Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the complex of trypsin with a highly potent synthetic inhibitor at 0.97 Å resolution

Manashi Sherawat,[a] Punit Kaur,[a] Markus Perbandt,[b] Christian Betzel,[b] William A. Slusarchyk,[c,d] Gregory S. Bisacchi,[c,e] Chiehying Chang,[c] Bruce L. Jacobson,[c,f] Howard M. Einspahr[c,g] and Tej P. Singh[a]*

[a]Department of Biophysics, All India Institute of Medical Sciences, New Delhi 110029, India, [b]Department of Biochemistry and Molecular Biology, University of Hamburg c/o DESY, Building 22a, 22603 Hamburg, Germany, [c]Bristol–Myers–Squibb Pharmaceutical Research Institute, PO Box 4000, Princeton, NJ 08543-4000, USA, [d]19 Richmond Drive, Skillman, NJ 08558, USA, [e]AstraZeneca, 35 Gatehouse Drive, Waltham, MA 02451, USA, [f]Saqqara Consultants, 2934 Avenida Pimentera, Carlsbad, CA 92009, USA, and [g]PO Box 6395, Lawrenceville, NJ 08648-0395, USA

Correspondence e-mail: tps_aiims@hotmail.com

Received 29 November 2006
Accepted 9 February 2007

PDB Reference: trypsin–inhibitor complex, 2ayw, r2aywsf.

The structure of the complex formed between bovine β-trypsin and the highly potent synthetic inhibitor 2-{3′-[5′-methoxy-2′-(*N*-*p*-diaminomethylphenyl)amido]-1′-pyrido}-5-(*N*-2″-*t*-butylethanol)amidobenzoic acid ($C_{28}H_{32}N_5O_6$) has been determined at 0.97 Å resolution. X-ray intensity data were collected to 0.97 Å under cryocooled conditions. The structure was refined anisotropically using *REFMAC*5 and *SHELX*-97 to R_{cryst} factors of 13.4 and 12.6% and R_{free} factors of 15.7 and 16.3%, respectively. Several regions of the main chain and side chains that have not been previously observed were clearly defined in the present structure. H atoms are indicated as significant peaks in an $|F_o - F_c|$ difference map, which accounts for an estimated 35% of all H atoms at the 2.5σ level. The C, N and O atoms are definitively differentiated in the electron-density maps. The amido part of the inhibitor occupies the specificity pocket and the remainder fills the remaining part of the ligand-binding cleft and interacts with the enzyme through an extensive network of hydrogen bonds. The inhibitor distorts the stereochemistry of the catalytic triad, Ser195–His57–Asp102, thereby blocking the proton-relay process of the active site by preventing the formation of the crucial hydrogen bond between His57 $N^{\delta1}$ and Asp102 $O^{\delta1}$.

1. Introduction

Serine proteinases are ubiquitous in living organisms: they are not only involved in physiological processes, but are also responsible for a number of diseases. The hyperproteolytic activity of these enzymes is often a serious problem, a fact that makes these proteins important chemotherapeutic targets in blood coagulation, fibrinolysis, kinin formation, complement activation, digestion, reproduction and phagocytosis (Neurath, 1984). The serine proteinases follow a common mechanism of action involving three amino-acid residues. These are serine, which functions as a proton donor, histidine, which behaves as a proton donor and acceptor in different steps of the reaction, and aspartic acid, the role of which is to assure the correct orientation of the histidine residue so that it can support nucleophilic attack by serine.

The functional significance of the catalytic triad and the oxyanion hole in catalysis have been fully established. In order to appreciate the nature of the proton-relay mechanism as well as the proton path and its precise location, the positions of H atoms must be determined accurately. Therefore, it is important to study the crystal structures of serine proteinases at atomic and subatomic resolutions (Kuhn *et al.*, 1998; Betzel *et*

Table 1
Overall statistics of data processing.

Values in parentheses correspond to the highest resolution shell.

PDB code	2ayw
X-ray source	BW7B
Temperature (K)	100
Wavelength (Å)	0.91
Space group	$P3_221$
Unit-cell parameters (Å)	$a = b = 68.55$, $c = 73.33$
V_M (Å^3 Da^{-1})	2.1
Solvent content (%)	39.0
Resolution range (Å)	20.00–0.97 (1.00–0.97)
Total No. of observations	225251
No. of unique reflections	116164
Completeness (%)	100 (99.9)
R_{sym}† (%)	6.5 (40.1)
Mean $I/\sigma(I)$	18.2 (3.1)

† $R_{sym} = \sum |I(h)_i - \langle I(h)\rangle| / \sum \langle I(h)_i\rangle$, where $I(h)_i$ is the intensity of the ith replicate of the reflection h and $\langle I(h)\rangle$ is the mean value and the summations are over all R_{merge} replicates, then over all h.

al., 2001). Atomic resolution is considered to be achieved if the crystal diffracts to better than 1.2 Å (Sheldrick, 1990). For the determination of solvent molecules, particularly beyond the first hydration shell, it is necessary to determine the protein structure at a very high resolution and especially at low temperature. Additionally, the occurrence of disorder and multiple conformations in the main and side chains may be associated with the function of the macromolecule and very high-resolution structures are necessary in order to discern these effects.

Trypsin is a prominent member of the serine proteinase family. Its mechanism of action has been studied in detail (Blow, 1976; Kraut, 1977; Huber & Bode, 1978; Kossiakoff & Spencer, 1981; Moult *et al.*, 1985), but the exact relationship between the catalytic residues and the sequence of events during the reaction have not been fully elucidated, especially for the catalytic serine. The accepted reaction mechanism proceeds *via* a tetrahedral intermediate. This is hard to observe owing to its transient nature (West *et al.*, 1990), but a number of model compounds have been reported to form covalently bound tetrahedral adducts at the serine (Chen *et al.*, 1995; Rypniewski *et al.*, 1993; Mac Sweeney *et al.*, 2000). Several authors have also reported enzyme–product complexes (Harel *et al.*, 1991), acyl-enzyme intermediates (Strynadka *et al.*, 1992) or even a trapped tetrahedral intermediate (Yennawar *et al.*, 1994). The substrate-binding sites of trypsin and related enzymes have been characterized using naturally occurring inhibitors (Bode & Huber, 1992; Ibrahim & Pattabhi, 2005) that interact with the enzyme in a substrate-like manner according to the standard mechanism (Laskowski & Kato, 1980) and the structures of their complexes with the enzymes have been determined in detail (Tsunogae *et al.*, 1986; Bode *et al.*, 1990; Turk *et al.*, 1991; Lee *et al.*, 1994, 1997; Li *et al.*, 1994; Scheidig *et al.*, 1997; Sandler *et al.*, 1998; Katz *et al.*, 1998, 2003; Krishnan *et al.*, 1998; Whitlow *et al.*, 1999; Koepke *et al.*, 2000; Toyota *et al.*, 2001; Dullweber *et al.*, 2001; Rauh *et al.*, 2004; Schmidt *et al.*, 2005). The crystal structures of trypsin complexes with autoproteolytic cleaved peptide fragments and covalently bound inhibitors at atomic resolution have revealed the precise protonation state of the catalytic residues (Schmidt *et al.*, 2003). In addition, a complex of trypsin with a heterochiral peptide has recently been reported in which the peptide binds through a calcium-binding loop of trypsin (Shamaladevi & Pattabhi, 2005). In order to further illuminate both the binding conformations of the inhibitors in the binding pocket of trypsin and the path of the protons in the mechanism of action of trypsin, we have crystallized trypsin in the presence of the synthetic inhibitor 2-{3′-[5′-methoxy-2′-(*N*-*p*-diaminomethylphenyl)amido]-1′-pyrido}-5-(*N*-2″-*t*-butylethanol)amidobenzoic acid (ONO). The crystals of the present complex diffract to 0.97 Å resolution and the atomic resolution structure shows the presence of two molecules of ONO and one molecule each of benzamidine, MES and glycerol. We find that the molecules of ONO are involved in extensive intermolecular interactions with trypsin. The positions of H atoms in the catalytic triad have been clearly observed. The structure also shows that the inhibitor is extremely tightly bound to the enzyme.

2. Experimental procedures

2.1. Crystallization, data collection and data processing

The samples of bovine β-trypsin used in these experiments were purchased from Sigma–Aldrich (USA). The protein was purified using a gel-filtration chromatography column (100 × 2 cm) packed with Sephadex G-75. The purified samples were dissolved at about 20 mg ml^{-1} in a buffer consisting of 0.9 *M* ammonium sulfate, 5 m*M* MES pH 6.0, 1 m*M* $CaCl_2$ and 2.5 m*M* benzamidine. The inhibitor dissolved in 175 m*M* 100% DMSO was added to this solution in a tenfold molar excess and incubated for 3–4 h at room temperature. Crystals for use in diffraction experiments were grown by the hanging-drop vapour-diffusion method. In these experiments, 15 µl droplets of the protein–inhibitor solution diluted to 10 mg ml^{-1} with a solution containing 0.9 *M* ammonium sulfate, 5 m*M* MES pH 6.0, 1 m*M* $CaCl_2$ and 2.5 m*M* benzamidine were used. 3 µl reservoir solution was also added to the 15 µl droplets. Crystals grew to full size after about a week. The crystal used for data collection was a prismatic chunk of approximately 0.5 × 0.3 × 0.2 mm. It was flash-cooled by dipping it into a cryoprotectant consisting of the reservoir solution brought to 20%(*v*/*v*) in glycerol, extracting it with a loop and then placing it in a N_2 cold stream at 100 K. The crystal diffracted to 0.97 Å under cryocooled conditions. The crystal form is trigonal, space group $P3_221$, with unit-cell parameters $a = b = 68.55$, $c = 73.33$ Å. All diffraction data were measured at a wavelength of 0.91 Å using the synchrotron beamline BW7B at EMBL, Hamburg, Germany. The data were collected in three sets with different resolution ranges: high, medium and low. A total of 225 251 reflections were observed and the merged data, consisting of 116 164 unique reflections to 0.97 Å resolution, corresponded to a completeness of 99.9%. The data-collection and processing details are summarized in Table 1.

Table 2
Summary of crystallographic refinement.

	*CCP*4, *REFMAC*	*SHELX*
Refinement		
Resolution limits (Å)	20.0–0.97	20.0–0.97
No. of reflections	116164	116164
Working set (99%)	114997	114997
Test set (1%)	1167	1167
R_{cryst}† (%)	13.4	12.9
R_{free}‡ (%)	15.9	16.3
Model		
Protein atoms	1629	1629
Amidobenzoic acid atoms (2 molecule)	78	78
Benzamidine atoms (1 molecule)	9	9
Glycerol atoms (1 molecule)	6	6
2-(*N*-Morpholino)ethanesulfonic acid atoms (1 molecule)	11	11
Calcium ion	1	1
Water molecules	449	449
Total No. of H atoms	1603	1603
Observed H atoms	560	560
R.m.s. deviations from ideal values		
Bond lengths (Å)	0.012	0.008
Bond angles§	1.660	0.017
Dihedral angles§	6.727	0.026
Average *B* factors (Å^2)		
Wilson plot	15.0	15.0
All atoms	13.9	12.3
Main-chain atoms	8.6	9.2
Water molecules	29.0	32.7
Residues in most preferred regions (%)	87.2	87.2
Residues in additionally allowed regions (%)	12.8	12.8

† $R_{cryst} = \sum |F_o(h) - F_c(h)| / \sum F_o(h)$, where $F_o(h)$ and $F_c(h)$ are the observed and calculated structure-factor amplitudes for reflection h. ‡ R_{free} was calculated against 1% of the complete data set excluded from refinement. § Given in degrees in *REFMAC* and Å in *SHELXL*.

2.2. Structure solution

The atomic resolution structure of the complex of trypsin with a designed inhibitor was solved using the molecular-replacement method as implemented in the program *AMoRe* (Navaza, 2001) from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The rotation function was calculated with data between 10 and 3 Å resolution and with a Patterson search radius of 20 Å. A solution was obtained using a search model defined by the coordinates of the hemiacetal complex between leupeptin and trypsin (Kurinov & Harrison, 1996; PDB code 1jrt).

2.3. Refinement

The model was initially refined against all the diffraction data using *REFMAC*5 (Murshudov *et al.*, 1997) from the *CCP*4*i* v.4.2 software suite (Collaborative Computational Project, Number 4, 1994) with isotropic temperature factors. At the end of the first stage of refinement, the *R* factor fell to 24.6% and difference Fourier $|F_o - F_c|$ and Fourier $|2F_o - F_c|$ maps were computed. These maps showed strong difference density near the binding site of trypsin and also at three intermolecular locations. Two molecules of ONO (ONO1 and ONO2) were built into the electron density at the binding site of the enzyme, while benzamidine, MES and glycerol molecules were placed into the difference densities present at the interstitial sites. In addition to these densities, a large spherical density was also observed at the 6σ level, which was interpreted as a calcium ion. The expanded model was further refined with isotropic atomic temperature factors including ligand models defined in standard geometry with full occupancies.

For comparison purposes, the structure was also refined in parallel against diffraction intensities rather than structure-factor amplitudes using *SHELXL* (Sheldrick & Schneider, 1997) by the conjugate-gradient algorithm, this time including anisotropic atomic temperature factors for non-H atoms. H atoms generated either in their riding positions or, in the case of mobile atoms, by the electron-density search algorithm were incorporated in the refinement with *SHELXL* (Sheldrick & Schneider, 1997). For both refinements, water molecules were modeled by application of *ARP* with default settings (Lamzin & Wilson, 1993). The significance level for defining new waters in the $|F_o - F_c|$ map was determined automatically. It initially started at 3σ and approached the 4σ level at the end of the refinement. Water molecules were removed from the model if their associated density was less than 0.5σ in the $|3F_o - F_c|$ map. The models were inspected against $|3F_o - F_c|$ and $|F_o - F_c|$ maps and were adjusted using the program *O* (Jones *et al.*, 1991) running on a Silicon Graphics O2 workstation.

At this stage, the H atoms at many aromatic residues and C$^\beta$ atoms were clearly seen in the $|F_o - F_c|$ maps. In the *REFMAC*5-generated map for the protein molecule alone, a total of 276 positive peaks above 3σ in the $|F_o - F_c|$ maps were

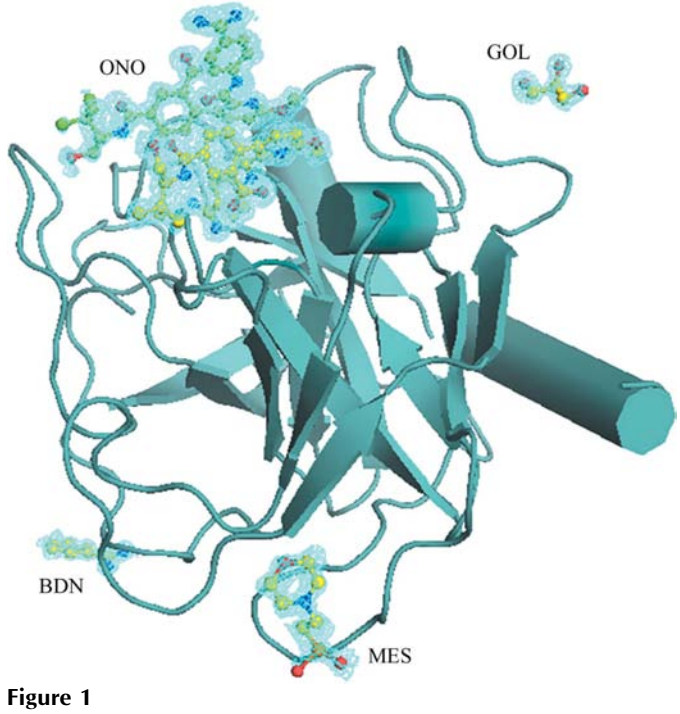

Figure 1
The $|F_o - F_c|$ map contoured at 2.5σ showing the electron densities for two molecules of the inhibitor 2-{3′-[5′-methoxy-2′-(*N*-*p*-diaminomethylphenyl)amido]-1′-pyrido}-5-(*N*-2″-*t*-butylethanol)amidobenzoic acid (ONO), benzamidine (BDN), glycerol (GOL) and 2-(*N*-morpholino)-ethanesulfonic acid (MES). ONO1 is shown in yellow, whereas ONO2 is indicated in green. The figure was drawn with *PyMOL* (DeLano, 2002).

located at idealized H-atom positions. These peaks, representing approximately 17% of the 1603 H atoms expected in the protein molecule, are clustered predominantly in either hydrophobic or good hydrogen-bond-forming regions. Peaks were also observed for most C^{β} H atoms in 146 out of 159 residues excluding Gly and Ala residues. Peaks at the 2σ level accounted for an estimated 35% of all H atoms in the protein molecule. Similar results were obtained from the *SHELXL*-generated map.

The inclusion of H atoms gave enhanced details of other H atoms in the regions of interest that had not been modeled or were detectable only below the 3σ threshold in previous electron-density maps. These coordinates were used for anisotropic refinement in both *SHELXL* and *REFMAC*5. At the end of the anisotropic refinement by *REFMAC*5, the $R_{\rm cryst}$ and $R_{\rm free}$ factors converged to 13.4 and 15.7%, respectively. Similarly, in the refinement with *SHELXL* the $R_{\rm cryst}$ and $R_{\rm free}$ factors converged to 12.9 and 16.3%, respectively (Table 2). At the end of the anisotropic refinement by both methods, the quality of the electron density for the H atoms produced by both refinements improved remarkably. In fact, in the final analysis there is little to distinguish between the results of the two refinements (Table 2). These results clearly indicated that the anisotropic refinements with atomic resolution data sets using both *REFMAC*5 and *SHELXL* produce similar qualities of results. For the sake of brevity, the subsequent structure analysis will be based on the model refined with *REFMAC*5. The atomic coordinates of the complex have been deposited in the Protein Data Bank with accession number 2ayw (for the *REFMAC*5 results).

3. Results and discussion

3.1. Quality of the structure

The protein model is of very high quality and includes all protein non-H atoms. Even the segments in alternate conformations were modeled without ambiguity. The complete chain is visible in contiguous $|3F_{\rm o} - F_{\rm c}|$ electron density above the 1.5σ level including fragments that assume two conformations. Moreover, almost all main-chain atoms are seen in $|3F_{\rm o} - F_{\rm c}|$ density at or above the 4σ level. Water molecules generally have very good $|3F_{\rm o} - F_{\rm c}|$ density. The two molecules of ONO and one molecule each of benzamidine, MES and glycerol could also be modeled in the electron densities with confidence (Fig. 1). As mentioned previously, one calcium ion was observed with strong electron density. The final refinement performed using the full-matrix algorithm in which all the positional parameters were varied provides the best estimate of the quality of the model. The standard uncertainties (s.u.s) characterizing the C—C bond distances (a value often used as a global indicator of the quality of small-molecule organic crystal structures) are as low as 0.006 Å and have a mean value of 0.011 Å. This is a remarkable result considering that s.u.s are calculated for all C—C bonds including those for side chains, where lack of specific interactions and high mobilities result in some cases in poor definition and consequently in high standard errors in the atomic coordinates. Most notably, individual heavy atoms (C, N and O atom) could be differentiated (Fig. 2). As seen from Fig. 2, the electron density in the red colour at the 4σ level clearly distinguishes the N atoms from C atoms as the density at N atoms is larger than that of C atoms (Deacon *et al.*, 1997). There are four salt bridges in the structure (Fig. 2). In Fig. 2, the electron density for C, N and O atoms can be distinguished based on the volume of the density contoured at 4σ. The carboxylate group in each of these salt bridges is expected to have a charge localized on $O^{\delta2}$ of the Asp or Asn residue. In Fig. 2(*a*), for example, $O^{\delta2}$ of Asp194 should be singly bonded to the C^{γ} and hydrogen bonded to the N atom of Ile16. The electron density for the C^{γ}—$O^{\delta2}$ bond is broken at the 4σ level, while that for the $O^{\delta1}$—C^{γ} bond is

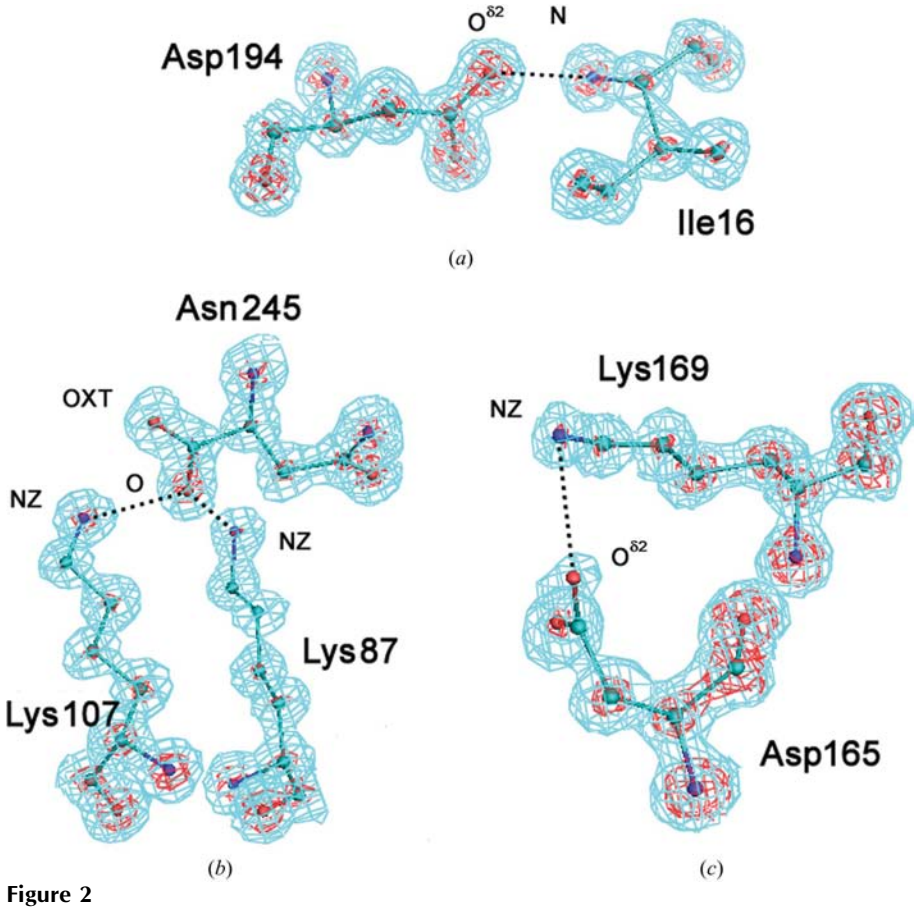

Figure 2
Interactions between (*a*) Asp194 and Ile16, (*b*) C-terminal Asn245 and Lys107 and Lys87 and (*c*) Asp165 and Lys169. The $|2F_{\rm o} - F_{\rm c}|$ electron-density map is contoured at 1.0σ (blue) and 4.0σ (red). The presence of the density towards one of the two carboxyl O atoms clearly shows the presence of a double bond in the group.

continuous, having comparable density to the carbonyl (C═O, backbone atoms) atoms of Asp194 and Ile16. Similar types of density patterns are seen for the other salt bridges.

3.2. Alternate conformation

The current structure has revealed five double-conformation residues (Fig. 3). All the residues with double conformations are found on the surface of the protein and are distributed in three clusters, with Ser37 and Lys60 at one site, and Ser113 and Arg117 at another site, while Asp165 is clustered with Ser166 and Ser167 (not shown) at the third site. The room-temperature crystal structure of the protein at 1.8 Å resolution did not reveal double conformations of these residues (Kurinov & Harrison, 1996), which is consistent with the observations of Esposito *et al.* (2000). By disorder in this context we do not mean lack of organized structure, but rather clearly visible alternate conformations.

3.3. Solvent structure

449 water molecules, as O atoms, were clearly determined. Water molecules were retained for which the electron density in the subsequent $|3F_o - F_c|$ maps persisted after the refinement cycles and which fulfilled the following criteria: (i) they were within 3.4 Å of the enzyme O or N atom or bound to neighboring water molecule with a good hydrogen-bonding geometry with *B* factors less than 45 Å^2 and (ii) their real-space correlation coefficients were above 65%. The average *B* factor for all the water molecules thus picked up is 29.0 Å^2.

3.4. C—H···O hydrogen bonds

An analysis of medium- and high-resolution structures revealed that C—H···O interactions are ubiquitous in macromolecular structures and contribute significantly to the free-energy difference between folded and unfolded states of proteins (Derewenda *et al.*, 1995). The sum of the C—H bond lengths and the van der Waals radii for H and O atoms is 3.7 Å. The aforementioned analysis showed a broad maximum at 2.65 Å in the distribution of C···O distances relative to that of C···C distances and a peak at a C···O distance shorter than 3.7 Å, suggestive of the effect of attractive interactions (Derewenda *et al.*, 1995). An analysis of the present structure clearly reveals the existence of many C—H···O distances less than 3.5 Å. As an example, it is clearly seen in Fig. 4 that His40 is involved in several C—H···O-type interactions. The His $C^{\varepsilon 2}$—H···O (carbonyl) Cys42 interaction is particularly important as it is present in all serine proteinases (Derewenda *et al.*, 1994).

3.5. Calcium-binding site

Serine proteinase requires calcium ions to stabilize the folding of the polypeptide chain that results in the optimal activity of trypsin. Thus, the level of calcium ions determines the state of proteolysis. Intracellular calcium levels are too low to produce active enzymes. Extracellular levels are higher and the enzyme is activated on secretion from the cell (Voordouw *et al.*, 1976). According to the high-resolution electron density, one calcium ion is clearly identified in the structure at an expected position as reported previously in β-trypsin structures (Johnson *et al.*, 1999; Deepthi *et al.*, 2001). It appears as a large spherical peak that is involved in coordination with two carboxyl O atoms, one each from Glu70 and Glu80, two carbonyl O atoms from Asn72 and Val75 and two water molecules. The position of Ca^{2+} is well defined with sixfold coordination and is situated in a highly polar region of the molecule.

3.6. Active-site geometry and binding of the inhibitor

One molecule of ONO (ONO1) occupies the specificity pocket and interacts extensively with the active site as well as

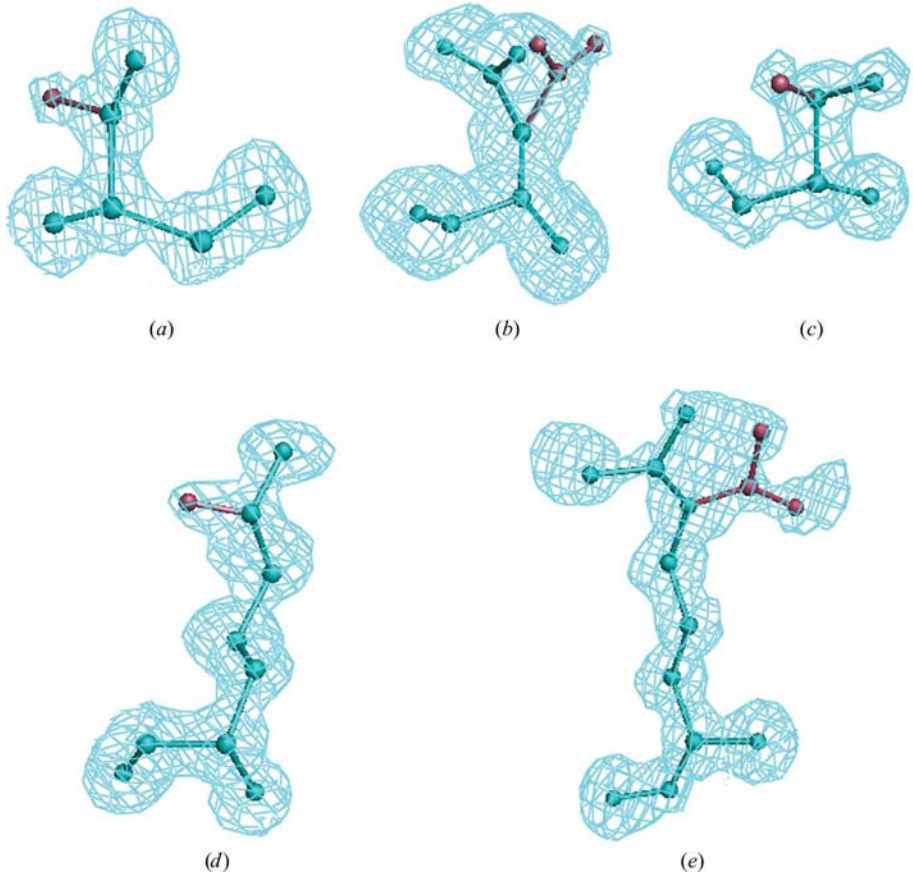

Figure 3
$|2F_o - F_c|$ electron-density map contoured at 1.2σ showing disordered residues with more than one conformation for the side chains: (*a*) Ser37, (*b*) Asp165, (*c*) Ser113, (*d*) Lys60, (*e*) Arg117.

Table 3
Interactions of the inhibitor with protein and water molecules.

INH1	Protein	Distance (Å)
N7	Ser214 O	3.31
	Ser195 O^{γ}	3.36
N9	Ser190 O	3.17
	Gly219 O	2.88
	Asp189 $O^{\delta2}$	2.83
	OW426	3.02
N10	Asp189 $O^{\delta1}$	2.94
	Ser190 O^{γ}	2.99
	Ser190 O	3.25
	OW192	2.95
O14	Gln192 $N^{\varepsilon2}$	2.94
O28	His57 $N^{\varepsilon2}$	2.66
	Ser195 O^{γ}	3.02
	OW194	2.98
	OW123	3.46
O29	Ser195 O^{γ}	2.72
	Gly193 N	3.03
	OW123	2.65
N31	OW205	3.26
	OW313	3.39
O32	Gln192 $N^{\varepsilon2}$	3.55
	OW259	2.73
O44	His57 $N^{\delta1}$	3.50
O67	OW429	2.51

INH2	Protein	Distance (Å)
O14	Gln192 $N^{\varepsilon2}$	2.88
N16	OW173	3.30
O28	OW173	2.76
N31	Gln192 $O^{\varepsilon1}$	3.09
O67	Gly148 O	2.46
	Asn143 $N^{\delta2}$	3.35

INH1	INH2	Distance (Å)
N16	O44	3.65

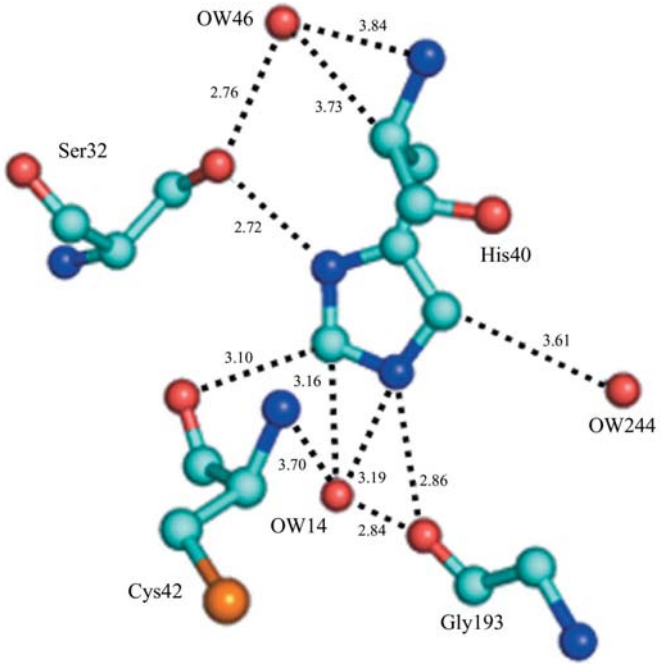

Figure 4
Example of C—H···O interactions involving His40 in the structure.

with other residues in the vicinity of the active site (Table 3; Figs. 1 and 5), while a large part of the second inhibitor molecule ONO2 is exposed to the surface (Fig. 6). The amino group of the diaminomethylphenylamido moiety of ONO1 sits in the specificity pocket and forms several hydrogen bonds to the side chains of Asp189 and Ser190, anchoring it firmly at the specificity site. In this orientation, the carboxyl group of the amidobenzoic acid moiety interacts with His57 $N^{\varepsilon2}$ through a strong hydrogen bond. The carboxyl group O atoms also form hydrogen bonds with Ser195 O^{γ}, Gly193 N and two other water molecules OW123 and OW194. It is important to note here that His57 $N^{\delta1}$ is deprotonated and $N^{\varepsilon2}$ is protonated. As indicated by the electron density (Fig. 7), Asp102 forms a clear hydrogen bond not to His57 $N^{\delta1}$, which is the commonly observed interaction and which is a necessary requirement for the catalytic process, but instead to His57 NH. Furthermore, His 57 $N^{\varepsilon2}$, being protonated, forms a hydrogen

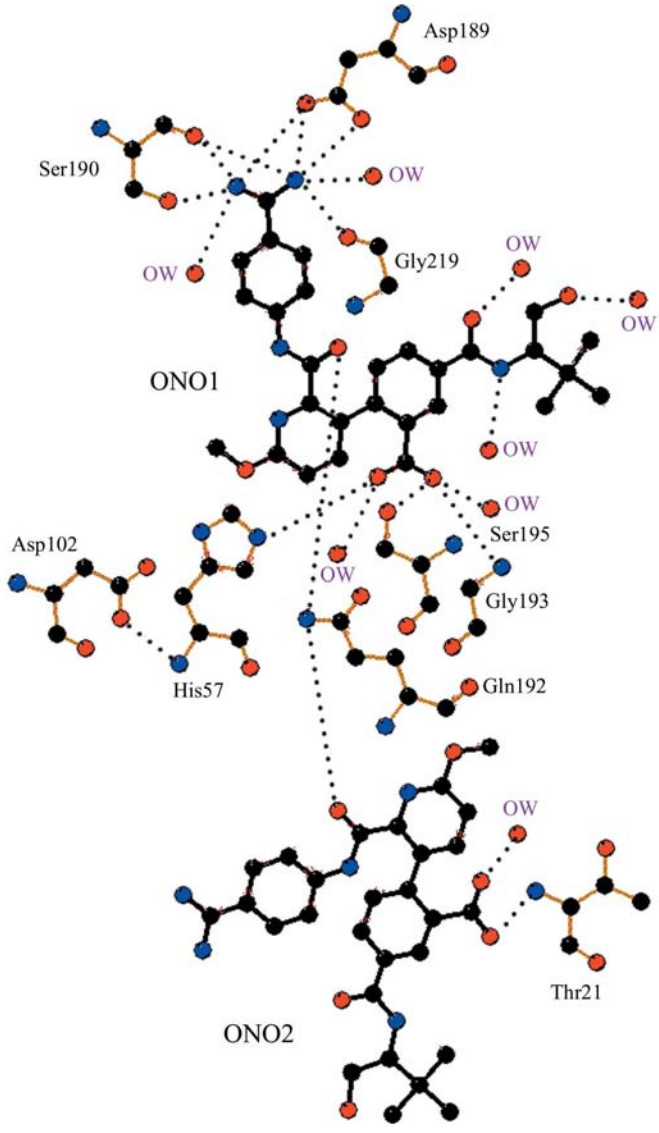

Figure 5
Schematic *LIGPLOT* representation (Wallace *et al.*, 1995) showing the interactions between trypsin, ONO1, ONO2 and a symmetry-related trypsin molecule.

bond to the ONO carboxylate instead of accepting a hydrogen bond from Ser O^{γ}, which is also a necessary requirement for the catalytic process. The electron density also shows that the Asp102 is relatively poorly aligned geometrically with His57. It appears that the interactions of the inhibitor with His57 and Ser195 perturb the stereochemistry of the catalytic triad. The number of interactions observed between trypsin and the current inhibitor is larger than found in other complexes of trypsin with synthetic and natural Bowman–Birk and Kunitz-type inhibitors (Tsunogae *et al.*, 1986; Bode *et al.*, 1990; Turk *et al.*, 1991; Lee *et al.*, 1994, 1997; Li *et al.*, 1994; Scheidig *et al.*, 1997; Sandler *et al.*, 1998; Katz *et al.*, 1998, 2003; Krishnan *et al.*, 1998; Whitlow *et al.*, 1999; Koepke *et al.*, 2000; Toyota *et al.*, 2001; Dullweber *et al.*, 2001; Rauh *et al.*, 2004; Schmidt *et al.*, 2005), indicating that its potency is likely to be higher than other known inhibitors.

The second molecule of the inhibitor ONO2 interacts with the first molecule through Gln192 $N^{\varepsilon 2}$. It also interacts with a symmetry-related protein molecule (Fig. 5). Both inhibitor molecules also form several hydrogen bonds to solvent water molecules (Table 3).

The molecule of benzamidine is located at the interstitial site and its primary role in the present structure appears to be to stabilize the packing of the molecules in the crystal through hydrophobic interactions.

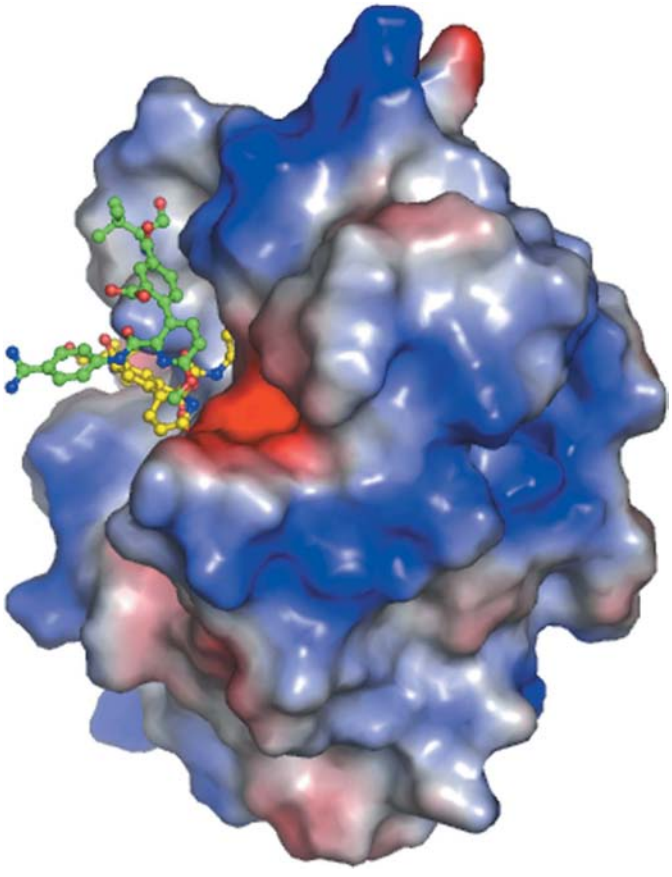

Figure 6
Representation of the binding cavity. The inhibitor molecule ONO1 is almost completely buried in the binding pocket, whereas ONO2 is on the surface of the enzyme molecule, interlinking the two symmetry-related trypsin molecules.

4. Conclusions

This atomic resolution structure allows the identification of the locations of the H atoms of the active-site residues and the definition of the precise nature of the interactions between the enzyme and the inhibitor molecules. An explanation is also provided for the differences in the affinities of the inhibitors ONO and benzamidine when they compete to interact with the protein; ONO binds to trypsin at the active site, whereas benzamidine is relegated to a nonspecific interstitial site. It has been observed for the first time that the inhibitor causes protonation at $N^{\varepsilon 2}$ and deprotonation at $N^{\delta 1}$ of His57, while Ser O^{γ} remains protonated owing to the influence of the inhibitor. This process of eliminating the bonds interrupts the normal hydrogen bonding between His57 $N^{\delta 1}$ and the carboxylic group O atom $O^{\delta 1}$ of Asp102 and between His57 $N^{\varepsilon 2}$ and Ser195 O^{γ}. As a result, Asp102 forms a hydrogen bond to His57 NH. This is an important observation and indicates the potency of the inhibitor in deactivating the enzyme completely. To repeat, the inhibitor not only binds to the trypsin firmly at the binding site, but also inactivates the enzyme completely by altering its protonation–deprotonation sites.

The present high-resolution refinement of the structure has allowed better definition of the molecular interactions involving H atoms and has provided direct visualization of these atoms in the electron-density maps. This improvement in quality makes possible future studies to analyze nonclassical interactions such as, for example, C—H···O interactions.

The structure determination also shows that only one molecule of the inhibitor ONO1 is bound to trypsin at the active site. The second molecule is associated with the first molecule through the protein chain, occupies the surface at the outer side of the active site and interacts with the symmetry-related protein molecules. Benzamidine, another inhibitor of trypsin, is located at the interstitial site, away from

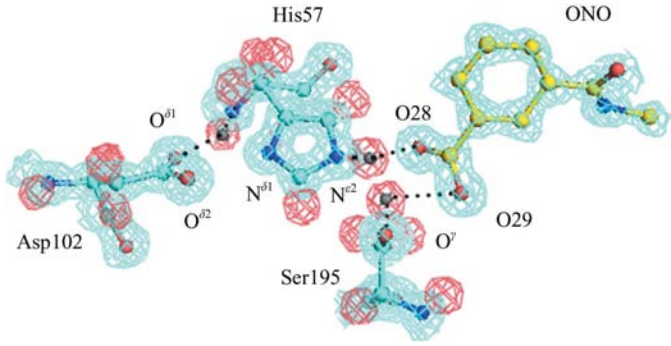

Figure 7
Interactions of the residues of the catalytic triad with the carbonyl group of the inhibitor ONO1, showing model H atoms superimposed on the $|F_o - F_c|$ map (model H phases) at 2σ (blue). The H-atom densities (red, 2σ in difference map) are clearly seen between the carbonyl group O atom O29 and Ser195 O^{γ}, O atom O28 and His57 $N^{\varepsilon 2}$ and His57 N and Asp102 $O^{\delta 2}$.

the active site. Overall, on the basis of the examination of the intermolecular interactions with the inhibitor and the enzyme, it may be possible that ONO is the most effective inhibitor of trypsin observed to date, It appears to fit ideally into the active site while interacting with His57 and Ser195 simultaneously. The strong hydrophobic base of the inhibitor also utilizes the hydrophobic surface of the protein effectively. The structure of the complex provides a solid structural basis for understanding the mechanism of inhibition of trypsin and may open new avenues for the design of more effective inhibitors of trypsin and other serine proteinases.

This project was supported by a joint grant from the Deutscher Akademischer Austauschdienst (DAAD), Germany and by a grant from the Department of Science and Technology (DST), New Delhi in terms of the Project-Based Personnel Exchange Programme 2000.

References

Betzel, C., Gourinath, S., Kumar, P., Kaur, P., Perbandt, M., Eschenburg, S. & Singh, T. P. (2001). *Biochemistry*, **40**, 3080–3088.

Blow, D. M. (1976). *Acc. Chem. Res.* **9**, 145–152.

Bode, W. & Huber, R. (1992). *Eur. J. Biochem.* **204**, 433–435.

Bode, W., Turk, D. & Sturzebecher, J. (1990). *Eur. J. Biochem.* **193**, 175–182.

Chen, Z., Li, Y., Mulichak, A. M., Lewis, S. D. & Shafer, J. A. (1995). *Arch. Biochem. Biophys.* **322**, 198–203.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Deacon, A., Gleichmann, T., Kalle, A. J., Price, H., Raftery, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997). *J. Chem. Soc. Faraday Trans.* **93**, 4305–4312.

Deepthi, S., Johnson, A. & Pattabhi, V. (2001). *Acta Cryst.* D**57**, 1506–1512.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.

Derewenda, Z. S., Derewenda, U. & Kobos, P. M. (1994). *J. Mol. Biol.* **241**, 83–93.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). *J. Mol. Biol.* **252**, 248–262.

Dullweber, F., Stubbs, M. T., Musil, D., Sturzebecher, J. & Klebe, G. (2001). *J. Mol. Biol.* **313**, 593–614.

Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A. A. & Mazzarella, L. (2000). *J. Mol. Biol.* **297**, 713–732.

Harel, M., Su, C. T., Frolow, F., Ashani, Y., Silman, I. & Sussman, J. L. (1991). *J. Mol. Biol.* **221**, 909–918.

Huber, R. & Bode, W. (1978). *Acc. Chem. Res.* **11**, 114–122.

Ibrahim, B. S. & Pattabhi, V. (2005). *J. Mol. Biol.* **348**, 1191–1198.

Johnson, A., Gautham, N. & Pattabhi, V. (1999). *Biochim Biophys Acta*, **1435**, 7–21.

Jones, T. A., Zou, J.-Y., Cowan, S. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Katz, B. A., Clark, J. M., Finer-Moore, J. S., Jenkins, T. E., Johnson, C. R., Ross, M. J., Luong, C., Moore, W. R. & Stroud, R. M. (1998). *Nature (London)*, **391**, 608–612.

Katz, B. A., Elrod, K., Verner, E., Mackman, R. L., Luong, C., Shrader, W. D., Sendzik, M., Spencer, J. R., Sprengeler, P. A., Kolesnikov, A., Tai, V. W.-F., Hui, H. C., Breitenbucher, J. G., Allen, D. & Janc, J. W. (2003). *J. Mol. Biol.* **329**, 93–120.

Koepke, J., Ermler, U., Warkentin, E., Wenzl, G. & Flecker, P. (2000). *J. Mol. Biol.* **298**, 477–491.

Kossiakoff, A. A. & Spencer, A. (1981). *Biochemistry*, **20**, 6462–6474.

Kraut, J. (1977). *Annu. Rev. Biochem.* **46**, 331–358.

Krishnan, R., Zhang, E., Hakansson, K., Arni, R. K., Tulinsky, A., Lim-Wilby, M. S., Levy, O. E., Semple, J. E. & Brunck, T. K. (1998). *Biochemistry*, **37**, 12094–12103.

Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.

Kurinov, I. V. & Harrison, R. W. (1996). *Protein Sci.* **5**, 752–758.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

Laskowski, M. & Kato, I. (1980). *Annu. Rev. Biochem.* **49**, 593–626.

Lee, A. Y., Smitka, T. A., Bonjouklian, R. & Clardy, J. (1994). *Chem. Biol.* **1**, 113–117.

Lee, S. L., Alexander, R. S., Smallwood, A., Trievel, R., Mersinger, L., Weber, P. C. & Kettner, C. (1997). *Biochemistry*, **36**, 13180–13186.

Li, Y., Huang, Q., Zhang, S., Liu, S., Chi, C. & Tang, Y. (1994). *J. Biochem. (Tokyo)*, **116**, 18–25.

Mac Sweeney, A., Birrane, G., Walsh, M. A., O'Connell, T., Malthouse, J. P. & Higgins, T. M. (2000). *Acta Cryst.* D**56**, 280–286.

Moult, J., Sussman, F. & James, M. N. (1985). *J. Mol. Biol.* **182**, 555–566.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.

Neurath, H. (1984). *Science*, **224**, 350–357.

Rauh, D., Klebe, G. & Stubbs, M. T. (2004). *J. Mol. Biol.* **335**, 1325–1341.

Rypniewski, W. R., Hastrup, S., Betzel, C., Dauter, M., Dauter, Z., Papendorf, G., Branner, S. & Wilson, K. (1993). *Protein Eng.* **6**, 341–348.

Sandler, B., Murakami, M. & Clardy, J. (1998). *J. Am. Chem. Soc.* **120**, 595–596.

Scheidig, A. J., Hynes, T. R., Pelletier, L. A., Wells, J. A. & Kossiakoff, A. A. (1997). *Protein Sci.* **6**, 1806–1824.

Schmidt, A. E., Chand, H. S., Cascio, D., Kisiel, W. & Bajaj, S. P. (2005). *J. Biol. Chem.* **280**, 27832–27838.

Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003). *J. Biol. Chem.* **278**, 43357–43362.

Shamaladevi, N. & Pattabhi, V. (2005). *J. Biomol. Struct. Dyn.* **22**, 635–642.

Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.

Strynadka, N. C., Adachi, H., Jensen, S. E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K. & James, M. N. (1992). *Nature (London)*, **359**, 700–705.

Toyota, E., Ng, K. K., Sekizaki, H., Itoh, K., Tanizawa, K. & James, M. N. (2001). *J. Mol. Biol.* **305**, 471–479.

Tsunogae, Y., Tanaka, I., Yamane, T., Kikkawa, J., Ashida, T., Ishikawa, C., Watanabe, K., Nakamura, S. & Takahashi, K. (1986). *J. Biochem. (Tokyo)*, **100**, 1637–1646.

Turk, D., Sturzebecher, J. & Bode, W. (1991). *FEBS Lett.* **287**, 133–138.

Voordouw, G., Milo, C. & Roche, R. S. (1976). *Biochemistry*, **15**, 3716–3724.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *Protein Eng.* **8**, 127–134.

West, J. B., Hennen, W. J., Lalonde, J. L., Bibbs, J. A., Zhong, Z., Meyer, E. F. Jr & Wong, J. H. (1990). *J. Am. Chem. Soc.* **112**, 5313–5320.

Whitlow, M., Arnaiz, D. O., Buckman, B. O., Davey, D. D., Griedel, B., Guilford, W. J., Koovakkat, S. K., Liang, A., Mohan, R., Phillips, G. B., Seto, M., Shaw, K. J., Xu, W., Zhao, Z., Light, D. R. & Morrissey, M. M. (1999). *Acta Cryst.* D**55**, 1395–1404.

Yennawar, N. H., Yennawar, H. P. & Farber, G. K. (1994). *Biochemistry*, **33**, 7326–7336.