = carboxymethylcysteine, < E = pyroglutamic acid.

SATIs and the inhibitors from the cultivated species is evident from Fig. 3.

*ESI-MS of noncovalent complex between trypsin and reactive-site-modified SATI-I*

Recently, a number of groups have demonstrated that ESI-MS can be used to probe non- covalent interactions between a range of proteins and small molecules such as enzyme–inhibitor complexes [21, 22]. Thus, a part of this study, ESI-MS was used to monitor the non-covalent complex of SATI-I-mod with trypsin. Figure 6 shows the ESI-mass spectrum of a 1:1 molar mixture at pH 7 of trypsin and SATI-I-mod. There are two major peaks in the spectrum due to trypsin at $M_r$ 23319.6 (expected 23316.0) and the trypsin–inhibitor complex at $M_r$ 26539.3 (expected 26535.0); there was also a peak (not shown) due to the SATI-I-mod at $M_r$ 3236.0 (expected 3236.8, from Table 1). The intensity of the trypsin–inhibitor complex was pH dependent, with no complex observed below pH 4.

The intensity of the complex was monitored as a function of the ratio of inhibitor added in order to determine the stoichiometry of binding in an analogous fashion to standard methods, e.g. spectrophotometric assaying techniques [23]. Figure 7 shows a plot of the *m/z* 805 peak from the trypsin–inhibitor complex (i.e. the $33^+$ ion) as a function of the molar ratio of inhibitor to trypsin. As is evident from Fig. 7, the amount of complex increases linearly up to a ratio of 1:1 SATI-I-mod:trypsin and at higher ratios the plot levels off, as would be expected for a 1:1 complex.

Table 3. Peak assignments for tandem-mass spectrum of RCm-4B, sequence < Glu–Arg–Gly–Arg–Ile–CmC–Pro–Arg, $M_r$ 1025.2

| Peak assignments* | Observed (*m/z*) | Calculated (*m/z*) |
|---|---|---|
| $[M + 2H]^{2+}$ | 513.8 | 514.1 |
| $[M + 2H - H_2O]^{2+}$ | 505.1 | 505.1 |
| $[M + 2H - CNHNH_2]^{2+}$ | 492.8 | 492.6 |
| $[M + 2H - Arg]^{2+}$ | 435.8 | 436.1 |
| $A_3$ | 296.6 | 297.4 |
| $Y''_2$ | 272.3 | 272.3 |
| $B_2$ | 267.8 | 268.3 |
| $Y''_2$-$NH_3$ | 254.3 | 255.3 |
| $A_2$ | 240.1 | 240.3 |
| $A_3$-Arg side-chain | 196.4 | 197.4 |
| $Y''_1$ | 175.2 | 175.2 |
| $A_2$-Arg side-chain | 139.8 | 140.3 |
| Immonium ion (CmC) | 133.9 | 134.2 |
| $B_1$ | 111.9 | 112.1 |
| $B_1$-$NH_3$ | 94.6 | 95.1 |
| Immonium ion (Ile) | 86.0 | 86.2 |
| $A_1$ | 84.0 | 84.1 |
| Immonium ion (Pro) | 70.0 | 70.1 |
| $CH_3CO_2H$ | 60.0 | |

*Nomenclature according to ref. [20].

## DISCUSSION

Squash family trypsin inhibitors show a high degree of sequence similarity with 10 residues, three disulphide bridges and a reactive site near the *N*-terminus consisting of Arg(Lys)5–Ile6 being totally conserved (Fig. 3). This conservation, with an Arg5–Ile6 reactive site, was evident in the three trypsin inhibitors from *S. australis* (SATI-I, II and III); hence, the SATIs clearly belong to this group of serine proteinase inhibitors.

There are, however, a number of sequence differences between the SATIs and the inhibitors isolated from other Cucurbitaceae. Thus, SATI-I, II and III have glutamine at position 23 which is shared only with BDTI-II from *Bryonia dioica*. In addition, SATI-I and BDTI-II share an arginine residue at position 9 which is occupied by the more common glutamic acid residue in SATI-II and SATI-III (Fig. 3). Typically, inhibitors from species within the same genus share unique residues [2]. Hence,

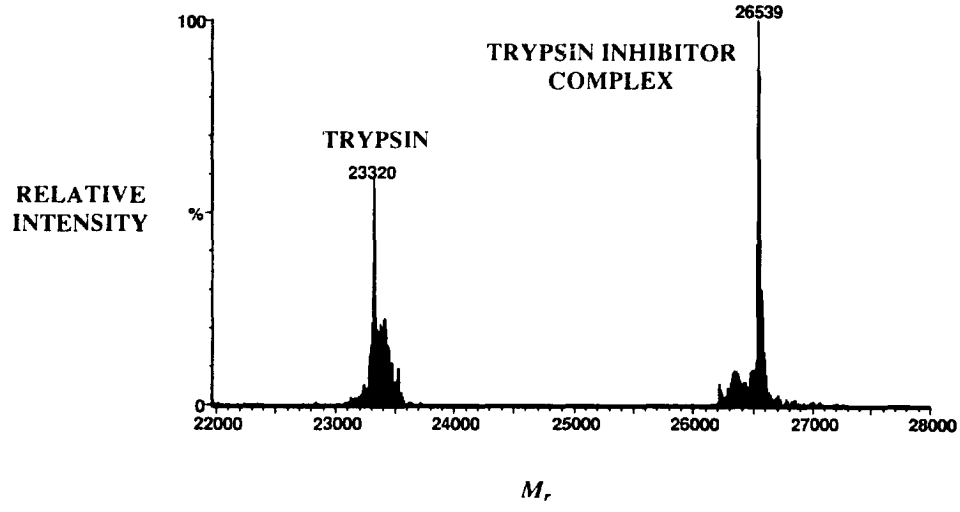


Fig. 6. Transformed ESI-mass spectrum of trypsin and SATI-I-mod mixed in a 1:1 molar ratio at pH 7.

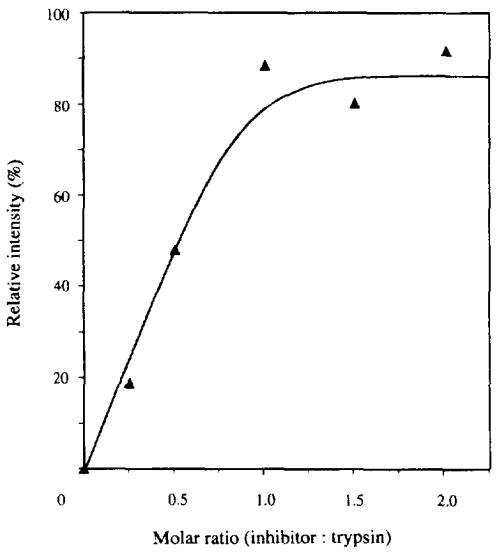


Fig. 7. Relative intensity of $m/z$ 805 ($33^+$) ion from the trypsin-SATI-I-mod complex observed in the ESI mass spectra of mixtures in which the molar ratio of SATI-I-mod to trypsin was varied.

the commonality of Gln23 between SATI-I, -II and -III and BDTI-II and Arg9, between SATI-I and BDTI-II is unusual because they are isolated from species of different genera. SATI-I has further unique residues in Thr21, Gly24 and Ser25. These features distinguish SATI-I not only from SATI-II and SATI-III, but from all squash family trypsin inhibitors that have been identified. Nonpolar residues (valine and isoleucine) occupy position 21 in the sequences of most other known squash family inhibitors, although the basic residue lysine is present in TKTI-III, TKTI-IV [24] and HMTI [25], and arginine in TGTI-II [25] (not included in Fig. 3). Hence, an uncharged polar threonine residue at position 21 of SATI-I is novel. SATI-II and SATI-III have the common isoleucine at residue 21. Similarly, Gly24 in SATI-I is interesting as most other squash family inhibitors have polar residues at this position, including SATI-II and SATI-III, which have a serine residue at this position. There is no residue at position 25 in SATI-II and III, a deletion that is shared only with MCTI-II (Fig. 3).

SATI-II and SATI-III are identical except for < Glu–Arg added to the *N*-terminus of SATI-III. < Glu at the *N*-terminus of SATI-III is present in only three other squash family inhibitors, MCTI-I, LLTI-I and LLDTI-II, all of which have two less residues in their *N*-terminal extensions (Fig. 3). SATI-II has a single residue (glycine) as its *N*-terminal extension, while SATI-III has a three-residue (< Glu–Arg–Gly) extension. It follows that SATI-II may be a product of tryptic cleavage at Arg2 in SATI-III prior to or during purification. A similar structural motif is observed in the inhibitor pairs of CPTI-II/III and CMTI-III/IV, i.e. the inhibitors of each pair have identical sequences except for an *N*-terminal extension of His–Glu–Glu in CMTI-IV and CPTI-III (see Fig. 3). Thus, CMTI-III and CPTI-II may also arise from limited proteolysis of CMTI-IV and CPTI-III, respectively [2]. It is conceivable that these extended termini and shortened inhibitor pairs could be expressed by individual genes, but it is more likely that the larger chain is expressed exclusively, and the shorter chain is produced by proteolysis of the synthesized parent [2]. The tertiary structural differences between the various SATIs and other squash family inhibitors is currently under investigation by detailed NMR spectroscopic studies. Such data may provide information on the structural role of these terminal regions.

During purification by trypsin-affinity chromatography, the three SATIs were cleaved at the reactive site bond, producing a further three corresponding species. The modification of the reactive site of squash family inhibitors is a feature of trypsin-affinity chromatography [4, 13] and is a general observation when naturally occurring trypsin inhibitors interact with trypsin [18, 26]. The resultant modified inhibitors retain antitryptic activity [18, 26], as was observed in this study. Squash family inhibitors have been isolated exclusively in the native form by using alternatives to trypsin-affinity chromatography, such as anhydrotrypsin-affinity chromatography [14, 24] or ion-exchange chromatography [8, 12]. Anhydrotrypsin is a catalytically inactive trypsin derivative with the active-site Ser195 modified to dehydroalanine [27]. The modified trypsin forms complexes with inhibitors without cleaving the reactive site bond of the inhibitor and can be easily immobilized on insoluble supports.

The SATI primary structures were determined by comparing sequence information obtained by standard protein-sequencing, MS-MS and amino acid analysis with the $M_r$ obtained by ESI-MS, and through alignment of the conserved residues to maximize similarity with other squash family inhibitors. The reduction and *S*-carboxymethylation of the inhibitors which had been hydrolysed at the reactive-site Arg5–Ile6 bond by trypsin produced two peptide chains. MS-MS provided useful confirmation of the sequences of the resultant smaller *N*-terminal RCm-peptide fragments. For these arginine-rich peptides, the tandem mass spectral data alone were incomplete, as the spectra were dominated by fragment ions from the arginine residues rather than ions indicative of the entire sequence [28]. However, in the case of SATI-III, for which amino acid analysis data were also available, the MS-MS experiments allowed determination of the sequence of the blocked *N*-terminal peptide. Protein sequencing was not required, thus eliminating the need for prior treatment with pyroglutamate aminopeptidase to remove the blocked *N*-terminal residue.

Owing to retention of antitryptic activity in the reactive-site-modified species, SATI-I-mod could be effectively used to determine the binding ratio in the noncovalent complex formed with trypsin. The novel technique of ESI-MS was used to monitor this interaction (Figs 6 and 7). The gentle nature of the electrospray ionization process enabled the ready observation of the noncovalent complex between trypsin and the inhibitor.

As expected, SATI-I-mod formed a complex with bovine trypsin in a 1:1 molar ratio. It should also be possible to use ESI-MS to determine the binding constants for the trypsin–inhibitor complex; however, this requires a more detailed study in which the relative abundances of the trypsin, inhibitor and complex are accurately known. It is expected that the association equilibrium constants ($K_a$) for the SATIs with trypsin will be within the range of $10^9$–$10^{12}$ $M^{-1}$, which is typical of the squash family inhibitors [2, 7–9, 13].

## EXPERIMENTAL

*Materials.* All reagents and solvents were of analytical reagent grade or better.

*Purification of inhibitors.* Succulent fruit (3.286 kg 71% w/w $H_2O$) from *S. australis* was collected from local, disturbed subtropical rainforest areas at Dunmore, N.S.W., Australia, dried at 40° to constant weight (535 g) and then pulverized and defatted ×4 with 1.5 l of $Me_2CO$ at 4°. At each extraction, the mixt. was clarified by centrifugation. The green $Me_2CO$ supernatants were combined (6 l) and allowed to evap. in a fumehood. The $Me_2CO$-sol. content of the pulverized dried fruit was 17% (w/w). The defatted meal was then dried at 40° to constant wt (396 g) and homogenized ×3 in 800 ml of 0.1 M Na acetate buffer, pH 4. Again, each extract was clarified by centrifugation. The supernatants were combined (2.4 l) and then treated with $(NH_4)_2SO_4$ (90% satn) to ppt. all protein present. The pellet was collected by centrifugation and, after lyophilization, the pellet (41.1 g) was stored at −20°.

*Assay of protease inhibition.* Inhibition of bovine trypsin (type III) (Sigma) was assayed using *N*-benzoyl-L-arginine ethyl ester (BAEE) (Sigma) as substrate [29]. Trypsin was dissolved in 0.001 M HCl containing 0.01 M $CaCl_2$ at 1 mg $ml^{-1}$. This enzyme soln (30 μl) was incubated with the inhibitor in 600 μl 0.05 M Tris, pH 8.1, for 10 min at 25°. A 200 μl aliquot of this mixt. was pipetted into a cuvette containing 1 ml of 0.002 M BAEE and the change in *A* at 265 nm, 25°, was recorded. The rate of hydrolysis of BAEE was referenced against a soln containing 200 μl of 0.05 M Tris, pH 8.1, and 1 ml of the substrate soln. Typically, an increase in *A* of 0.2 AU was observed over 2.5 min.

Purified inhibitors of trypsin were assayed for anti-chymotrypsin activity using *N*-benzoyl-L-tyrosine ethyl ester (BTEE) (Sigma) as substrate [30]. Chymotrypsin (Sigma) was dissolved in 0.001 M HCl at 1 mg $ml^{-1}$. BTEE was dissolved in 50% aq. MeOH to a concn of 0.001 M. The enzyme soln (25 μl) was incubated with the inhibitor in 600 μl of 0.05 M Tris, pH 8.1, for 10 min at 25°. A 200 μl aliquot of this mixt. was pipetted into a cuvette containing 480 μl of the substrate soln and 500 μl of 0.05 M Tris, pH 8.1, and the change in *A* at 256 nm, 25°, was recorded. The rate of hydrolysis of BTEE was referenced against a soln containing 480 μl of substrate soln and 700 μl of 0.05 M Tris, pH 8.1. Typically, an increase in *A* of 0.2 AU was observed over 2.5 min.

*Immobilization of trypsin on cyanogen bromide activated Sepharose 4B.* Sepharose 4B (Pharmacia) was activated with cyanogen bromide by a slight modification of the method of ref. [31]. Settled wet gel (50 ml) was activated with 11.5 g cyanogen bromide in 11.5 ml MeCN. Bovine trypsin (type III) (Sigma) (400 mg, 4 mg $ml^{-1}$ coupling buffer) was added to the activated gel and the coupling reaction allowed to proceed at 4° for 16 hr with gentle rocking. An estimate of unbound trypsin (*ca* 150 mg) in the coupling soln was made by assaying the trypsin soln before and after coupling to the gel. Unbound trypsin was washed from the gel before the remaining activated groups on the agarose were blocked by treatment with 1 M 2-ethanolamine, pH 8, for 2 hr at 25° [32]. After thorough washing, the gel was packed into a Pharmacia C column (2.6 × 15 cm) and stored at 4° in 0.05 M Tris, pH 7, containing 0.01 M $CaCl_2$ and 0.02% (w/v) $NaN_2$.

*Crude purification of inhibitors by trypsin-affinity chromatography.* The trypsin-Sepharose 4B affinity column was equilibrated with 0.05 M Tris, pH 8.1, prior to loading the sample of lyophilized $(NH_4)_2SO_4$-precipitated protein (2 g) which had been dissolved in a minimal amount of equilibration buffer. Unbound protein was washed from the column with $H_2O$ and then 0.2 M NaCl. Inhibitors were decoupled from the immobilised trypsin with 0.01 M HCl. A flow rate of 2 ml $min^{-1}$ was used and elution was monitored at 280 nm. All peaks eluted from the column were immediately assayed for antitryptic activity and those containing inhibitor were freeze-dried. Inhibitors eluted by acid were neutralized with 0.2 M NaOH prior to lyophilization. The column was stored in the buffer described above.

*Purification, reduction and S-carboxymethylation of inhibitors.* Inhibitors were purified by reversed-phase HPLC on a C18 Alltech Macrosphere 300 5-μm analytical column (0.46 × 25 cm) with a solvent system of 0.1% (v/v) TFA in $H_2O$ (solvent A) and 0.1% (v/v) TFA in MeCN (Solvent B). A flow rate of 1 ml $min^{-1}$ was used and the eluted protein was monitored by measuring *A* at 229 nm. Individual inhibitors were isolated from the crude affinity-purified inhibitor mixture using a linear gradient of 17–32% B over 18 min. Larger quantities of inhibitor were sepd on a C18 Alltech Macrosphere 300 5-μm semi-prep. column (1 × 25 cm) using these conditions but with a flow rate of 5 ml $min^{-1}$. Inhibitors not sepd by the elution gradient above were isolated using slower linear gradients, such as 21–28% B over 45 min and 19–27% B over 160 min, on the analytical HPLC column.

HPLC-purified inhibitors were reduced with dithiothreitol and alkylated with iodoacetic acid according to the method of ref. [33]. A protein concn of 2 mg $ml^{-1}$ was used in 350 μl of reaction buffer containing 6 M guanidinium chloride, 0.1 M Tris and 0.001 M EDTA, pH 8.3. Immediately after quenching the reaction with mercaptoethanol, the reduced and *S*-carboxymethylated (RCm) peptides were sepd by HPLC using a linear gradient of 0–60% B over 60 min with the system described above, fitted with the analytical column used previously.

*Mass spectrometry.* The $M_r$ of each HPLC-purified inhibitor and peptides from the RCm-inhibitors was determined by ESI-MS using a triple quadrupole instrument with the solvent delivered at 5 $\mu$l min$^{-1}$. Samples (*ca* 20 pmol $\mu$l$^{-1}$ for each inhibitor) were dissolved in 50% aq. MeCN containing 1% $HCO_2H$; 10 $\mu$l was injected for each analysis. The skimmer potential was set at 50 V and the resolution to 1 mu. The mass spectrometer was scanned at a rate of 1 sec per 100 mu and data from 10–20 scans were summed to obtain representative spectra. Sequences of the smaller RCm-*N*-terminal peptide fragments were determined by MS-MS. For these expts, Ar was used as collision gas and the collision energy was either 100 or 200 eV, as indicated below. The concns of peptide samples were higher (*ca* 100 pmol $\mu$l$^{-1}$) than those used for ESI-MS expts.

The binding of one of the inhibitors to trypsin was also investigated by ESI-MS. For these expts, a sample of sequence-grade bovine trypsin (Sigma) was dissolved in 0.01% TFA at 4 pmol $\mu$l$^{-1}$. This soln (20 $\mu$l) was then combined at pH 4 with varying amounts of inhibitor solution (in ratios ranging from 0.25:1 to 2:1, inhibitor: trypsin) and the resulting solns dried in a vacuum centrifuge. Each dried sample was then redissolved in 20 $\mu$l of 50% aq. MeCN (pH of $\approx$7) prior to analysis. The pH of the final soln was critical to the success of the analysis.

*Protein sequencing.* Amino acid sequencing of inhibitors was performed on a pulsed liquid-phase protein sequencer with an on-line analyser. Samples were added to a polybrene-treated glass fibre disc and sequenced using standard procedures. To remove salt and contaminants, some samples were adsorbed onto ProBlott™ and the membrane washed with $H_2O$ prior to sequencing in the Blott™ cartridge on the instrument.

*Amino acid analysis.* Amino acid analyses were performed using the AminoQuant™ procedure which prederivatizes the amino acids with *o*-phthalaldehyde and proline with 9-fluorenylmethyl chloroformate, then separates each derivative by reversed-phase HPLC. After the protein had been hydrolysed to amino acids and dried, an aliquot of the hydrolysate in buffer was derivatized on the instrument prior to injecting onto the column.

*Acknowledgements*—Financial support for this work was provided by the Australian Research Council. R.M. was in receipt of an Australian Postgraduate Research Award. The assistance of Assoc. Prof. Roger Truscott and Mr Andrew Aquilina with some aspects of protein purification is gratefully acknowledged.

## REFERENCES

1. Laskowski, M. J. (1986) in *Nutritional and Toxicological Significance of Enzyme Inhibitors in Foods.* (Friedman, M., ed.), p. 1. Plenum Publ. Corp., New York.
2. Wieczorek, M., Otlewski, J., Cook, J., Parks, K., Leluk, J., Wilimowska-Pelc, A., Polanowski, A., Wilusz, T. and Laskowski, M. (1985) *Biochem. Biophys. Res. Commun.* **126**, 646.
3. Telford, I. R. (1982) in *Flora of Australia*, Vol. 8, p. 158. Australian Govt. Publ. Service, Australia.
4. Polanowski, A., Wilusz, T., Nienartowicz, B., Cieslar, E., Slominska, A. and Nowak, K. (1980) *Acta Biochim. Polon.* **27**, 371.
5. Otlewski, J., Whatley, H., Polanowski, A. and Wilusz, T. (1987) *Biol. Chem. Hoppe-Seyler* **368**, 1505.
6. Polanowski, A., Otlewski, J., Leluk, J., Wilimowska-Pelc, A. and Wilusz, T. (1988) *Biol. Zent. Bl.* **107**, 45.
7. Hara, S., Makino, J. and Ikenaka, T. (1989) *J. Biochem.* **105**, 88.
8. Hamato, N., Takano, R., Kamei-Hayashi, K. and Hara, S. (1992) *Biosci. Biotech. Biochem.* **56**, 275.
9. Matsuo, M., Hamata, N., Takano, R., Kamei-Hayashi, K., Yasuda-Kamatani, Y., Nomoto, K. and Hara, S. (1992) *Biochim. Biophys. Acta* **1120**, 187.
10. Bode, W., Greyling, H. J., Huber, R., Otlewski, J. and Wilusz, T. (1989) *FEBS Letters* **242**, 285.
11. Leluk, J., Otlewski, J., Wieczorek, M., Polanowski, A. and Wilusz, T. (1983) *Acta Biochim. Polon.* **30**, 127.
12. Otlewski, J., Polanowski, A., Leluk, J. and Wilusz, T. (1984) *Acta Biochim. Polon.* **31**, 267.
13. Favel, A., Mattras, H., Coletti-Previero, M. A., Zwilling, R., Robinson, E. A. and Castro, B. (1989) *Int. J. Pep. Prot. Res.* **33**, 202.
14. Hojima Y., Pierce, J. V. and Pisano, J. J. (1982) *Biochemistry* **21**, 3741.
15. Wilusz, T., Wieczorek, M., Polanowski, A., Denton, A., Cook, J. and Laskowski, M. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 93.
16. Kupryszewski, G., Ragnarsson, U., Rolka, K. and Wilusz, T. (1986) *Int. J. Pep. Prot. Res.* **27**, 245.
17. Le-Nguyen, D., Nalis, D. and Castro, B. (1989) *Int. J. Pep. Prot. Res.* **34**, 492.
18. Laskowski, M. and Kato, I. (1980) *Annu. Rev. Biochem.* **49**, 593.
19. Busch, K. L., Glish, G. L. and McLuckey, S. M. (1988) in *MS/MS: Techniques and Applications of Tandem Mass Spectrometry*, p. 1. VCH, New York.
20. Roepstorff, P. and Fohlman, J. (1984) *Biomed. Mass Spectrom.* **11**, 601.
21. Ganem, B., Li, Y. T. and Henion, J. D. (1991) *J. Am. Chem. Soc.* **113**, 7818.
22. Gale, D. C., Goodlett, D. R., Light-Wahl, K. J. and Smith, R. D. (1994) *J. Am. Chem. Soc.* **116**, 6027.
23. Green, N. M. and Work, E. (1953) *Biochem. J.* **54**, 347.
24. Qian, Y., Tan, F., Qi, Z. and Chi, C. (1990) *Sci. China Ser. B Chem. Life Sci. Earth Sci.* **33**, 599.
25. Ling, M. H., Qi, H. and Chi, C. (1993) *J. Biol. Chem.* **268**, 810.
26. Ozawa, K. and Laskowski, M. (1966) *J. Biol. Chem.* **241**, 3955.
27. Ako, H., Foster, R. J. and Ryan, C. A. (1972) *Biochem. Biophys. Res. Commun.* **47**, 1402.

28. Kilby, G. W. and Sheil, M. M. (1993) *Org. Mass Spectrom.* **28**, 1417.
29. Schwert, G. W. and Takenaka, Y. (1955) *Biochim. Biophys. Acta* **16**, 570.
30. Hummel, B. C. W. (1959) *Can. J. Biochem.* **37**, 1393.
31. March, S. C., Parikh, I. and Cuatrecasas, P. (1974) *Anal. Biochem.* **60**, 149.
32. Hill, C. R., Thompson, L. G. and Kenney, A. C. (1989) in *Protein Purification Methods: a Practical Approach* (Harris, E. L. V. and Ange, S., eds), p. 286. IRL Press, Oxford.
33. Allen, G. (1981) in *Sequencing of Proteins and Peptides: Laboratory Techniques in Biochemistry and Molecular Biology* (Work, T. S. and Burdon, R. H., eds), 1st edn, p. 30. Elsevier Science Publishers B.V., Amsterdam.