Volume 297, number 1,2, 143–146 FEBS 10645 February 1992
© 1992 Federation of European Biochemical Societies 00145793/92/$5.00

Amino acid sequencing of a trypsin inhibitor by refined 1.6 Å X-ray crystal structure of its complex with porcine β-trypsin

Qichen Huang[a], Shengping Liu[a], Youqi Tang[a], Fuyue Zeng[b] and Ruiqing Qian[b]

[a]*Institute of Physical Chemistry, Peking University, 100871 Beijing, PR China* and [b]*Shanghai Institute of Organic Chemistry, Academia Sinica, 200032 Shanghai, PR China*

Received 24 September 1991, revised version received 5 December 1991

The stoichiometric complex formed between porcine β-trypsin and the *Momordica charantia*, Linn. *Cucurbitaceae* trypsin inhibitor-A (MCTI-A) was crystallized and its X-ray crystal structure determined using molecular replacement method. The primary sequence and topology of the inhibitor was determined by recognizing the electron density and refined to a final *R* value of 0.167 (7.0–1.6 Å) with RMS deviation of bond lengths from standard values 0.012 Å. The sequence was compared with those obtained by other groups and was found to be similar to the squash proteinase inhibitor. Its spatial structure and the conformation of its primary binding segment from Cys-3I (P3) to Glu-7I (P3') which contains the reactive scissile bond Arg-5I C-Ile-6I N were also very similar with other squash family proteinase inhibitors.

Momordica charantia, Linn. trypsin inhibitor-A; Trypsin; Complex; Crystallography sequencing; Active site geometry

1. INTRODUCTION

There are many plant proteinase inhibitors which play important roles in the plant's life [1–4]. X-Ray crystallography provides a unique method to obtain a detailed description of the spatial structure and interactions between proteinase and inhibitor.

Three trypsin inhibitors (MCTI-A, -B, -C) were purified from Bitter Gourd seeds (*Momordica charantia*, Linn. *Cucurbitaceae*). Among them MCTI-A is very resistant to heat or acid denaturation and exhibits strong inhibition towards trypsin ($K_i = 1.7 \times 10^{-10}$ M). The amino acid compostition, primary sequence and the disulfide topology were unknown at the start of this study. We have determined its primary and spatial structure by X-ray crystallography. Comparison of our results with those of other groups shows that MCTI-A is a member of squash family inhibitors [4–9]. We report the structure determination of the complex of this inhibitor with porcine β-trypsin [10,11]. The final coordinates will be deposited at the Protein Data Bank, Brookhaven National Laboratory.

2. MATERIALS AND METHODS

MCTI-A inhibitor was isolated and purified from the seeds of Bitter Gourd (Guangzhou Seeds Co., China) (Qian, R., personal communication). Porcine β-trypsin was provided by Shanghai Institute of Biochemistry, Academia Sinica. MCTI-A and porcine β-trypsin were dissolved in 0.01 M HAc-NaAc buffer (pH 4.3) at a ratio of 1:5.4 (w/w), total protein concentration was 2% (w/v). Crystals were obtained using 'hanging drop' vapour diffusion technique. Precipitant solution was 0.2 M sodium phosphate buffer containing 35% saturation NaCl, pH 6.0–7.0. The reservoir solution was the same as the precipitant one, but NaCl was 85% saturation. Trigonal crystals of space group $P3_221$ were obtained at 24°C within about one week; the cell constants are $a=b=62.65$ Å, $c=124.31$ Å, $\alpha=\beta=90°$, $\gamma=120°$. The crystal contains one complex molecule per asymmetric unit and diffracts to almost 1.4 Å.

X-Ray intensity data were collected with several different detectors. They were further processed and merged by means of the PROTEIN program package [12] (see Table I).

The rotation of the trypsin component was determined by fast rotation function program FROT [13] using the refined bovine β-trypsin model [14]. The positioning of the trypsin molecule in the unit cell [15] was achieved by a translation function search method with programs written by E.E. Lattmann [16] (modified by J. Deisenhofer and R. Huber).

Different residues of porcine β-trypsin from bovine β-trypsin [10] and the inhibitor model were built into electron density using program TOM/FRODO [17,18]. The complex model was refined with the energy restraint crystallographic refinement procedures in program XPLOR [19,20]. After the crystallographic *R* value had dropped to 0.26, group and restraint B values were refined on 7.0–1.6 Å resolution data. Solvent molecules were inserted at stereochemically reasonable positions where the difference electron density exceeded 3σ. Finally, the individual atomic *B* values were also refined. The final *R* value (defined as $R = \Sigma||F_o|-|F_c||/\Sigma|F_o|$) for 31,697 unique reflections within 7.0–1.6 Å resolution is 0.167. The final refinement characteristics are given in Table II.

3. RESULTS

The primary sequence of MCTI-A was determined from the electron density maps. It has 28 amino acid residues and from its topology it is clear that MCTI-A

Correspondence address: Q. Huang, Institute of Physical Chemistry, Peking University, 100871 Beijing, PR China.

Abbreviations: MCTI, *Momordica charantia*, Linn. trypsin inhibitor; RMS, root-mean-square; B, isotropic temperature factor; inhibitor residues are indicated by an I after the residue number.

Table I

Reflection data

Set	Resolution Å	Equipment	Reflections measured	Reflections observed ($f>2\sigma$) (unique)	Observed/ Possible (%)	R_{merg} (%)
1	2.70	Rikagu AFC-5R diffractometer	9,346	7084	86	7.5
2	1.78	Huber rotation camera	32,439	19405	70	6.9
3	1.60	UCSD Mark-II area detector	223,007	32214	85	6.8
Total	1.60			32,238	85	8.3

$R_{merg} = \Sigma|I_i - \langle I \rangle|/\Sigma I_i$

is a member of squash proteinase inhibitor family [4–9] (see Fig. 1). There are 5 amino residues whose assignments are uncertain (R1, S2, T11, K19 and A24) because of weak and irregular side chain electron density. We place these residues according to the lengths of the side chain electron density at 1σ contour level. It is impossible to distinguish between Asp and Asn, Glu and Gln, Val and Thr, because they have the same side chain shape. We define these amino residues according to the homologies with other known squash inhibitors [4–9].

After finishing our work we noticed the results obtained by Saburo Hara et al. on three proteinase inhibitors from Bitter Gourd [5]. Compared with their results we found that our MCTI-A is very similar to their MCTI-II. Three out of four different residues of our results are of shorter side chains than theirs and all differences occur where there is some doubt in assignment of the side chain electron density. These differences may be caused by different sources of proteins, or by errors in analysis. Chemical analysis of our sample is in progress.

Porcine β-trypsin differs from bovine β-trypsin in 39 residues. These residues can be fitted to electron density with the exception of the Arg-125, which is disordered. Residue 27 of porcine trypsin appears to be Ile and not Val [10].

In the Fourier map the inhibitor's primary binding segment 3I(P3')–8I(P3) is clear and the atoms have low B values. Whole inhibitor shows continuous electron density although it is weak at Arg-1I position, the N-terminus. There are three strong fixed solvent molecules of low B values, namely 24W, 65W, 72W, and these can be considered as integral constituents of the inhibitor structure. The trypsin chain can be completely traced from its C-terminus to its N-terminus, but with several side chain atoms being undefined (all involved in crystal contacts). The calcium ion is at almost the same position as found by others [22].

Our inhibitor model is similar in structure to the inhibitor from pumpkin seeds *Cucurbita maxima* (CMTI-I) determined by W. Bode et al. [23]. Fig. 2 shows the Cα drawing of the complex, with the projection direction close to that shown in [23]. The inhibitor roughly exhibits the shape of a flat-iron, with the base plate made up of the polypeptide chain from Met-8I on, and the grip formed by the protruding reactive site loop containing the scissile peptide bond Arg-5I C–Ile-6I N.

Table II

Final model parameters of the MCTI-A-porcine β-trypsin structure

Number of active protein atoms	1,917
Number of active solvent atoms	136
RMS standard deviation from target values	
Bond lengths	0.012 Å
Bond angles	2.6°
Torsion angles	28.5°
Improper	1.1 Å
Resolution range	7.0–1.6 Å
Number of unique reflections used for refinement	31,697
R value	0.167*
Estimated mean coordinate error from R value according to Luzzati [21]	0.18 Å

*Without omitting reflection data.

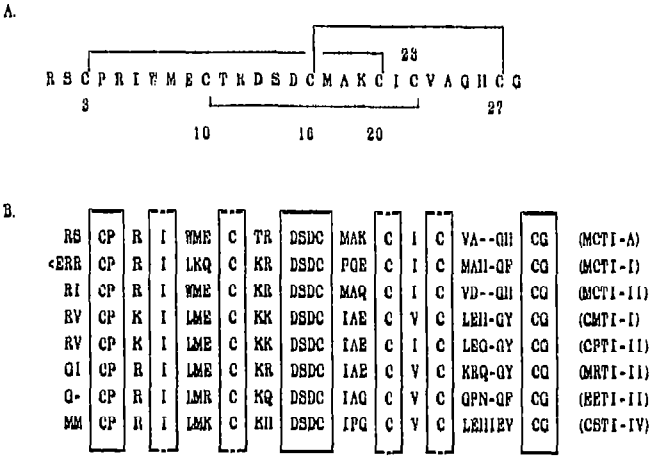

Fig. 1. Primary sequence and disulfide topology of MCTI-A. (A) Primary sequence and disulfide topology of MCTI-A. (B) Primary sequence comparison between MCTI-A and other trypsin inhibitors of squash family. MCTI-A = *Momordica charantica, Linn.* inhibitor A; MCTI-I = *Momordica charantica Linn.* inhibitor I; MCTI-II = *Momordica charantica, Linn.* inhibitor II; CMTI-I = *Cucurbita maxima* trypsin inhibitor II; CPTI-II = *Cucurbita pepo* trypsin inhibitor II; MRTI-I = *Momordica repens* trypsin inhibitor I; EETI-I = *Momordica elaterium* trypsin inhibitor II; CSTI-IV = *Cucumis sativa* trypsin inhibitor IV.

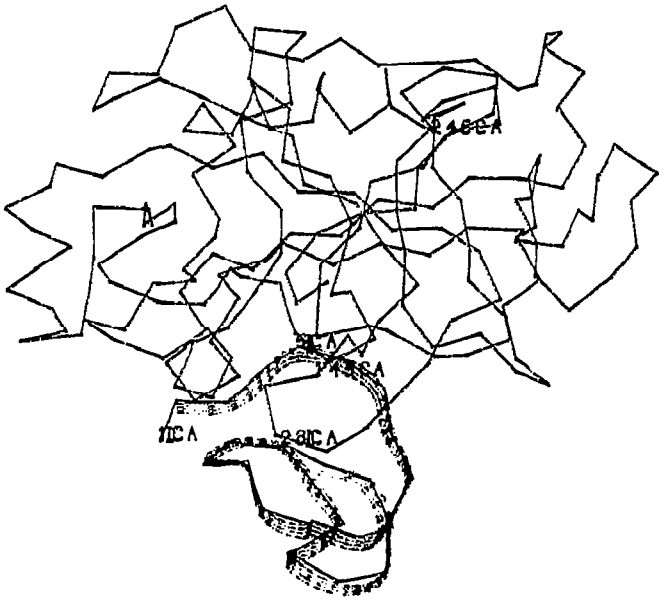

Fig. 2. Cα drawing of the complex formed between porcine β-trypsin and MCTI-A (highlighted by ribbon diagram).

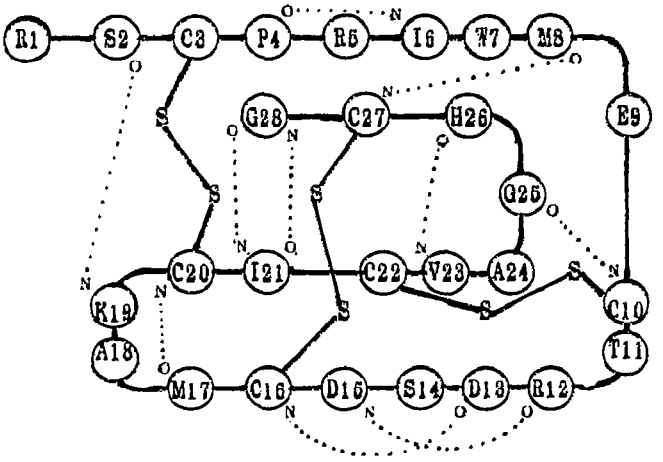

Fig. 3. Schematic representation of the polypeptide arrangement of MCTI-A and its disulfide connectivities. Inter-main chain hydrogen bonds displayed by dashed lines were selected according to XPLOR [20].

The inhibitor has no regular secondary structure, except a type-II tight turn at 17I–20I and a type I′ tight turn at 23I–26I. One residue (Ala-18I) at the type-II turn shows the conformation outside the allowed region of $\varphi - \psi$ [24], with φ=39.7°, ψ=−126.9°.

Fig. 3 shows the hydrogen bond scheme of the inhibitor. The hydrogen bond between Arg-1I O and Cys-3I N observed in [23] was not observed in our results. Because of the high temperature factors of the side chain atoms of Arg-1I, the hydrogen bonds between them and the core of the inhibitor reported by W. Bode are not confirmed by our result. There are 5 pairs of intermolecular main chain–main chain hydrogen bonds between the inhibitor and trypsin. As observed in other trypsin– inhibitor complexes [22,23,25], these hydrogen bonds are between Cys-3I (P3) and Gly-216, between Arg-5I (P1) and Ser-214, Gly-193, Ser-195, between Trp-7I and Phe-41. The complete structure of the inhibitor is illustrated in Fig. 4.

The side chain of P1-residue Arg-5I is accommodated in the S1-pocket of trypsin with almost identical geometry to that reported in [22,23,25]. The distal guanidyl group is engaged in pairs of hydrogen bonds with Ser-190 OG, Asp-189 OD_1 and OD_2, and its NE atom in hydrogen bonds with solvent 58W OW, which is further hydrogen bonded to Gln-192 OE_1 and Cys-3I O. The conformation of the contacting area of the inhibitor and trypsin is the same as that observed by others [22,23,25], especially the P3′–P3 segment of the inhibitor and the S3′–S3 of the trypsin. The RMS deviation of equivalent atoms of MCTI-A and Mung bean inhibitor lysine fragment [25] from 3I to 7I is 0.31 Å, but from residue 8I onwards the 2 folds are different. The scissile peptide bond Arg-5I C-Ile-6I N is normal, and no distortion about the planarity is observed. The distance between Ser-195 OG and Arg-5I is 2.57 Å [22,23].

4. DISCUSSION

This structure analysis shows that it is possible to sequence an unknown protein by X-ray crystallog-

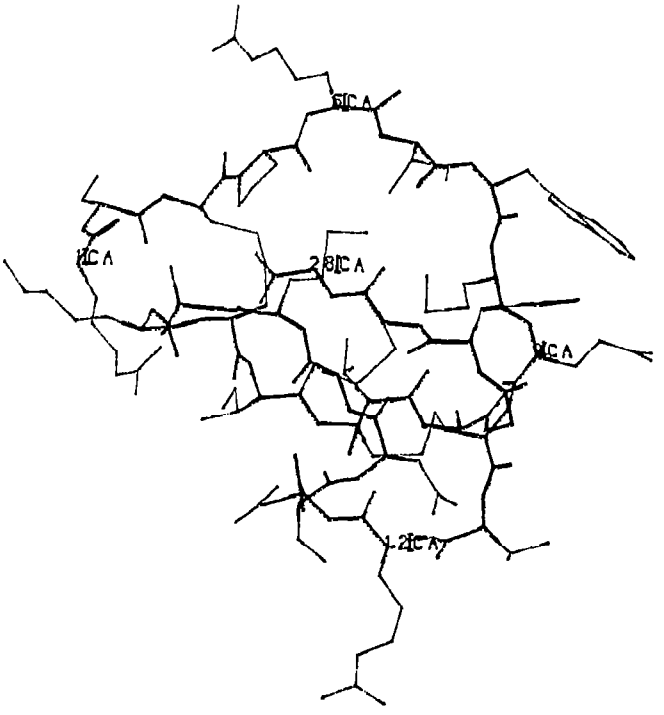

Fig. 4. Complete structure of MCTI-A. No solvent molecules are shown.

raphy. The high resolution of our reflection data and the tight binding between the inhibitor and trypsin were important in the success of this project. The result shows that is is easy to determine a porcine β-trypsin's structure based on bovine β-trypsin using molecular replacement method, though there are 39 different residues between them. Comparison of our MCTI-A sequence and MCTI-II obtained by Saburo Hara et al. [5] shows that these two inhibitors may be the same, since they have almost the same sequence.

In spite of the small size of this inhibitor, the interaction pattern is almost the same as that observed in other related complexes. There is considerable similarity in structure to Cucurbita Maxima inhibitor, CMTI-I, another member of the squash trypsin inhibitor family, though it is a residue shorter than CMTI-I. Like other squash inhibitors, MCTI-A contains a high amount of Cys amino residues (>20%). As observed in other high Cys content inhibitors such as Mung bean inhibitor [3,25], MCTI-A lacks any regular secondary structure. MCTI-A has the disulfide connectivities where the first three cysteine residues are covalently linked with the last three in consecutive order. These make an important contribution to stability of this inhibitor. The inhibition site Arg-5I between 2 pairs of disulfide bridges: Cys-3I–Cys-20I and Cys-10I–Cys-22I, at the second site beside the first disulfide pair. Those residues from Cys-3I to Cys-10I, Cys-20I–Cys-22I and two pairs of disulfide bridges can be thought of as a 12–14 amino acid residue ring, like those observed in Birman-Birk type inhibitors [3,25]. Since the relative disorder of the N-terminus, Arg-1I, the hypothesis suggested by W. Bode that the salt bridge of the Arg-1I side chain with the carboxy-terminus existing in CMTI-I promotes stability, was not supported by our results [23].

MCTI-A is suitable for synthetic studies because of its small size and strong trypsin inhibition. In our laboratory, the complex of trypsin and Mung bean inhibitor, a kind of Birman-Birk type inhibitor and its Lys active fragment have been studied [25]. Several small peptides consisting of a fragment of Mung bean inhibitor have been synthesized in our group. They show relatively weak inhibition of trypsin compared with the natural inhibitor. The crystals of trypsin and these synthesized peptides have been obtained. Combined with MCTI-A structure, the results should extend our understanding of trypsin–inhibitor complexes.

REFERENCES

[1] Ryan, C.A. (1981) in: The Biochemistry of Plants, vol. 6, pp. 351–369, Academic Press, New York.
[2] Chi, C. et al. (1982) in: Proteins in Biology and Medicine (Ralph, B., Chih, L. and Robert, H. eds.) pp. 341–362, Academic Press, New York.
[3] Chang, H., Chi, C. et al. (1979) Scientia Sinica, Vol. XXII no. 12, 1443–1454.
[4] Dung, L., Dominique, N. and Bertrand, C. (1989) Int. J. Peptide Protein Res. 34, 492–497.
[5] Saburo Hara et al. (1989) J. Biochem. 105, 88–92.
[6] Polanowski, A. et al. (1980) Acta Biochim. Polon. 27, 371–382.
[7] Wieczorek, M. et al. (1985) Biochem. Biophys. Res. Commun. 126, 646–653.
[8] Otlewski, J. et al. (1987) Biol. Chem. Hoppe-Seyler 369, 1505–1507.
[9] Siemion, I.Z., Wilusz, T. and Polanowski, A. (1984) Mol. Cell. Biochem. 60, 159–161.
[10] Hermodson, M.A., Ericsson, L.H., Newzath, H. and Walsh, K.A. (1973) Biochemistry 12, 3146.
[11] Sweet, R., Wright, H., Janin, J., Chothia, C. and Blow, D.M. (1974) Biochemistry, 13, 4212–4228.
[12] Steigeman, W. (1974) Ph.D. Thesis, TU München.
[13] Crowther, R. (1972) in: The Molecular Replacement Method, vol. 10 (Rossmann, M.G. ed.), Gordon and Breach, New York.
[14] Bode, W. and Schwager, P. (1975) J. Mol. Biol. 98, 693–717.
[15] Crowther, R. and Blow, D. (1967) Acta Crystallogr. 23, 544.
[16] Lattman, E. (1985) in: Methods in Enzymology, vol 115 (Wyckoff, H.W., Hirs, C.W. and Timasheff, N. eds.) pp. 55–77.
[17] Jones, T.A. (1978) J. Appl. Crystallogr. 11, 268–272.
[18] Pflugrath, J., Saper, M. and Quiocho in: Methods and Application in Crystallographic Computing (Halls, S. and Ashiaka, T. eds.) p. 407, Clarendon Press, London.
[19] Brunger, A., Kuriyan, J. and Karplus, M. (1987) Science 235, 458–460.
[20] Jack, A. and Levitt, M. (1978) Acta Crystallogr. A34, 931.
[21] Luzzati, V. (1952) Acta Cryst. 5, 802–810.
[22] Marquart, M., Walter, J., Deisenhofer, J., Bode, W. and Huber, R. (1983) Acta Crystallogr. B39, 480–490.
[23] Bode, W., Greyling, H.J., Huber, R., Otlewski, J. and Wilusz, T. (1989) FEBS Lett. 242, 285–292.
[24] Ramakrishnan, C. and Ramachandran, G.N. (1965) Biophys. J. 5, 909–933.
[25] Tang Youqi et al. (1986) Acta Physico-Chimica Sinica 2, 195–198.