Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the bifunctional inhibitor of trypsin and α-amylase from ragi seeds at 2.9 Å resolution

S. Gourinath, A. Srinivasan and T. P. Singh*

Department of Biophysics, All India Institute of Medical Sciences, New Delhi 110029, India

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

Received 27 February 1998
Accepted 30 April 1998

PDB Reference: trypsin inhibitor, 1blu.

The crystal structure of a bifunctional inhibitor of α-amylase and trypsin from the seeds of ragi (Indian finger millet, *Eleusine coracana* Gaertneri) has been determined by an X-ray diffraction method. The inhibitor consists of 122 amino acids with five disulfide bridges and belongs to the plant α-amylase/trypsin-inhibitor family. This is the first crystal structure determination of a member of this family. The protein, purified from the seeds of ragi, has a molecular mass of 13300 Da with a p*I* of 10.3. Crystals were grown by a microdialysis method using ammonium sulfate as precipitant. The improved purification protocol and the modified crystallization conditions enabled reproducible growth of the crystals. The cell parameters are a = 41.2, b = 47.4, c = 55.9 Å. The intensity data were collected to 2.9 Å resolution, and the crystal structure was determined using the molecular-replacement method. The structure was refined using the *X-PLOR* and *CCP*4 program packages to a conventional R factor of 21%. The structure contains four α-helices between residues 19–29, 37–51, 56–65 and 90–95, and two short antiparallel β-strands between residues 67–70 and 73–75.

1. Introduction

Protein inhibitors of proteinases are widespread in nature. The primary structures of numerous inhibitors of different classes of proteinases have been determined and the information has revealed a number of families of inhibitors whose members are related by extensive homology and topological relationships between their disulfide bridges and the location of reactive (inhibitory) sites. On the basis of their sequence identities, they can be divided into various distinct classes (Laskowski & Kato, 1980; Laskowski, 1986; Laskowski *et al.*, 1987; Richardson, 1991; Strobl *et al.*, 1995). Three-dimensional structures of several of these inhibitors have been reported (Bode & Huber, 1992). They contain identical binding loops but their overall conformations are very different.

The proteinase inhibitors are found in bacteria as well as in eukaryotic cells. These inhibitors are also widely distributed in the plant kingdom. Most of the plant inhibitors are present in the seeds of various plants, but are not necessarily restricted to this part of the plant. The inhibitors are diverse in number and in specificity toward various proteolytic enzymes.

The α-amylase/trypsin-inhibitor family, which is a member of the cereal inhibitors, is a relatively new class of plant inhibitors. So far, 24 members of this family have been sequenced and the average number of residues found in them is about 120 (Strobl *et al.*, 1995). This class of inhibitors contains high numbers of cysteine residues and four or five disulfide bridges (Maeda *et al.*, 1983; Poerio *et al.*, 1991). The

© 1999 International Union of Crystallography
Printed in Great Britain – all rights reserved

members of the cereal inhibitor family are either inhibitors of mammalian α-amylase or trypsin. So far, the structures of only three α-amylase inhibitors of plant origin have been determined (Zemke *et al.*, 1991; Oda *et al.*, 1997; Nakaguchi *et al.*, 1997). The primary sequence of ragi α-amylase/trypsin inhibitor (RATI) has been determined (Campos & Richardson, 1983). It contains 122 amino acids with ten cysteine residues forming five intramolecular disulfide linkages. RATI shares between 22 and 66% sequence identity with other members of the family (Strobl *et al.*, 1995). Earlier studies suggest that it binds to trypsin and α-amylase simultaneously, as the inhibitor is capable of forming a ternary complex with α-amylase and trypsin (Shivraj & Pattabiraman, 1981; Alagiri & Singh, 1993).

This is an important class of plant inhibitors. To date, no crystal structure of any member of this family has been determined. Therefore, we decided to determine the three-dimensional structure of RATI by an X-ray diffraction method. We have previously reported the preliminary X-ray crystallographic data (Srinivasan *et al.*, 1991), but the crystals could not be obtained reproducibly. Therefore, it was necessary to alter the purification protocol and modify the crystallization conditions to generate crystals reproducibly. Recently, the nuclear magentic resonance (NMR) structure of a cloned RATI has also been reported (Strobl *et al.*, 1995). In the NMR model, residue 70 was taken as Ser whereas the chemical-sequence data indicate it to be either Pro or Ser. The trypsin-binding loop of RATI was reported to adopt the 'canonical' substrate-like conformation, which is highly conserved among the serine proteinase inhibitors. The binding loop was found to be stabilized by two adjacent helices. The RATI fold is expected to be general for all members of the RATI family because conserved residues among the members of the family form the core of the structure. In this paper, we report the new purification protocol, successful crystallization conditions and new crystal data, describe the use of NMR models for determining the initial phases for X-ray structure analysis, and discuss the three-dimensional structure of the protein.

2. Experimental

2.1. Crystallization and data collection

The initial steps of the purification of RATI were carried out as described previously (Shivraj & Pattabiraman, 1980; Srinivasan *et al.*, 1991). To this purification protocol, an ion-exchange chromatography step with a pH-gradient elution was added. The protein was bound at pH 5.0, using 2 m*M* acetate buffer and 75 m*M* NaCl. The bound protein was eluted with a pH gradient from pH 5.0 to 11.0 using 5 m*M* NaH_2PO_4 and 2.5 m*M* Na_3PO_4 solutions. The lyophilized sample of the protein was 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}. It was equilibrated in microdialysis set-ups with 1.15 *M* ammonium sulfate in 0.4 *M* sodium phosphate buffer (pH 8.0). The square-pyramidal-shaped crystals of RATI grew within 2–3 weeks. The improved purification procedure and modified crystallization conditions led to reproducible crystal growth. Typical crystal dimensions were of the order of 0.3 × 0.2 × 0.15 mm. Although these crystals were relatively small, they diffracted well up to 2.9 Å resolution. The diffraction data were collected on a MAR Research imaging-plate scanner using graphite-monochromated Cu *Kα* radiation generated by a rotating-anode generator (Rigaku, RU200) operating at 100 mA, 40 kV, with a focal point of 0.3 × 3 mm. Each data image covered a rotation angle of 2.0° and was exposed for 20 min. The intensity data were processed using the *DENZO* and *SCALEPACK* program packages (Otwinowski, 1993; Minor, 1993). A number of rather weak reflections could not be estimated, thus limiting the overall completeness to 86%. The results from the data collection are given in Table 1. As seen from Table 1, the unit-cell dimensions of the present crystal form are very different to those reported earlier (Srinivasan *et al.*, 1991). These crystals, with 40.2% solvent content, represent one of the tightly packed crystal forms.

Table 1
Characteristics of crystal data.

Space group	$P2_12_12_1$
Unit cell (Å)	a = 41.2, b = 47.4, c = 55.9
Cell dimensions of the form reported previously† (Å)	a = 30.5, b = 56.3, c = 73.6
V (Å^3)	1.1×10^5
V_m (Å^3 Da^{-1})	2.07
Z (number of molecules in the unit cell)	4
Solvent (%)	40.2
Resolution (Å)	2.9
$R_{\rm merge}$‡ (%)	7.8
Total observations	1876
Unique observations	2218
Multiplicity	8.9
Completeness (%)	86
Completeness of data in 3.4–2.9 Å resolution cell (%)	87

† Srinivasan *et al.* (1991). ‡ $R_{\rm merge} = \sum(|\langle I\rangle - I_i|)/\sum I_i$, summed over all symmetry-related reflections.

2.2. Structure solution and refinement

In the Protein Data Bank, no other structure was found that could readily lead to derivation of the initial phases using the molecular-replacement method. The NMR structure of RATI using the cloned protein has been reported (Strobl *et al.*, 1995). Use of the solution structure derived from NMR experiments as a search model was considered as a possible starting point for attempting the crystal structure determination by molecular replacement. This methodology has yet to develop into a regular approach for structure determination but a few recent reports (Brünger *et al.*, 1987; Braun *et al.*, 1989; Baldwin *et al.*, 1991; Janes *et al.*, 1994; Weiss *et al.*, 1995; Hyvonen *et al.*, 1995; Müller *et al.*, 1995; Wilmanns & Nilges, 1996; Mittl *et al.*, 1998) have established the feasibility of utilizing an NMR-derived model as a starting point for solving an X-ray crystal structure by molecular replacement. With the use of NMR techniques, many sets of conformers are obtained in solution. The individual NMR structures may not be suffi-

ciently accurate to solve the rotation and translation functions uniquely and, additionally, an average of various NMR structures may give a distorted model. Nevertheless, the approaches which seem to work in finding suitable NMR models for crystal structure determination are: (i) use of the restrained minimized mean NMR structure (Brünger *et al.*, 1987; Baldwin *et al.*, 1991; Wilmanns & Nilges, 1996) and (ii) use of a model with the atoms of the backbone and those of the side chains of the interior part of the molecule (Braun *et al.*, 1989). There are not enough examples to suggest the superiority of one approach over the other. Moreover, most of the examples available so far were based on known crystal structures. Therefore, both approaches for crystal structure determination using NMR solution structures may need to be investigated.

We have determined the crystal structure of RATI by the molecular-replacement method with the program *AMoRe* (Navaza, 1994) using an NMR solution model (Strobl *et al.*, 1995). There were 20 NMR structures of this protein available. These models were deployed for calculation of rotation and translation functions. Rigid-body refinements were also attempted for all the models. In all calculations, reflections were used in the resolution range 15.0–4.50 Å. None of the 20 NMR models gave a distinct rotation solution. The translation and fitting calculations from the first 19 models did not give any visible indication of a correct solution. The correct solution emerged from the 20th NMR model. The correct solution stood at the 13th position in the rotation-function calculations with a correlation coefficient (CC) of 35.3% as defined in the program *AMoRe* (Navaza, 1994). This solution moved to first position after calculation of the translation function, and had an *R* factor of 49.7% and a CC of 34.2%. The model was further refined as a rigid body to an *R* factor of 46.8% and a CC of 43.6%. The structure of RATI was refined by a combined iterative procedure that included model rebuilding with the program *O* (Jones *et al.*, 1991). Molecular-dynamics refinement using *X-PLOR* (Brünger, 1993), restrained least-squares refinement (Hendrickson & Konnert, 1981) and map improvement using $(2F_o - F_c)$ and $(F_o - F_c)$ Fourier calculations using the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) were carried out. The electron-density maps obtained from the starting correct solution indicated characteristic features for all aromatic side chains and for most parts of the backbone. Furthermore, the correct placements of disulfide bridges were also indicated. The subsequent model building and refinement were carried out from the correctly placed side chains and the backbone. The remaining side-chain atoms were slowly introduced into the model. There was an uncertainty in the sequence of RATI about the residue at position 70. The sequence data reported this residue as Pro/Ser (Campos & Richardson, 1983). The cloning was performed with Ser, for

Table 2
Refinement statistics.

Resolution limits (Å)	17.0–2.9
Protein atoms	912
R factor	0.21
Free *R* factor	0.31
Positional error estimate (Å)	0.20

Stereochemistry	Estimated deviations
Bond length (1−2) (Å)	0.024
Angle distance (1−3) (Å)	0.046
Planar distance (1−4) (Å)	0.068
Chiral volumes (Å^3)	0.19
Planar torsion angle (°)	5.6
Deviation from planes	0.030

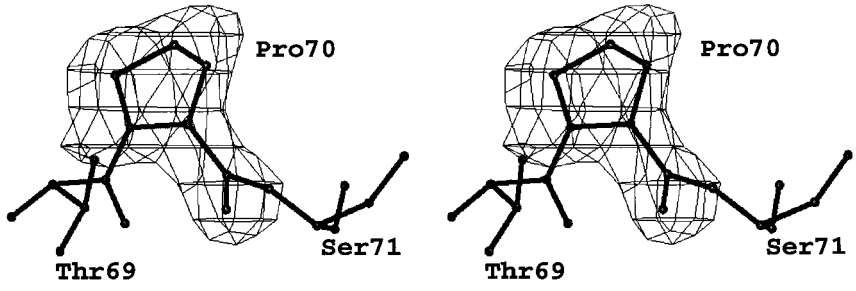

Figure 1
$(F_o - F_c)$ difference electron-density stereo map of residue 70 after the residue was removed in the OMIT map. Electron density is contoured at 1.5σ.

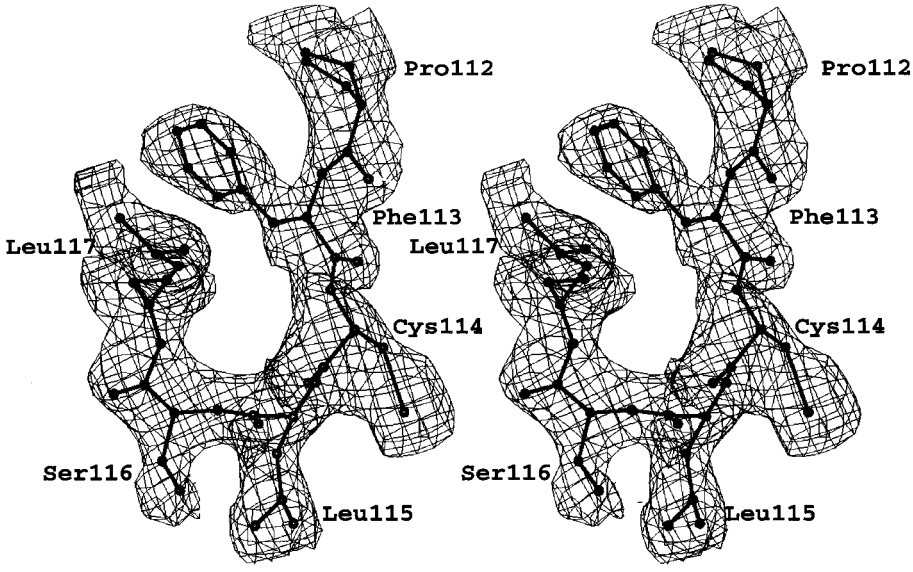

Figure 2
A representative stereo plot of the final $(2F_o - F_c)$ electron-density map of RATI contoured at 1σ.

which the NMR structure has been reported (Strobl *et al.*, 1995). To resolve this ambiguity, an OMIT map was calculated by deleting residue 70. The OMIT map (Fig. 1) clearly indicated the presence of proline at position 70. The refinement was continued with simulated annealing using *X-PLOR* (Brünger, 1993) on a Silicon Graphics workstation. Initially, the temperature factors were set uniformly to 15 Å^2. After each manual intervention using the program *O* (Jones *et al.*, 1991), the structure significantly improved after 40 cycles of restrained least-squares refinement and slow cooling from 3000 to 300 K. After an additional 80 cycles of least-squares refinement, 20 cycles of individual *B*-factor refinement were performed. The final *R* factor for all data was 0.21 and the free *R* factor was 0.31 (Table 2).

3. Results and discussion

3.1. Quality of the refined model

The refined model contains 912 non-H protein atoms accounting for all 122 residues. It gives an *R* factor of 21% for all 2218 unique reflections in the resolution range 17.0–2.9 Å. The r.m.s. deviations from ideal stereochemistry are 0.024 Å for bond lengths and 1.6° for bond angles. The average positional error is estimated to be 0.20 Å from the Luzzati plot (Luzzati, 1952). All the backbone torsion angles for non-glycine residues lie within the allowed regions of the Ramachandran plot (Ramakrishnan & Ramachandran, 1965). The entire polypeptide chain of RATI is represented by a reasonably defined electron density. A representative stereo plot of the final electron-density map ($2F_o - F_c$) for RATI is illustrated in Fig. 2. The average *B* factor for all 912 atoms is 24.7 Å^2.

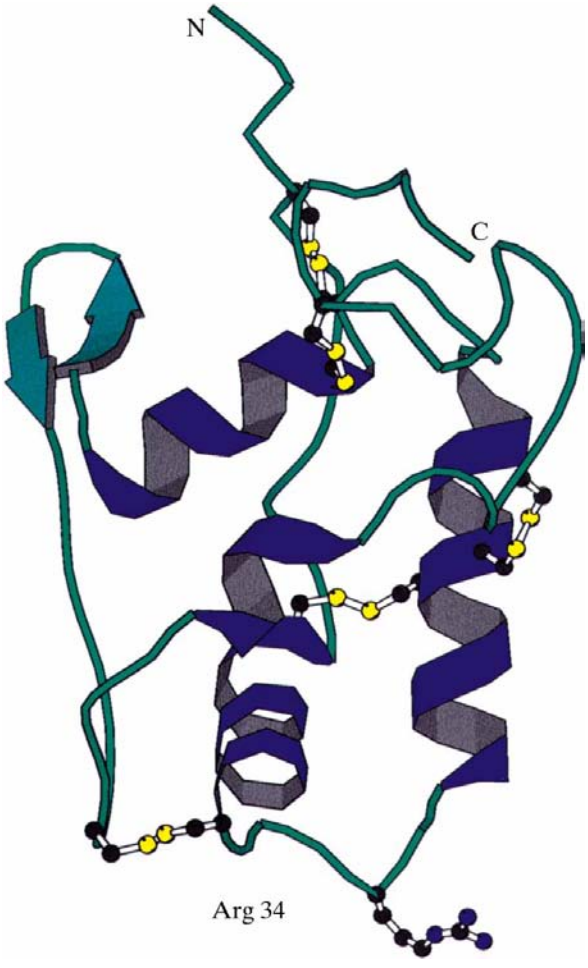

Figure 3
A view of the protein structure drawn using the program *MOLSCRIPT* (Kraulis, 1991). The disulfide links are shown in yellow. The trypsin binding loop with Arg34 is also indicated.

3.2. Overall protein structure

As shown in Fig. 3, RATI has a low secondary-structure content with four α-helices formed between residues 19–29 (α1), 37–51 (α2), 56–65 (α3) and 90–95 (α4), and two antiparallel β-strands between residues 67–70 (α1) and 73–75 (α2). The total secondary-structure content is 35% α-helix and 7% β-pleated sheet. The helices form the core of the molecule in a simple up-and-down topology. The helices are packed in an extremely tight manner. They are connected by large flexible loops. Two different criteria have been used to define the range of each helix. One is based on the Cα conformation (Levitt & Greer, 1977) which is consistent with a visual inspection of the molecule, while the other is based on the hydrogen-bonding pattern.

The polypeptide chain of RATI has five disulfide bridges (Cys6—Cys55, Cys20—Cys44, Cys29—Cys85, Cys45—Cys103 and Cys57—Cys114). The disulfide bonds are well spread in the structure (Fig. 3) and are positioned to hold various loops and helices together. The low secondary-structure content in this protein is compensated by a relatively large number of disulfide bonds. The residues adjacent to the disulfide bonds have relatively low *B* factors. It is clearly seen that the disulfide bonds impart stability to the structure by bringing together various secondary-structure elements, such as helix–helix, helix–loop and loop–loop.

3.3. Trypsin-binding loop

The trypsin-binding loop corresponds to Cys29—Ala36 and is well defined in the electron density. An entirely new structural motif has been observed for the trypsin-binding loop in RATI (Fig. 3). The flexible loop is held like a string between two antiparallel α-helices. The loop is also held at one end by a disulfide bridge involving Cys29—Cys85. The helices holding the loop are also held together by a disulfide bond Cys20—Cys44. The conformation of the binding loop corresponds to the conformation observed for other serine proteinase inhibitors, where the conformation of residues surrounding the scissile bond enables tight binding of the inhibitor to the target proteinase in a substrate-like manner. The scissile bond (Arg34—Leu35) protrudes from the surface of the molecule. The position of Arg 34 is indicated in Fig. 3.

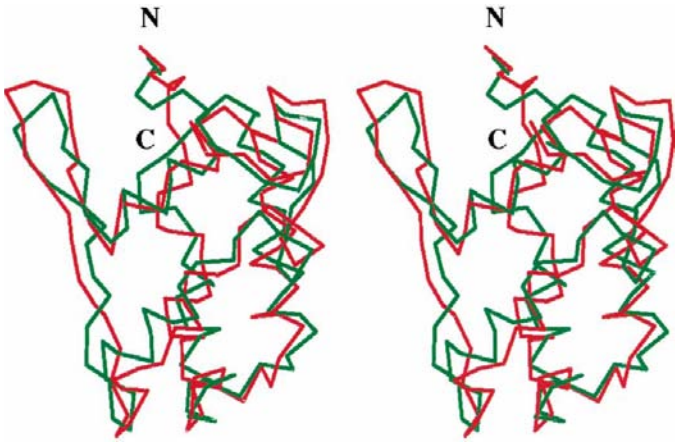

Figure 4
Stereoscopic superposition of the Cα backbone representation of RATI (red line) and the NMR model (green line).

3.4. α-Amylase binding site

The α-amylase binding site is not yet fully established, as relatively little is known about the mechanism by which protein inhibitors of α-amylases interact with their target enzymes. The only structural data available so far are from the three-dimensional crystal structure of the complex of porcine α-amylase and tendamistat (Wiegand *et al.*, 1995). The chemical-modification experiments indicate that a lysine residue or residues may be involved in the binding to mammalian α-amylase (Shivraj & Pattabiraman, 1981; Alagiri & Singh, 1993). There are only two Lys residues in RATI, namely Lys41 and Lys96, and the side chains of both are placed in close proximity, on the opposite side to the trypsin-binding loop, thus making it a possible site for α-amylase binding.

3.5. Comparison of X-ray and NMR structures

The comparison of the global molecular architecture in single crystals (present structure) and in solution, where the solution structure is characterized by a group of 20 conformers calculated by a hybrid distance geometry simulated-annealing method (Holak *et al.*, 1988, 1989) with the program *X-PLOR* 3.1 (Brünger, 1993), is shown in Fig. 4. The backbone atoms Cα of the NMR model (the model which was used to derive the initial phases) are superimposed on the X-ray structure. Though the secondary structures are preserved in the two states, the average displacements of the backbone atoms are significantly large (r.m.s displacement for Cα atoms = 2.5 Å). Larger deviations occur in the peripheral and loop regions. The r.m.s. deviations for backbone atoms within the 20 NMR structures is less than that observed between the X-ray structure and any one of the 20 NMR models (r.m.s. deviation for Cα atoms for various NMR models = 0.8 ± 0.2 Å).

The authors thank S. Karthikeyan for his help with the computational work and for many helpful discussions. The financial support from the Department of Science and Technology, New Delhi, is gratefully acknowledged. SG thanks the Council of Scientific and Industrial Research, New Delhi, for a fellowship.

References

Alagiri, S. & Singh, T. P. (1993). *Biochim. Biophys. Acta*, **1203**, 77–84.
Baldwin, E. T., Weber, I. T., Charles, R. S., Xuan, J. C., Appella, E., Yamada, M., Matsushima, K., Edwards, B. F. P., Marius Clore, G., Groneborn, A. M. & Wlodawer, A. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 502–506.
Bode, W. & Huber, R. (1992). *Eur. J. Biochem.* **204**, 433–451.
Braun, W., Epp, O., Wuthrich, K. & Huber, R. (1989). *J. Mol. Biol.* **206**, 669–676.
Brünger, A. T. (1993). *X-PLOR Version 3.1 Manual.* New Haven: Yale University.
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.
Campos, F. A. P. & Richardson, M. (1983). *FEBS Lett.* **152**, 300–304.
Collaborative Computational Project, Number 4. (1994). *Acta Cryst.* D**50**, 760–763.
Hendrickson, W. A. & Konnert, J. H. (1981). *Biological Structure, Function, Conformation and Evolution*, Vol. 1, edited by R. Srinivasan, pp. 43–47. London: Pergamon.
Holak, T. A., Gondol, D., Otlewski. J. & Wilusz, T. (1989). *J. Mol. Biol.* **210**, 635–648.
Holak, T. A., Kearsley, S. K., Kim, Y. & Prestegard, J. H. (1988). *Biochemistry*, **27**, 6135–6142.
Hyvonen, M., Macias, M. J., Nilges, M., Oschkinat, H., Saraste, M. & Wilmanns, M. (1995). *EMBO J.* **14**, 4676–4685.
Janes, R. W., Peapus, D. H. & Wallace, B. A. (1994). *Nature Struct. Biol.* **1**, 311–319.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, M. Jr (1986). *Toxicological Significance of Enzyme Inhibitors in Foods*, edited by M. Freedman, pp. 1–17. New York: Plenum.
Laskowski, M. Jr & Kato, I. (1980). *Ann. Rev. Biochem.* **49**, 593–626.
Laskowski, M. Jr, Kato, I., Kohr, W. J., Park., S. J., Tashiro, M. & Whatley, H. E. (1987). *Cold Spring Harbor Symp. Quant. Biol.* **52**, 545–553.
Levitt, M. & Greer, J. (1977). *J. Mol. Biol.* **114**, 181–239.
Luzzati, P. V. (1952). *Acta Cryst.* **5**, 802–810.
Maeda, K., Kakabayashi, S. & Matsubara, H. (1983). *J. Biochem.* **94**, 865–870.
Minor, W. (1993). *XDISPLAYF* Program. Purdue University, West Lafayette, USA.
Mittl, P. R. E., Chene, P. & Grutter, M. G. (1998). *Acta Cryst.* D**54**, 86–89.
Müller, T., Oehlenschlager, F. & Buehner, M. (1995). *J. Mol. Biol.* **247**, 360–372.
Nakaguchi, T., Arakawa, T., Philo, J. S., Wen, J., Ishimoto, M. & Yamaguchi, H. (1997). *J. Biochem. (Tokyo)*, **121**, 350–354.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Oda, Y., Matsunaga, T., Fukuyama, K., Miyazaki, T. & Morimoto, T. (1997). *Biochemistry*, **36**, 13503–13511.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Poerio, E., Caporale, C., Carrano, L., Pucci, P. & Buonocore, V. (1991). *Eur. J. Biochem.* **199**, 595–600.
Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophysics J.* **5**, 909–933.
Richardson, M. (1991). *Methods Plant Biochem.* **5**, 259–305.
Shivraj, B. & Pattabhiraman, T. N. (1981). *Biochem. J.* **193**, 29–36.

Shivraj, B. & Pattabiraman, T. N. (1980). *Ind. J. Biochem. Biophys.* **17**, 181–185.
Srinivasan, A., Raman, A. & Singh, T. P. (1991). *J. Mol. Biol.* **22**, 1–2.
Strobl, S., Muhlhahn, P., Bernstein, R., Wiltscheck, R., Maskos, K., Wunderlich, M., Huber, R., Glockshuber, R. & Holak, T. A. (1995). *Biochemistry*, **34**, 8281–8293.
Weiss, M. S., Anderson, D. H., Raffioni, S., Brandshaw, R. A., Ortenzi, C., Luporini, P. & Eisenberg, D. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 10172–10176.
Wiegand, G., Epp, O. & Huber, R. (1995). *J. Mol. Biol.* **247**, 99–100.
Wilmanns, W. & Nilges, M. (1996). *Acta Cryst.* D**52**, 973–977.
Zemke, K. J., Muller-Fahrnow, A., Jany, K. D., Pal, G. P. & Saenger, W. (1991). *FEBS Lett.* **279**, 240–242.