Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the bifunctional inhibitor of trypsin and α-amylase from ragi seeds at 2.2 Å resolution

S. Gourinath,[a] Neelima Alam,[a] A. Srinivasan,[a] Ch. Betzel[b] and T. P. Singh[a]*

[a]Department of Biophysics, All India Institute of Medical Sciences, New Delhi-110029, India, and [b]Institute of Physiological Chemistry c/o DESY, Notkestrasse 85, 22603, Hamburg, Germany

Correspondence e-mail: tps@medinst.ernet.in

Received 11 June 1999
Accepted 20 December 1999

PDB Reference: trypsin/α-amylase inhibitor, 1b1u.

The crystal structure of a bifunctional inhibitor of α-amylase and trypsin (RATI) from ragi seeds (Indian finger millet, *Eleusine coracana* Gaertneri) has been determined by X-ray diffraction at 2.2 Å resolution. The inhibitor consists of 122 amino acids, with five disulfide bridges, and belongs to the plant α-amylase/trypsin inhibitor family. The crystals were grown by the microdialysis method using ammonium sulfate as a precipitating agent. The structure was determined by the molecular-replacement method using as models the structures of Corn Hageman factor inhibitor (CHFI) and of RATI at 2.9 Å resolution determined previously. It has been refined to an *R* factor of 21.9%. The structure shows an r.m.s. deviation for C^{α} atoms of 2.0 Å compared with its own NMR structure, whereas the corresponding value compared with CHFI is found to be 1.4 Å. The r.m.s. difference for C^{α} atoms when compared with the same protein in the structure of the complex with α-amylase is 0.7 Å. The conformations of trypsin-binding loop and the α-amylase-binding N-terminal region were also found to be similar in the crystal structures of native RATI and its complex with α-amylase. These regions differed considerably in the NMR structure.

1. Introduction

Protein inhibitors of proteinases are widespread in nature. The α-amylase/trypsin inhibitor family is a relatively new class of plant inhibitors. So far, 25 members of this family have been sequenced and the average number of residues found in them is 120 (Strobl *et al.*, 1995). This class of inhibitors contains a high number of cysteine residues, with 4–5 disulfide bridges (Maeda *et al.*, 1983; Poerio *et al.*, 1991). The members of the cereal inhibitor family are inhibitors of either α-amylase or trypsin. The bifunctional inhibitor from ragi (Indian finger millet, *E. coracana*) is the only inhibitor which inhibits both α-amylase and trypsin simultaneously. Earlier studies have clearly indicated that the inhibitor is capable of forming a ternary complex with α-amylase and trypsin (Shivraj & Pattabhiraman, 1981; Alagiri & Singh, 1993).

We have previously reported the crystal structure of RATI at 2.9 Å resolution (Gourinath *et al.*, 1999). An earlier structure solution was derived from an NMR model (Strobl *et al.*, 1995). We now report its structure at 2.2 Å resolution. The present structure was determined by the molecular-replacement method using as a model Corn Hageman factor inhibitor (CHFI), which has a sequence identity of 67%. The results of the 2.2 Å structure analysis are presented here.

2. Experimental

2.1. Crystallization and data collection

RATI was purified as described previously (Shivraj & Pattabhiraman, 1980; Srinivasan *et al.*, 1991; Gourinath *et al.*, 1999). Purified samples of the protein were dissolved in 20 m*M* sodium phosphate buffer pH 8.0, containing 0.3 *M* ammonium sulfate, to a final protein concentration of 10–15 mg ml^{-1}. The solution was equilibrated in microdialysis setups with 1.15 *M* ammonium sulfate in 0.4 *M* sodium phosphate buffer pH 8.0. Square-pyramidal crystals of RATI grew within 2–3 weeks. Typical crystal dimensions were approximately 0.3 × 0.2 × 0.15 mm. The crystals diffracted to 2.2 Å resolution. The data were collected on the EMBL beamline BW-7B at DESY, Hamburg with wavelength λ = 1.0 Å using a MAR Research MAR345 imaging-plate scanner and a 2.0° rotation for each image. A single crystal was mounted in a nylon loop and flash-frozen in a nitrogen stream at 100 K using 20–25% glycerol as a cryo-protectant while maintaining the ionic strength of the precipitant. The data were processed using *DENZO* and *SCALEPACK* program packages (Otwinowski & Minor, 1997). The results of the data collection and processing are given in Table 1. The data indicated a large mosaicity of 1.9°, presumably owing to freezing.

2.2. Structure determination and refinement

The structure reported earlier at 2.9 Å (Gourinath *et al.*, 1999) was determined using the NMR model of the cloned inhibitor (Strobl *et al.*, 1995). The structure of Corn Hageman factor inhibitor (CHFI) has been reported recently at 1.9 Å resolution (Behnke *et al.*, 1998). It has a 67% sequence identity with RATI. Therefore, it was decided to use the CHFI X-ray model as well as the 2.9 Å structure of RATI (Gourinath *et al.*, 1999) to determine the crystal structure of RATI using the new data set at 2.2 Å resolution. Both models gave the same solution. Subsequently, the solution obtained using

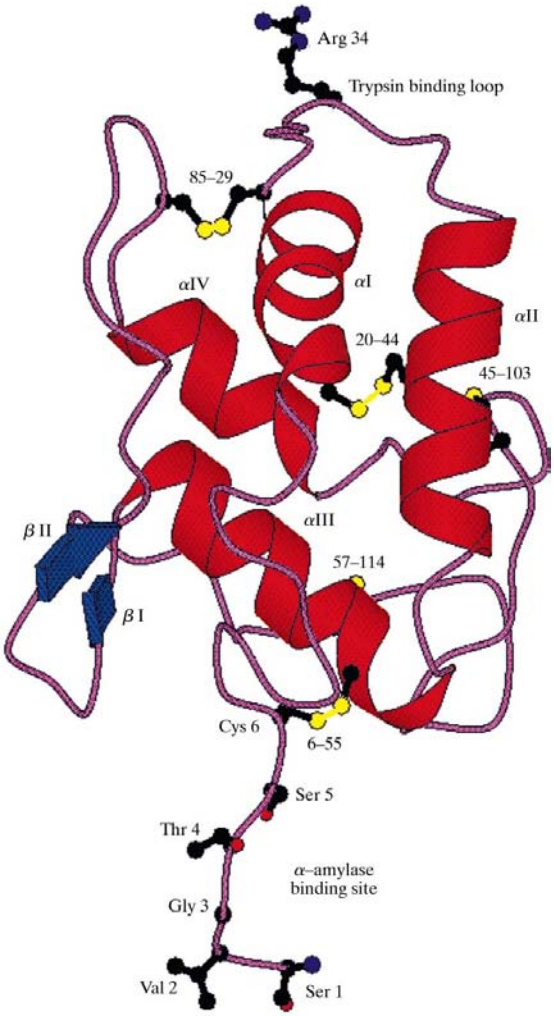

Figure 1
Ribbon diagram of RATI indicating secondary structures and binding sites for trypsin and α-amylase. The figure was drawn using *MOLSCRIPT* (Kraulis, 1991).

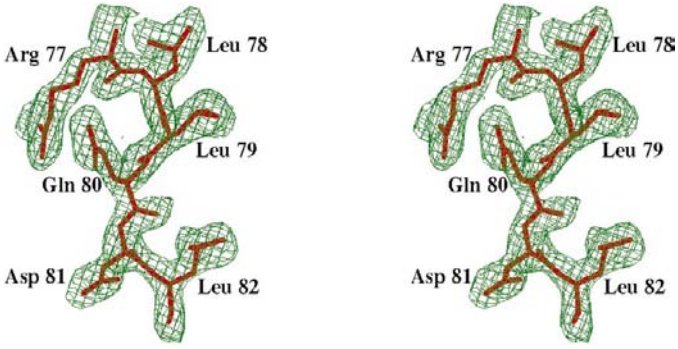

Figure 2
Stereoview of a section of the electron-density map ($2F_o - F_c$). The electron-density map is contoured at the 1.5σ level. The figure was obtained using the program *O*.

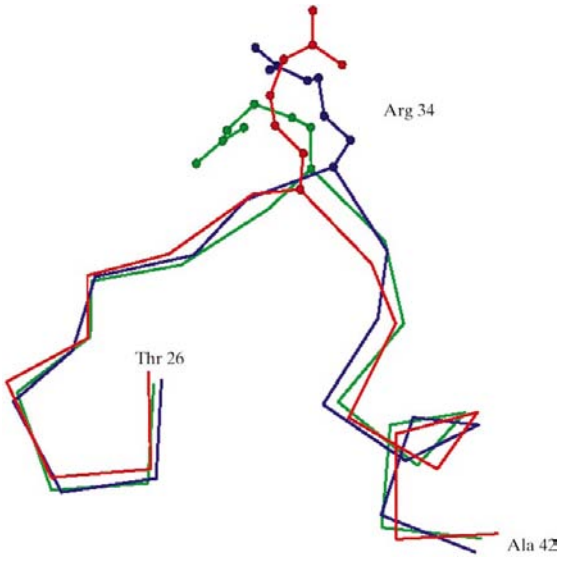

Figure 3
Superimposition of the C traces of the trypsin-binding loop of RATI from the native crystal structure (blue), crystal structure of complex (green) and the NMR structure in solution (red). The orientations of the side chain of Arg34 are also indicated in all three states.

Table 1
Crystal data for RATI.

Space group	$P2_12_12$
Unit-cell dimensions (Å)	a = 47.4, b = 54.5, c = 40.4
Resolution range (Å)	10.0–2.2
No. of observed reflections	57852
No. of unique reflections	5094
Completeness (%)	91
Highest resolution shell (2.3–2.2 Å)	83.5
R_{sym} (%)	7.2
Highest resolution shell	16.2
Mosaicity as determined by data processing (°)	1.9
V_m (Å^3 Da^{-1})	1.96
Solvent content (%)	37
Z (No. of molecules in the unit cell)	4

Table 2
Refinement statistics.

Refinement	
Resolution limits (Å)	10.0–2.2
No. of reflections	5094
R factor (%)	21.9
Free R factor (%)	27.7
R factor for all reflections	22.4
Overall G factor	0.15
Model	
Protein atoms	911
Water molecules	54
R.m.s. deviations from ideal values	
Bond lengths (Å)	0.009
Bond angles (°)	1.2
Dihedral angles (°)	26.9
Improper angles (°)	0.70
Average B factors (Å^2)	
Main-chain atoms	28.2
Side-chain atoms	27.7
All atoms	28.0

crystal structure model of CHFI was pursued. The rotation function was calculated using diffraction data in the resolution range 12–4 Å. The first 20 peaks of the rotation solution were used to compute the translation function. The first peak of the rotation solution appeared as a distinct peak in the output of the translation search, having a correlation coefficient of 42% and an R factor of 50%. In the subsequent rigid-body refinement, the correlation coefficient increased to 46% and the R factor fell to 48%. The solution was applied to the model coordinates using *LSQKAB* (Collaborative Computational Project, Number 4, 1994) and the transformed coordinates were used for the refinement.

After molecular replacement, two rounds of refinement using *X-PLOR* (Brünger *et al.*, 1987) following by manual rebuilding using the program *O* (Jones *et al.*, 1991) were performed on the protein structure. The refinement protocol in *X-PLOR* consisted of a rigid-body refinement (40 cycles) using all the data, followed by simulated annealing, a conventional positional refinement (50 cycles) and a restrained B-factor refinement (40 cycles). The simulated-annealing runs involved heating to 3000 K and then cooling in increments of 25 K, with 50 steps of molecular-dynamics simulation of 0.5 fs at each temperature. At this stage, the R factor and R_{free} [for 10% excluded data set; R_{free} reflections were selected using the *X-PLOR* program *FREEF(AG)*] were reduced to 0.25 and 0.35, respectively. The water molecules were located from a difference electron-density ($F_o - F_c$) map calculated at the 2.5σ level. Initially, they were only built into density peaks that clearly exhibited the appropriate shape. Those water molecules were retained for which the electron density in the subsequent $2F_o - F_c$ maps persisted after the refinement cycles and if they fulfilled the following criteria: (i) they were within 3.4 Å of the protein O or N atoms (or bound water) with good hydrogen-bonding geometry, (ii) their B factors were less than 45 Å^2 and (iii) they had real-space correlation coefficients above 65%. A total of 54 water molecules were included in the model. The refinement with further cycles and model building resulted in improvement of the model; the final R factor was 21.9% (Table 2).

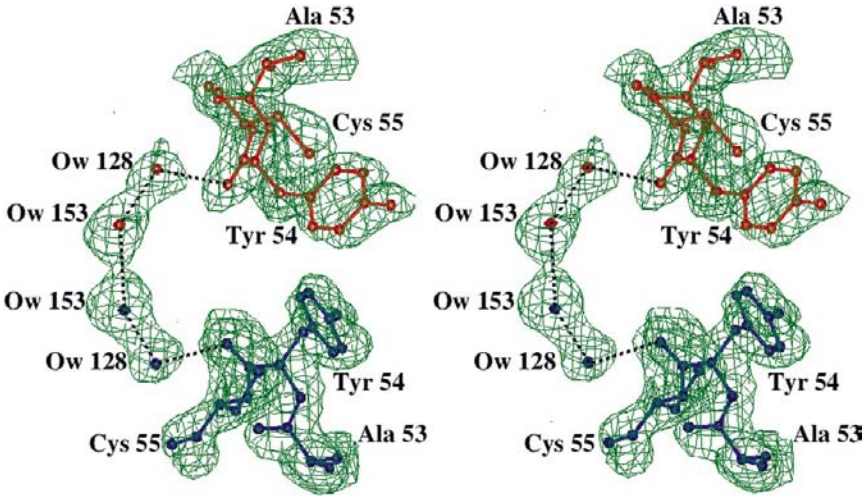

Figure 4
The electron density for a section of symmetry-related molecules showing a typical intermolecular interaction through solvent waters. The electron-density contours are drawn at the 1.5σ level.

3. Results and discussion

3.1. Quality of the refined model

The final model (Fig. 1) contains 911 protein atoms and 54 water molecules. The model has r.m.s. deviations of 0.009 Å for bond lengths and 1.2° for bond angles. The main-chain and side-chain atoms have well defined electron densities (Fig. 2). The present model at 2.2 Å resolution shows a remarkable improvement over the earlier reported structure at 2.9 Å resolution. The r.m.s. deviations in bond lengths improved from 0.02 to 0.009 Å and those of bond angles from 2.6 to 1.2°. The conformations of the trypsin-binding loop (26–42) and the N-terminal region (1–10) did not change

Table 3
The conformation of the trypsin-binding loop.

		X-ray structure of native RATI		X-ray structure of complex		NMR structure in solution		Trypsin inhibitors (Bode & Huber, 1992)	
Residue No.	Position	φ	ψ	φ	ψ	φ	ψ	φ	ψ
31	P4	−98.5	116.2	−102.8	131.3	−102 ± 3	157 ± 3		
32	P3	−107.2	149.7	161.7	−179.3	120 ± 3	180 ± 4	$-140 < \varphi < -120$	$140 < \psi < 170$
33	P2	−87.2	146.9	−61.3	160.5	−65 ± 1	143 ± 3	$-100 < \varphi < -60$	$139 < \psi < 180$
34	P1	−86.5	31	−87.7	67.7	−114 ± 2	45 ± 5	$-120 < \varphi < -95$	$9 < \psi < 50$
35	P1′	−125.2	150.6	−142.3	161.1	−75 ± 1	173 ± 3	$-100 < \varphi < -60$	$139 < \psi < 180$
36	P2′	−64.7	124.8	−51.0	124.4	−92 ± 1	62 ± 2	$-140 < \varphi < -99$	$70 < \psi < 120$
37	P3′	−61.4	−13.7	−56.8	−39.7	−50 ± 3	−39 ± 3	$-140 < \varphi < -99$	$70 < \psi < 120$
38	P4′	−76.5	−53.6	−58.4	−39.2	−61 ± 2	−37 ± 2		

significantly. The overall structure showed an r.m.s. difference of 1.3 Å. The corresponding r.m.s. differences between ten residues of N-terminus (α-amylase binding site) and the trypsin-binding loop (26–42) were 1.4 and 0.8 Å, respectively. The distribution of main-chain dihedral angles (φ, ψ) shows that more than 80% of the residues were found in the most favoured regions of the Ramachandran plot (Ramachandran & Sasisekaran, 1968).

3.2. Overall protein structure

RATI is folded into four α-helices (αI–αIV) and two short β-strands (βI–βII; Fig. 1). The α-helices are arranged in pairs in an antiparallel sense. Similarly, β-strands are also organized as two antiparallel strands. The helices follow an up-and-down topology, while β-strands are held with the help of a tight β-turn of type I.

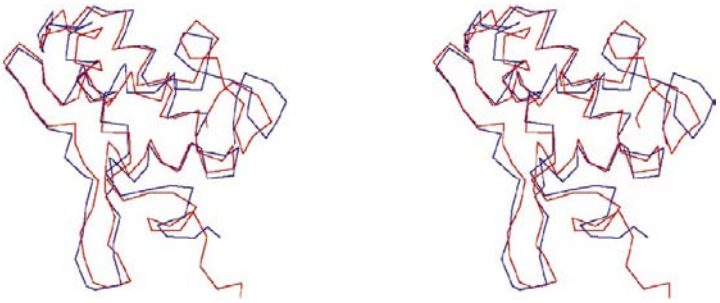

Figure 5
Stereoview of the superimposition of the C^{α} tracing of X-ray structure (red) and average NMR structure (blue) of RATI. The NMR structure contains residues 5–115. The r.m.s. shift for the C^{α} atoms is 2.0 Å.

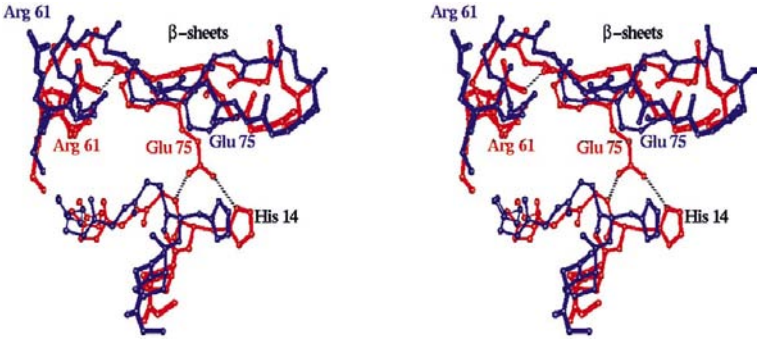

Figure 6
Stereoview showing some of the typical differences between the X-ray and NMR structures.

3.3. Trypsin-binding site

The binding site for trypsin occurs at a loop supported by two helices, αI and αII. This pair of helices is firmly held together by a disulfide bridge, Cys20–Cys44. The conformation of the trypsin-binding loop (Table 3) is very similar to the conformations observed for the trypsin-binding loops in other trypsin inhibitors (Bode & Huber, 1992), suggesting that the trypsin inhibitor binding site can be constructed using different scaffolding. The trypsin-binding loops (26–42) in native RATI and in the complex with α-amylase have been superimposed (Fig. 3). The r.m.s. difference for the main-chain atoms in the two structures is 1.0 Å. The orientation of Arg34 in the native crystal structure of RATI is similar to that of the complex with α-amylase, whereas it is oriented in the opposite direction in the NMR structure (Fig. 3).

3.4. α-Amylase-binding site

The α-amylase-binding site was not clearly established, as only relatively little is known about the mechanism by which protein inhibitors of α-amylases interact with their target enzymes. The only structural data available on the interactions between α-amylase and its inhibitor was from the crystal structure of the complex of porcine α-amylase and tendamistat (Wiegand *et al.*, 1995). The inhibitor tendamistat binds in an extended groove in the porcine pancreatic α-amylase. The triplet Trp-Arg-Tyr from the inhibitor is involved in the key interactions with the amylase. In the case of RATI, chemical modification experiments had indicated that a free amino group could be involved in the binding to mammalian α-amylase (Shivraj & Pattabhiraman, 1981; Alagiri & Singh, 1993). Recently, the crystal structure of a complex formed between α-amylase and RATI was reported (Strobl *et al.*, 1998). It showed that the free N-terminus was involved in amylase binding, which was completely different from all known inhibitors. In another recent crystal structure of a complex formed between barley amylase subtilisin inhibitor (BASI)

Table 4
Differences between X-ray and NMR structures of RATI.

	X-ray structure	NMR structure
βII	73–76	73–75
β-turn I	69–72	Highly distorted
	99–102	Absent
	107–110	Highly distorted
β-turn II	7–10	Absent
Fig. 6	Glu75 interacts with His14	Absent
	Arg61 interacts with Gly66	Absent

and barley amylase (Vallee *et al.*, 1998), the interactions were very different to those observed in the present structure, as the binding of BASI to α-amylase created a cavity exposed to the external medium and shaped to accommodate an extra calcium ion. This feature seems to contribute to the inhibitory effect, as the key amino-acid side chains of the active site were in direct contact with water molecules, which were in turn ligated to the calcium ions. In summary, it clearly meant that the rational prediction of a synthetic inhibitor with high binding constant was a difficult task.

3.5. Solvent structure

54 water molecules have been found in the crystal structure of RATI. The solvent content of only 37% in the RATI crystals suggested a rather tight packing of the protein molecules. There were few direct interactions involving protein atoms in the structure. It appeared that the small interstitial regions provided suitable conditions for solvent molecules to interact with neighbouring protein molecules, leading to a network of water-mediated intermolecular contacts (Fig. 4). This could explain the large observed mosaicity in RATI crystals.

3.6. Comparison of X-ray and NMR structures

The present structure of RATI was superimposed with the average structure derived from 20 NMR structures of RATI reported earlier (Strobl *et al.*, 1995), using *LSQMAN* (Kleywegt & Jones, 1996) to minimize the r.m.s. distance of the two structures. This gave an r.m.s. difference of 2.0 Å for the C^{α} atoms (Fig. 5). An overall displacement including the side chains was found to be 2.5 Å. The two α-helices αIII and αIV are longer in the X-ray structure than in the NMR models. The secondary structures were more clearly defined in the X-ray analysis. The important differences between these structures are presented in Table 4 and Fig. 6.

3.7. RATI in the complex with α-amylase

The structure of a complex formed between RATI (cloned protein) and α-amylase from yellow meal worm has been reported very recently (Strobl *et al.*, 1998). The most significant difference between the two structures is observed in the N-terminal region, which is involved in the binding with α-amylase (Fig. 7). The r.m.s. differences for the C traces between the first ten residues of the two structures was 1.4 Å. The r.m.s. difference of the C traces for the full structures is 0.7 Å. The structure of the molecule in the free state is similar to its conformation in the complexed state. The residues (1–10) at the N-terminus of RATI in the complex adopt a 3_{10}-helical conformation, while in the native RATI they exist as a loose structure like a wagging tail. The N-terminal residues in the native structure are involved in three intermolecular hydrogen bonds: Ser1 $O^{\gamma}\cdots$Ala106 N (symmetry related), 3.2 Å; Ser5 $O^{\gamma}\cdots$Thr107 O (symmetry related), 3.0 Å; Cys6 N$\cdots$Ile108 O (symmetry related), 3.2 Å. In the complex, the N-terminal part of RATI is involved in extensive interactions with α-amylase. The superposition of first ten residues of RATI in the native structure and in the complex is shown in Fig. 8.

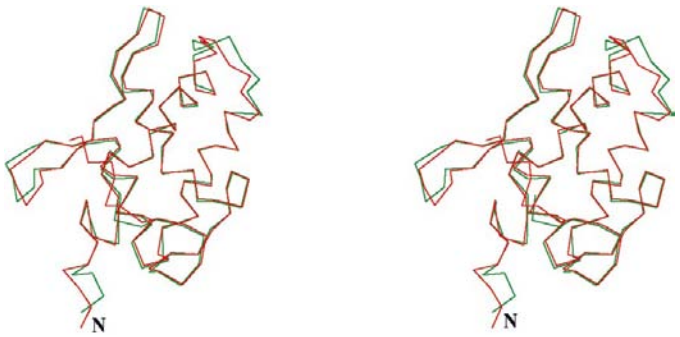

Figure 7
Stereoview of the superimposition of C^{α} traces of native RATI (present structure; red) and RATI in the complex with α-amylase (green). The N-terminal region which is involved in the binding with α-amylase adopts a slightly different conformation in the complex.

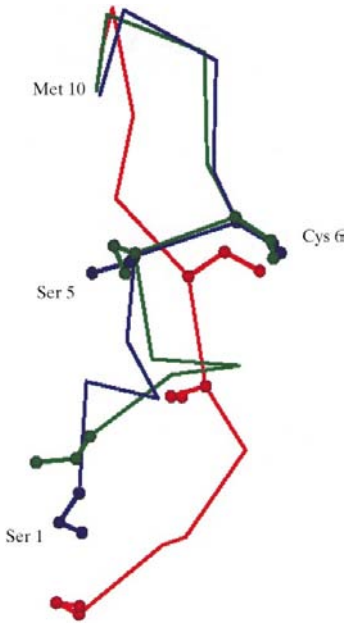

Figure 8
Superimposition of the C^{α} traces of the α-amylase binding region of the native (blue) and the complex (green) crystal structures and the NMR structure in solution (red).

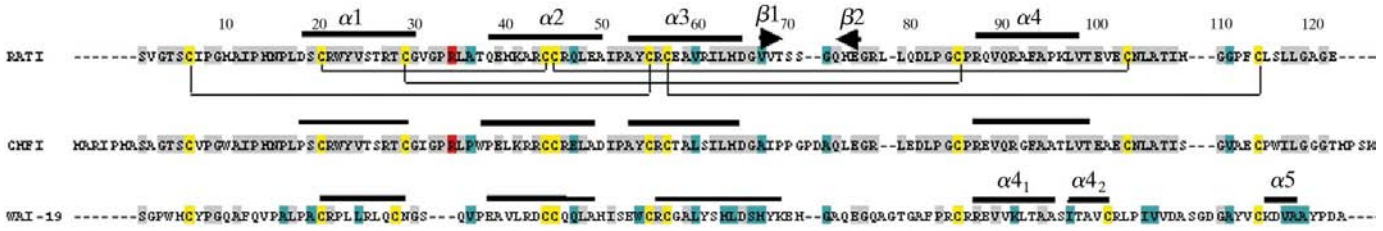

Figure 9
The sequence comparison of RATI, CHFI and WAI-0.19. The identical residues in the sequences are indicated in green. RATI and CHFI have 67% sequence identity, while RATI and WAI-0.19 show only 24% sequence identity. The cysteine residues are indicated in yellow. All the three structures have identical distributions of disulfide bridges.

3.8. Structural comparison with other cereal inhibitors

So far, crystal structures of only three cereal inhibitors have been determined: RATI (present work), Corn Hageman factor inhibitor (CHFI) and 0.19 wheat α-amylase inhibitor (WAI-0.19). RATI has a sequence identity of 67% with CHFI and only 24% with WAI-0.19 (Fig. 9). The superimposition of C^{α} traces of RATI and CHFI shows an r.m.s. deviation of 1.4 Å. The corresponding r.m.s. deviation with WAI-0.19 is 1.7 Å. The two antiparallel β-sheets between residues 67 and 69, and 73 and 75 are absent in both CHFI and WAI-0.19. A critical interaction involving the side chain of Arg61 and the backbone of Gly66 in RATI is missing in CHFI, as Arg61 is replaced by Ser in CHFI. In RATI, Met40 interacts with Pro33 to stabilize the trypsin-binding loop. This interaction is missing in CHFI, as Met40 is replaced by Leu. In the structure of RATI, Glu102 forms hydrogen bonds with Tyr23 (OH) and Arg27 (NH), while Phe93 and Tyr23 are involved in hydrophobic interactions. On the other hand, in WAI-0.19, as the Phe and Tyr are missing, Gln27 moves in to the protein core, thus disturbing the canonical structure of the trypsin-binding loop.

3.9. Comparison with other proteins

The overall structural topology of RATI has been found to be similar to a hydrophobic protein from soya bean (HPS; Baud *et al.*, 1993) and the maize non-specific lipid-transfer protein (M-NSLTP; Shin *et al.*, 1995). The pattern of disulfide bridges is slightly different, but three of the four α-helices match very closely. The sequence homology between RATI and these other proteins is only ~20%. The structure of RATI has also been found to be similar to 2S seed storage protein (Rico *et al.*, 1996). The sequence alignment of RATI with 2S protein shows that six out of eight Cys residues are well aligned.

4. Conclusions

The core of RATI consists of a globular four-helix motif with a simple up-and-down topology. The helices are formed between residues 18 and 30, 38 and 50, 53 and 65, and 87 and 98, and two short antiparallel β-sheets are formed between residues 67 and 70, and 73 and 75. There are two wide loops between residues 31 and 36, and 77 and 86. There are no similarities between the secondary-structure topology of RATI and other members of serine-proteinase protein inhibitors for which three-dimensional structures are known. The scaffold stabilizing the trypsin-binding loop of RATI exhibited a new motif, quite different from other serine-proteinase inhibitors, in which its binding loop is positioned between the two α-helices. The crystal structure of RATI in its native and complexed form with α-amylase showed overall similar conformations, with an r.m.s. difference of 0.7 Å. The X-ray structure in single crystals and NMR models in solution show an overall similar folding. However, significant differences were observed in the loop regions and peripheral parts of the protein. The lengths of the helices were found to be longer in the X-ray structure than in the NMR analysis. The surface residues in the two structures showed large displacements (up to 4.0 Å). The structures of CHFI and WAI-0.19 differed from the structure of RATI by r.m.s. deviations of 1.4 and 1.7 Å, respectively, for the C^{α} atoms. This indicated that its own NMR structure showed larger deviations than the X-ray structures of other family members. The difference between the two structures reflects the potential difference of the structures in two states.

Diffraction data for this study were collected at EMBL using beamline BW-7B of HASYLab at DESY, Hamburg. The beamtime extended by the EMBL Outstation, Hamburg is gratefully acknowledged. SG thanks the German Academic Exchange Service, New Delhi for the award of a fellowship for the data collection and the Council of Scientific and Industrial Research, New Delhi for the fellowship.

References

Alagiri, S. & Singh, T. P. (1993). *Biochim. Biophys. Acta*, **1203**, 77–84.
Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S. & Lehman, M. (1993). *J. Mol. Biol.* **231**, 877–887.
Behnke, C. A., Yee, V. C., Trong, I. L., Pedersen, L. C., Stenkamo, R. E., Kim, S. S., Reeck, G. R. & Telle, D. C. (1998). *Biochemistry*, **37**, 15277–88.
Bode, W. & Huber, R. (1992). *Eur. J. Biochem.* **204**, 433–451.
Brünger, A. T., Campbell, R. T., Marius Clore, G., Gronenborn, A. M., Karplus, M., Petsko, G. A. & Teeter, M. M. (1987). *Science*, **235**, 1049–1053.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Gourinath, S., Srinivasan, A. & Singh, T. P. (1999). *Acta Cryst.* D**55**, 25–30.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 826–828.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Maeda, K., Kakabayashi, S. & Matsubara, H. (1983). *J. Biochem.* **94**, 865–870.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Poerio, E., Caporale, C., Carrano, L., Pucci, P. & Buonocore, V. (1991). *Eur. J. Biochem.* **199**, 595–600.
Ramachandran, G. N. & Sasisekaran, V. (1968). *Adv. Protein Chem.* **23**, 283–438.
Rico, M., Bruix, M., Gonzalez, C., Monsalve, R. I. & Rodriue, R. (1996). *Biochemistry*, **35**, 15672–15682.
Shin, D. H., Lee, J. Y., Hwang, K. Y., Kim, K. K. & Suh, S. W. (1995). *Structure*, **3**, 189–199.
Shivraj, B. & Pattabhiraman, T. N. (1980). *Ind. J. Biochem. Biophys.* **17**, 181–185.
Shivraj, B. & Pattabhiraman, T. N. (1981). *Biochem. J.* **193**, 29–36.
Srinivasan, A., Raman, A. & Singh, T. P. (1991). *J. Mol. Biol.* **22**, 1–2.
Strobl, S., Maskos, K., Wiegand, G., Huber, R., Gomis-Ruth, F. X. & Glockshuber, R. (1998). *Structure*, **6,** 911–21.
Strobl, S., Muhlhahn, P., Bernstein, R., Wiltscheck, R., Maskos, K., Wunderlich, M., Huber, R., Glockshuber, R. & Holak, T. A. (1995). *Biochemistry*, **34**, 8281–8293.
Vallee, F., Kadziola, A., Bourne, Y., Juy, M., Rodenburg, K. W., Svensson, B. & Haser, R. (1998). *Structure*, **6**, 649–59.
Wiegand, G., Epp, O. & Huber, R. (1995). *J. Mol. Biol.* **247**, 99–100.