Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Atomic resolution structure of squash trypsin inhibitor: unexpected metal coordination

Ram Thaimattam,[a] Ewa Tykarska,[a] Andrzej Bierzynski,[b] George M. Sheldrick[c] and Mariusz Jaskolski[a,d]*

[a]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland, [b]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland, [c]Institute of Inorganic Chemistry, University of Goettingen, Germany, and [d]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

Correspondence e-mail: mariuszj@amu.edu.pl

Received 18 March 2002
Accepted 3 July 2002

PDB Reference: squash trypsin inhibitor, 1lu0, r1lu0sf.

CMTI-I, a small-protein trypsin inhibitor, has been crystallized as a 4:1 protein–zinc complex. The metal is coordinated in a symmetric tetrahedral fashion by glutamate/glutamic acid side chains. The structure was solved by direct methods in the absence of prior knowledge of the special position metal centre and refined with anisotropic displacement parameters using diffraction data extending to 1.03 Å. In the final calculations, the main-chain atoms of low B_{eq} values were refined without restraint control. The two molecules in the asymmetric unit have a conformation that is very similar to that reported earlier for CMTI-I in complex with trypsin, despite the Met8Leu mutation of the present variant. The only significant differences are in the enzyme-binding epitope (including the Arg5 residue) and in a higher mobility loop around Glu24. The present crystal structure contains organic solvent molecules (glycerol, MPD) that interact with the inhibitor molecules in an area that is at the enzyme–inhibitor interface in the CMTI-I–trypsin complex. A perfectly ordered residue (Ala18) has an unusual Ramachandran conformation as a result of geometrical strain introduced by the three disulfide bridges that clamp the protein fold. The results confirm deficiencies of some stereochemical restraints, such as peptide planarity or the N—C$^{\alpha}$—C angle, and suggest a link between their violations and hydrogen bonding.

1. Introduction

In recent years, protein crystallography has witnessed an accelerating increase in the number of atomic resolution structures (Dauter *et al.*, 1997; Longhi *et al.*, 1998). With advancements in crystallization methods, diffraction techniques (synchrotron radiation, faster and more sensitive detectors) and computing capabilities (faster computers and algorithms), it is not unusual today to reach a level in a macromolecular refinement that not long ago was reserved for small-molecule crystallography (Walsh *et al.*, 1998; Ridder *et al.*, 1999; Esposito, Vitagliano, Sica *et al.*, 2000; Esposito, Vitagliano, Zagari *et al.*, 2000; Jelsch *et al.*, 2000; Addlagatta *et al.*, 2001). Macromolecular structures at atomic and subatomic resolution provide an unsurpassed wealth of information that helps to detect details that would otherwise be overlooked or to rectify not fully adequate approximations based on lower resolution results. High resolution, particularly in combination with low temperature, is necessary to reveal the organization of solvent beyond the first hydration shell, to properly model multiple conformations or to capture subtle structural details, such as protonation states, that are often important for

adequate understanding of the mechanism of action of macromolecules. While experience in carrying out atomic resolution refinements is accumulating, it is still not at a level where full automation would be possible or indeed advisable. It seems that because of the unexpected that is revealed by almost every atomic resolution structure, each case should be handled individually and that only in combination with the investigator's ingenuity can the message be fully captured and comprehended. Still, there is an obvious need for establishing refinement standards that would provide protocol guidelines at different levels of atomic resolution or, for instance, at different packing coefficients. Also, it seems to be a generally accepted conclusion that the useful stereochemical standards used for restraining lower resolution structure refinements may not always be exact and that only information from very high resolution protein refinements can improve the situation (Sandalova *et al.*, 1999; Esposito, Vitagliano, Sica *et al.*, 2000; Esposito, Vitagliano, Zagari *et al.*, 2000; Addlagatta *et al.*, 2001). There is a need for accurate high-resolution structures of both small and large proteins. The former allow meticulous testing of various refinement variants and can even be refined by rigorous full-matrix least-squares treatment. The latter, on the other hand, are a test ground for the method and its limits; they also provide vast amounts of geometrical data for adequate statistical analysis. We have recently reported the ultrahigh-resolution (0.86 Å) structure of bovine pancreatic trypsin inhibitor, BPTI (Addlagatta *et al.*, 2001). Here, we report the crystal structure of a mutant (Met8Leu) of another trypsin inhibitor, from squash (*Cucurbita maxima*), CMTI-I. The resolution is lower, 1.03 Å, but still high enough to place the structure within the highest resolution category available in the Protein Data Bank (Berman *et al.*, 2000). The CMTI-I peptide is rather small at only 29 amino acids long, exactly half the size of BPTI. However, the asymmetric unit of the present crystal form contains two copies of the protein. In addition, the structure reveals an unexpected (and quite unusual for this type of inhibitor) coordination of zinc cations.

CMTI-I and other members of this family found in the seeds of *Cucurbitaceae* squash plants are the smallest known protein inhibitors of serine proteases (Wilusz *et al.*, 1983; Wieczorek *et al.*, 1985; Otlewski *et al.*, 1987; Polanowski *et al.*, 1987, 1988). Because of their short length, the squash-seed inhibitors lack regular secondary-structure elements and their compact fold is maintained by intramolecular disulfide bridges. In CMTI-I there are three such bridges connecting the N-terminal and C-terminal segments of the polypeptide chain in a consecutive order. CMTI-I (*C. maxima* trypsin inhibitor) binds bovine β-trypsin with a very high association constant of $3.2 \times 10^{11}\ M^{-1}$ (Wieczorek *et al.*, 1985). The structure of wild-type CMTI-I was first established by X-ray crystallography in a complex with bovine β-trypsin at 2.0 Å resolution (Bode *et al.*, 1989; PDB code 1ppe) and later with salmon trypsin at 1.8 Å resolution (Helland *et al.*, 1999; PDB code 2sta). Complexes of salmon and bovine trypsin with a similar plant inhibitor, CPTI, were also reported at 1.8 Å (2stb) and 1.5 Å (2btc), respectively (Helland *et al.*, 1999). (Chemically, CPTI is an Arg5Lys, Val21Ile mutant of CMTI.) As in other small-protein inhibitors of serine proteases such as BPTI, the enzyme-binding epitope is located on an exposed loop allowing efficient docking in the enzyme's binding cleft. In CMTI-I the binding loop covers residues 1–9, including the reactive-site peptide Arg5–Ile6. In the inhibitor–enzyme complexes this arginine residue penetrates deep into the S1 pocket of the enzyme, where it forms a salt bridge with a carboxylic acid. The crystal structures of the complexes also established that contacts with the enzyme involve residues Arg1–Leu7 and Glu9 of the binding loop but not Met8, the Met→Leu mutation site of the present variant. The structure of CMTI-I has also been studied by NMR spectroscopy by Holak *et al.* (1989) and Nilges *et al.* (1991). Recently, Zhukov *et al.* (2000) have used the NMR technique to study the Met8Leu mutant (PDB code 1bxj). Here, we report the crystal structure of the Met8Leu mutant of CMTI-I determined at 1 Å resolution using synchrotron radiation and cryogenic techniques.

2. Materials and methods

2.1. The protein

Recombinant CMTI-I with Met8 replaced with Leu was obtained by expression of a synthetic gene in *Escherichia coli*. It was isolated and purified to homogeneity by reversed-phase HPLC as described previously (Bolewska *et al.*, 1995). The association constants of the mutant inhibitor with β-trypsin and cathepsin G are virtually the same as those of native CMTI-I (Bolewska *et al.*, 1995).

2.2. Crystallization

Initial crystallization conditions were determined using Hampton Research Crystal Screen (Jancarik & Kim, 1991). During optimization of the crystallization conditions it was observed that the results critically depended on the concentration of zinc cations. The best crystals were obtained when the precipitating solution contained 5 m*M* $ZnSO_4$, 50 m*M* cacodylate buffer pH 6.5, 14% MPD and 8% PEG 8000. The crystals were grown using the vapour-diffusion method.

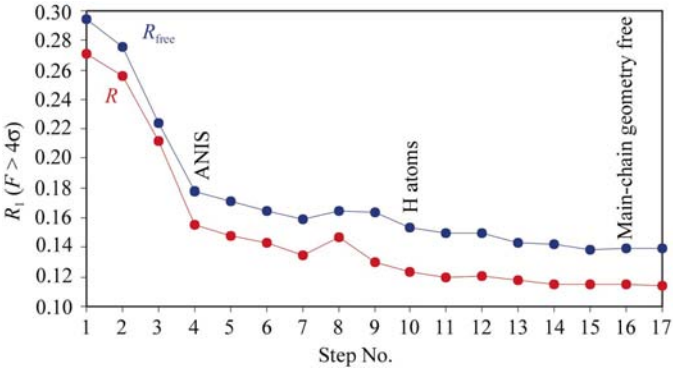

Figure 1
Plot of the R (lower line) and R_{free} (upper line) factors *versus* refinement round. The values plotted are for reflections with $F > 4\sigma(F)$, but the refinement included all reflections.

Hanging drops were prepared by mixing 1 µl of the precipitating buffer with equal amounts of the protein solution at a concentration of 12 mg ml^{-1}. Crystals appeared after 1–2 d and reached final dimensions of 0.40 × 0.15 × 0.10 mm after an additional 8 d.

2.3. Data collection and processing

X-ray diffraction data were measured at the EMBL BW7A beamline at the DORIS ring of the DESY synchrotron operated at λ = 0.8919 Å. The crystal was flash-frozen to 100 K after immersion in a cryoprotectant solution containing the mother liquor supplemented with glycerol (Teng, 1990). Using a MAR CCD detector, the data were collected in three runs. The maximum resolution, oscillation range and number of images were as follows: run 1, 1.03 Å, 0.5°, 200; run 2, 1.9 Å, 1.0°, 90; run 3, 2.9 Å, 2.0°, 45. The images were integrated in *DENZO* and the intensities were merged in the *SCALEPACK* program from the *HKL* package (Otwinowski & Minor, 1997). The isotropic crystal mosaicity parameter was refined to be 0.35°. Over 250 000 measurements were combined to produce 27 045 unique reflections, leading to high redundancy and nearly 100% completeness (Table 1). The overall $R_{\rm int}$ is 0.047 and is 0.151 in the last resolution shell. The crystal belongs to space group $P2_12_12$, with unit-cell parameters a = 47.28, b = 58.63, c = 19.32 Å. Calculation of the Matthews volume (Matthews, 1968) indicated, as most likely, two protein molecules in the asymmetric unit ($V_{\rm M}$ = 2.23 Å^3 Da^{-1}), corresponding to 45% solvent content. However, a dense packing of this small protein with three independent molecules could not be ruled out ($V_{\rm M}$ = 1.49 Å^3 Da^{-1}, 17.4% solvent content).

2.4. Structure solution by direct methods

Attempts to solve the structure by molecular replacement with *AMoRe* (Navaza, 1994) and *EPMR* (Kissinger *et al.*, 1999), using as a probe either the NMR model (1bxj) or the X-ray model of native CMTI-I from a complex with bovine trypsin (Bode *et al.*, 1989), were unsuccessful. Ultimately, the structure was solved by direct methods using the *ab initio* dual-space recycling procedure as implemented in *SHELXD* (Uson & Sheldrick, 1999). Because of the ambiguity of the asymmetric unit contents, the calculations were performed in parallel assuming two or three independent protein molecules. In retrospect, it is very puzzling that the first approach did not produce a solution even after many thousands of attempts. The second approach solved the structure in about 450 attempts, but the solution contained only two, not three, protein molecules. Surprisingly, the algorithm solved the structure in spite of a persistent 'uranium' peak located on the crystallographic twofold axis which was automatically rejected by the program. If it had been known that a heavy atom could be present on a special position, the input parameters could have been changed to retain and profit from it rather than rejecting it. Again the situation is clear in retrospect, as the structure turned out to be an unexpected protein–zinc complex. However, it must be stressed that the ability of the *ab initio* approach to solve a 58-amino-acid structure containing an electron-rich centre without *any* prior knowledge (or rather with incorrect presumptions) about the structure is an encouraging prognostic for the application of direct methods in macromolecular crystallography. The *E* map generated with the *ab initio* phases was of superb quality and included peaks corresponding to nearly all of the 444 non-H atoms of the two protein molecules. Some difficulties were encountered in model building at both Glu24 residues, but nevertheless a consistent tracing of both protein chains was easily achieved.

Table 1
Statistics of data collection and processing.

Values in parentheses are for the last resolution shell.

Crystal dimensions (mm)	0.40 × 0.15 × 0.10
Beamline	BW7A, EMBL
Wavelength (Å)	0.8919
Detector	MAR CCD
Temperature (K)	100
Resolution (Å)	15–1.03 (1.07)
Runs	3
Measured reflections	252201
Unique reflections	27045
$R_{\rm int}$	0.047 (0.151)
Redundancy	9.3 (3.5)
Completeness (%)	99.2 (98.3)
$\langle I/\sigma(I)\rangle$	30.7 (2.73)
Reflections with $I > 2\sigma$ (%)	83.7 (44.8)

2.5. Structure refinement

All structure-factor refinements were performed in *SHELXL* (Sheldrick, 1997; Sheldrick & Schneider, 1997). The initial refinement, including the two protein molecules (*A* and *B*) and the special position zinc ion, started with 2.5 Å data but the resolution was quickly extended in a stepwise procedure to include all reflections in the 10.0–1.03 Å range. Reflections in the range 15–10 Å (22 in total) were not included because of problems with an improperly centred beam stop. To account for diffuse solvent effects, a correction according to the Babinet principle was applied (Moews & Kretsinger, 1975). A randomly selected 5% of reflections (1345) were set aside for $R_{\rm free}$ testing (Brünger, 1992). The refinement included all reflections without any σ-cutoff, but the progress was monitored using R and $R_{\rm free}$ calculated for reflections with $F > 4\sigma(F)$. During the assembly of the protein model, 38 additional peaks with excellent density and clear interpretation were included as water molecules. At this stage of the refinement, the R and $R_{\rm free}$ values were 0.2707 and 0.2944, respectively (Fig. 1). Correction of the most obvious conformational errors and inclusion of an additional five water molecules reduced the R and $R_{\rm free}$ values to 0.2556 and 0.2757, respectively. From the very beginning, manual model adjustments, performed in *O* (Jones *et al.*, 1991), were based on $3F_o - 2F_c$ Fourier syntheses (for model maps), as recommended by Lamzin & Wilson (1997), and on $mF_o - DF_c$ syntheses (for difference maps), based on the coefficients recommended by Read (1997).

Addition of 13 fully occupied water molecules brought the R and R_{free} factors to 0.2121 and 0.2242, respectively. With the introduction of anisotropic displacement parameters, the residuals dropped by about 0.05 to 0.1542 and 0.1784, respectively. Introduction of a double conformation for the side chain of Arg5 of molecule A and addition of a few more water molecules and of a partially occupied S atom reduced the R and R_{free} values to 0.1475 and 0.1709, respectively. The introduction of the S atom was based on the observation of a strong difference-density peak on the twofold axis not far from the Zn^{2+} site and from the N-terminal ammonium groups. Together with some other difference peaks at about 1.65 Å, it was provisionally interpreted as a twofold symmetric sulfate ion introduced through $ZnSO_4$ in the crystallization buffer.

At step 6, the density maps for the Val21 and Tyr27 side chains in both molecules clearly showed the presence of alternate conformations. Electron density for the Leu23–Gly26 fragment of A suggested the possibility of double conformation for the main chain as well, but the maps in this area were generally poorly defined at this stage. Introduction of the above alternate conformation was accompanied by a drop of about 0.005 in both R and R_{free}. Conformational adjustments of the side chains in this region and of all lysine side chains plus inclusion of 12 water molecules decreased these values to 0.1344 and 0.1587, respectively. A similar situation could be seen in molecule B, but the electron-density maps were clearer. Consequently, the main chain and the side chains of residues Leu23–His25 in molecule B were modelled in two conformations and double conformations were introduced for the side chains of Val21 and Tyr27 and three high-B-value water molecules were deleted. These actions resulted in a noticeable increase in R and R_{free} (step 8). In round 9, some of the side chains in the double-conformation regions were manually adjusted in the maps and an additional 14 water molecules were included in the model. Introduction of geometrically generated riding H atoms decreased R and R_{free} by about 0.01 (step 10). The gradually improving electron-density maps in the double-conformation main-chain and side-chain region Leu23–His25 in B clearly suggested the existence of only a single conformation at this stage. In round 11, therefore, the Leu23–His25 region of molecule B was converted back to the original single conformation by removing the lower occupancy variant. This round also involved some remodelling in the solvent region, including addition of one partial MPD molecule and one partial glycerol molecule, and resulted in a decrease in the R and R_{free} factors to 0.1194 and 0.1495, respectively.

A similar situation, but at a later stage, was noted in the disordered region of molecule A. In round 12, only the original conformation of the fragment Leu23–Gly26 was retained, with two additional water molecules in the region of the deleted conformer. This action was accompanied by a slight increase in R, while R_{free} remained unchanged. The maps showed further significant improvement, supporting the presence of one dominating conformation. However, the two discrete $mF_o - DF_c$ peaks ($\sim$0.40 e Å^{-3}, 5.3σ level) near C^{β}(23) and C^{β}(24) found among the top four unaccounted-for residual peaks after the final refinement may indicate the presence of a minor component at very low occupancy.

A very high difference peak at 1.6 Å from $N^{\delta 1}$ and about 2.4 Å from the ring centre of His25 in molecule B, located about 1.4 Å above the ring plane, could be seen throughout the refinement. Its height progressively decreased with model improvement. For example, at round 11 it was 0.59 e Å^{-3} (8σ level) and was reduced to 0.37 e Å^{-3} (4.8σ level) after the final refinement.

In round 13, the occupancies of 30 water molecules were manually adjusted based on the electron-density maps. Two clearly visible O atoms of the twofold symmetric sulfate ion were also included. A nearby (1 Å) water molecule, evidently sharing the site with one of the sulfate O atoms, was made partial and complementary to that O atom. The occupancy of the sulfate ion was manually adjusted to 0.32. All solvent molecules, including one MPD and two glycerol molecules, were verified in the final cycles of the refinement, where the electron-density maps were much clearer in the solvent region.

As there was a sufficient number of experimental observations, in round 14 the five side chains in double conformations were assigned anisotropic displacement parameters. This caused a decrease in both residuals. The occupancies of these double conformations were fixed at the values refined in the previous runs. In runs 16–17, the BUMP restraints and the geometrical restraints on the main-chain segments without excessive displacement parameters ($B_{eq} < 15$ Å^2; Fig. 2a) were removed and the weights of the remaining stereochemical restraints were relaxed 1.25-fold (DEFS changed from the default values to 0.025 0.125 0.01 0.04; Sheldrick, 1997). The changes in R and R_{free} were negligible and the refinement remained stable and convergent.

The final refinement, after round 17 (Fig. 1) in which the test reflections were also included, was characterized by a data-to-parameter ratio of 5.2 and converged at R values of 0.1148 for 22 676 reflections with $F > 4\sigma(F)$ and 0.1206 for all reflections (26 990). The refinement statistics are summarized in Table 2. All refinements were performed using the conjugate-gradient algorithm. At the end of the refinement, one cycle of full-matrix minimization was calculated without any stereochemical restraints. The least-squares calculations were performed using the L.S. 1/BLOC 1 combination in *SHELX*97 (Sheldrick, 1997), which allowed estimation of the uncertainties in all positional parameters.

3. Results and discussion

3.1. Model quality

The protein model is of high quality and includes all atoms. Even the side chains in alternate conformations are generally modelled without ambiguity. The only exceptions are the side chains of Lys11 and Lys12 in both molecules and the Glu24 side chain in molecule A. The lysine side chains have clear $3F_o - 2F_c$ electron density up to C^{γ} at the 1.5σ level (0.85 e Å^{-3}). All these lysine side chains are exposed to

solvent and do not have any visible hydrogen-bonding partners or crystal contacts, although the possibility of hydrogen bonding cannot be ruled out for the N^{ζ} atom of Lys12 in molecule *B*, which is at a distance of 2.65 Å from Wat41, which in turn is hydrogen bonded (2.75 Å) to His25 $N^{\varepsilon 2}$ (*B*). Although the electron density is rather poor in this area, it is evident that a pair of carboxylic groups of Glu24 (*A*) form a short O···O (3.03 Å) contact across the twofold axis. The character of this interaction implies the presence of a single proton. Full protonation (two H atoms) is unlikely as it would typically result in a carboxylic acid dimer (Leiserowitz, 1976).

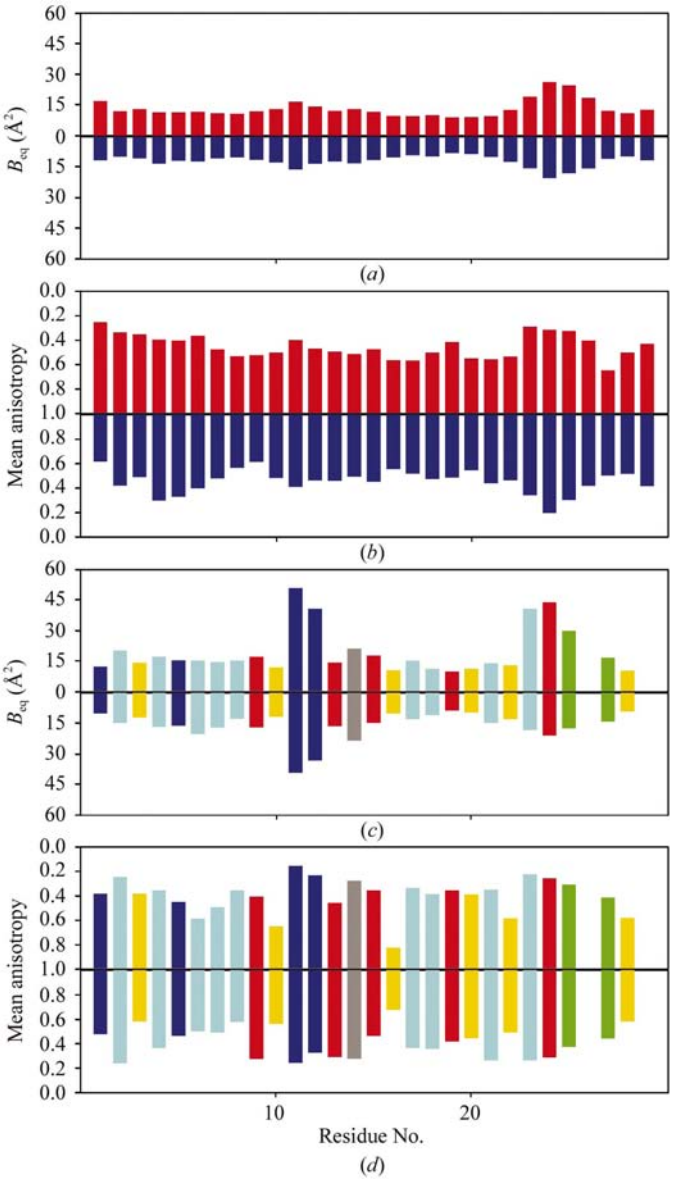

Figure 2
Analysis of the atomic displacement parameters. The equivalent isotropic displacement parameters B_{eq} (Å^2) are shown for the main chain and side chains in (*a*) and (*c*), respectively. The character of the side chains is colour-coded according to the convention of *SHELXPRO* (Sheldrick, 1997). Note that the pattern of B_{eq} is identical in both molecules (*A*, upper plots; *B*, lower plots). The mean anisotropy is illustrated in (*b*) and (*d*) for the main chain and side chains, respectively. Again, the pattern is similar in both molecules.

Table 2
Statistics of the refinement.

Program	*SHELX*97
Resolution (Å)	10.0–1.03
Reflections: total/R_{free}	26990/1345
Rejection criteria	None
Atoms	
Protein	476
Zn†	0.5
Water (full/partial)	46/30
Glycerol/MPD/sulfate†	2/1/1
Parameters refined	5179
R/R_{free}‡	
$F > 4\sigma(F)$	0.1141/0.1395
All reflections	0.1200/0.1449
R (including test reflections)	
$F > 4\sigma(F)$	0.1148
All reflections	0.1206
R.m.s. deviations from ideal	
Bonds (Å)	0.025
Angles (°)	2.7
Chiral volumes (Å^3)	0.106
Most favoured φ/ψ (%)	90.0

† Special position. ‡ Corresponds to step 17 in Fig. 1.

The complete main chain of both protein molecules is visible in contiguous $3F_o - 2F_c$ density at the 1.8σ level (1.03 e Å^{-3}), except for the Glu24–His25 region in molecule *A*. This region becomes contiguous at the 1.1σ level (0.63 e Å^{-3}). In the solvent region, the partially occupied MPD molecule has unambiguous $3F_o - 2F_c$ density at the 1.7σ level (0.97 e Å^{-3}). The first glycerol molecule is clearly seen at the 1.5σ level (0.85 e Å^{-3}) and the second one at the 1.3σ level (0.74 e Å^{-3}). The sulfate anion, with a very low occupancy of 0.32, is clearly visible at the 1.0σ level (0.57 e Å^{-3}) with $3F_o - 2F_c$ maxima located at the atomic positions.

90.0% of all residues are in the most favoured regions of the Ramachandran plot (Ramachandran *et al.*, 1963). The exceptions are Pro4 and Ala18 in both molecules. The Pro4 residues are in the additionally allowed regions. Pro4 precedes Arg5 on the enzyme-binding loop. The Ala18 residues of both molecules are characterized by positive φ angles but, as discussed below, they have excellent definition in electron-density maps and their departure from Ramachandran stereochemistry has good structural justification (see §3.8).

The quality of the model can also be assessed from the statistics of the derived geometrical parameters. For instance, the peptide C=O bond lengths (ranging from 1.19 to 1.28 Å, with a mean of 1.229 Å) are characterized by e.s.d.s between 0.01 and 0.03 Å, with a mean of 0.015 Å. The estimated standard deviation of the Zn—O distances is 0.007 Å. The r.m.s. deviation from the stereochemical standards (Engh & Huber, 1991) calculated for all bond lengths is 0.025 Å (Table 2). As noted earlier (Addlagatta *et al.*, 2001), this level of discrepancy is normal in high-resolution structures and is a convolution of model errors (reflected in the e.s.d. values) and the inadequacies of the standard targets themselves. In fact, when the refinement is well overdetermined, the restraints, when applied with the usual weighting level, have in practice very little influence on the optimization result in well ordered areas.

The highest and lowest peaks in the final $mF_o - DF_c$ map are 0.40 and -0.30 e Å^{-3}, respectively. In contrast to the situation in the 0.86 Å structure of BPTI (Addlagatta *et al.*, 2001), no difference electron density corresponding to H atoms could be seen in this 1.03 Å structure.

3.2. Atomic displacement parameters

In the refinement, all atoms were assigned anisotropic displacement parameters. Initially, partial occupancy atoms were treated isotropically in order not to proliferate the number of parameters. In the end, however, this approach was found counterproductive as it is the disordered regions that need additional parameters most, although the anisotropic tensor in such cases may have little physical meaning. The model includes riding H atoms generated on stereochemical principles. They were included in F_c calculations with B_{iso} proportional to the B_{eq} values of their carriers according to the rules adopted in *SHELX* (Sheldrick, 1997).

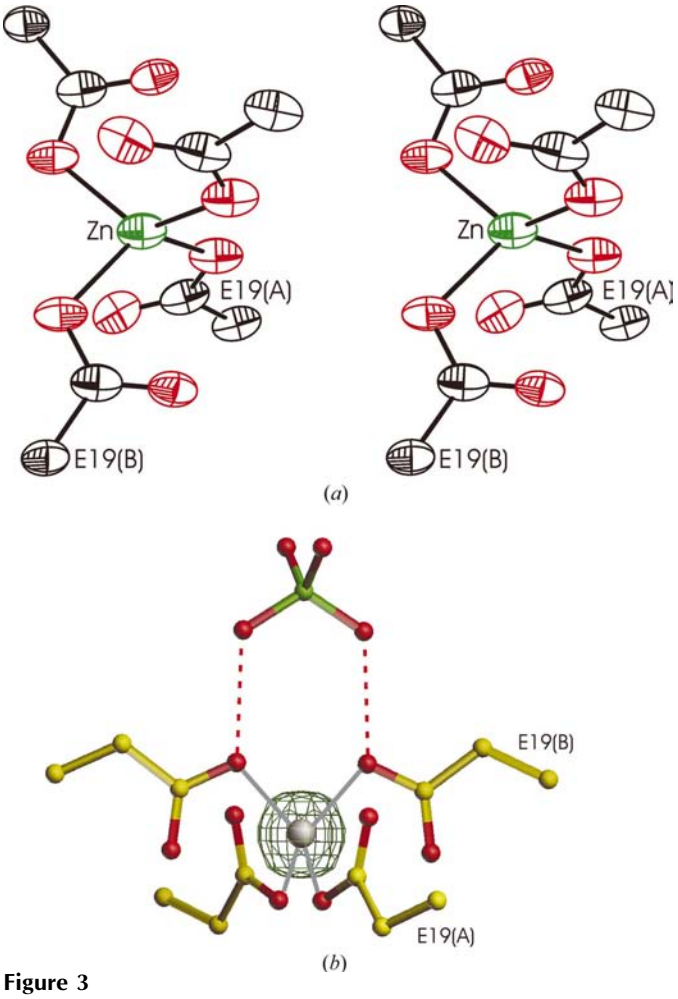

Figure 3
Coordination of the zinc cation by the Glu19 residues. (*a*) A stereoview of the anisotropic displacement ellipsoids drawn at 20% probability. The Zn^{2+} cation is on a twofold axis that runs across the paper. (*b*) The environment around Zn^{2+}, shown in its $3F_o - 2F_c$ electron density contoured at 1.99 e Å^{-3} (3.5σ). The zinc-coordinating O atom in molecule *B* must be protonated in order to form the hydrogen bond (broken line) to the sulfate anion, which is located on the same twofold axis as the Zn^{2+} cation.

The mean *B* factors for the main-chain, side-chain and solvent atoms are 12.5, 18.0 and 24.1 Å^2, respectively. From Fig. 2 it can be seen that the profile of the B_{eq} distribution in *A* and *B* is the same but that the absolute values for molecule *B* are systematically smaller. A similar situation is seen in the mean anisotropy profiles for both the main chains and the side chains. The cysteine residues, which are all involved in disulfide bridges, have smaller B_{eq} values and smaller mean anisotropies than the rest of the protein molecules. Also, the zinc-coordinating residues Glu19 have lower mean B_{eq} and mean anisotropy. The lowest B_{eq} in the protein model is found for the main chain of Glu19 in molecule *B*. Increased mobility is seen in the loop at Glu24–His25. This phenomenon was also observed in the complex between CMTI-I and trypsin (Bode *et al.*, 1989). The atoms forming the coordination complex undergo correlated motion in the *z* direction, as suggested by the elongated ellipsoids in Fig. 3(*a*). This rigid-body motion has only local character, limited to the Zn^{2+} cation and the Glu19 side chains.

3.3. Zinc coordination

The first peak on the *E* map was located on the twofold axis and was about three times higher than the next peak. Its intermolecular environment consisted of the Glu19 carboxylate groups from both protein molecules and their symmetric copies arranged in a tetrahedral fashion (Fig. 3). The peak height, coordination scheme and distances to the nearest O atoms (about 2 Å) leave no doubt that it represents a metal cation. The most likely candidate is zinc as it was present at a relatively high concentration in the crystallization medium and as its concentration had a strong effect on the crystallization outcome. The refinement results confirm this conclusion as the B_{eq} parameter for the zinc cation is 8.59 Å^2. The anisotropic ellipsoid of Zn^{2+} is slightly elongated along the twofold axis and there are some residual $mF_o - DF_c$ features (up to 0.3 e Å^{-3}) within 1 Å of the metal centre, but they are not even among the highest uninterpreted difference peaks. $3F_o - 2F_c$ electron density at the metal site can be seen up to 48σ (27.4 e Å^{-3}). The coordination tetrahedron is distorted primarily in the angular sense. The O—Zn—O′ angles generated by the twofold symmetry are much narrower (90.1 and 93.8°) than the remaining ones (116.7, 121.1°).

Coordinated unknown metal cations can be identified using the bond-valence parameters tabulated by Brese & O'Keeffe (1991). If we use the observed metal–oxygen distances, the bond-valence parameter R_{ij} calculated for a divalent cation is 1.734 Å, very close to the value of 1.704 Å listed by Brese and O'Keeffe for Zn^{2+}. If we use the reciprocal approach and estimate the experimental valence V_j of the assumed zinc cation, we obtain 1.843 (*i.e.* approximately 2), again consistent with our assumption. In view of all the above, one can be quite confident that the metal coordinated by the protein in the present crystal structure is indeed zinc.

There is an interesting issue concerning the protonation state of the Glu19 carboxyl ligands. Normally, one would assume that the carboxylic groups should be deprotonated, *i.e.*

$-COO^-$. However, the coordinating $O^{\varepsilon 2}$ atoms of the Glu19 *B* ligands are simultaneously at a hydrogen-bonding distance (2.68 Å) from one of the O atoms of the sulfate anion located nearby on the same twofold axis (Fig. 3*b*). These close contacts require the presence of H atoms, which can only be provided by the Glu19 carboxylic groups. It is true that the sulfate anion has fractional occupancy and that it is complemented, for interactions with the Glu19 carboxylate groups, by two copies of a partial water molecule. In theory, therefore, the protonation of the Glu19 group could also be of partial character. Chemical intuition suggests, however, that such a stable entity as a coordination complex should have a well defined composition. Therefore, we assume that the acid group of Glu19 in molecule *A* is deprotonated and that of Glu19 in molecule *B* is protonated. One should note that the pH of the crystallization solution (6.5) is not entirely incompatible with local protonation of carboxylate groups. Furthermore, introduction of two protons into a $Zn^{2+}(COO^-)_4$ complex balances out the electrostatic charges and reduces the repulsions between the negatively charged ligands. A similar situation with a zinc cation coordinated simultaneously by two carboxylic and two carboxylate groups has been reported previously in a small-molecule crystal (Diskin-Posner *et al.*, 2000), where metal coordination was used as a designer tool to construct a synthon for supramolecular self-assembly. It is interesting to note that the carboxylate (charged) ligands have shorter Zn—O distances [1.978 (7) Å compared with 2.004 (7) Å] and form a somewhat wider angle than the neutral species.

Glu19 has virtually the same conformation in *A*, *B* and the CMTI–trypsin complexes (Bode *et al.*, 1989; Helland *et al.*, 1999). A carboxylate O atom of this side chain accepts a hydrogen bond from the NH group of Val2 in both cases. These interactions seem to stabilize the protein structure (see §3.10). In the six NMR structures of the 3cti ensemble (Nilges *et al.*, 1991) there are equal populations of this residue pointing towards the Val2 NH group and pointing away. The N—H···O hydrogen bond between Val2 and Glu19 was not identified in the NMR structure of the present mutant (1bxj) because the first two N-terminal residues as well as the Glu19 side chain had flexible conformation and could not be identified without ambiguity. Nevertheless, the temperature coefficient of the NMR signal of the Val2 amide proton strongly suggested that it participates in (at least partially populated) intramolecular hydrogen bonding (Zhukov *et al.*, 2000).

3.4. The enzyme-binding epitope

The enzyme-binding loop is from Arg1 to Glu9. The central residue of the inhibitory loop is Arg5, which penetrates the enzyme active site to form a strong salt bridge (Bode *et al.*, 1989). In molecule *A*, Arg5 exists in two conformations with occupancies of 0.69 and 0.31. The minor alternative is similar to the single conformation observed in molecule *B*, while the major one corresponds to the conformation reported in the CMTI-I–trypsin complexes (Bode *et al.*, 1989; Helland *et al.*, 1999). The scissile peptide bond in molecule *A* is planar (178.8°), while in molecule *B* this ω torsion angle shows a significant deviation from planarity (171.4°). This peptide bond is almost planar in the CMTI/CPTI–enzyme complexes studied thus far.

With high precision of the stereochemical parameters, it is possible to assess the pyramidalization of the carbonyl C atom at the scissile peptide bond. Using an approach similar to that of Marquart *et al.* (1983) and Esposito, Vitagliano, Zagari *et al.* (2000), it is found that the pyramidalization parameter χ_C as defined by Winkler & Dunitz (1971) is −2.2° in molecule *A* and −2.6° in molecule *B*. As expected in the absence of an attacking nucleophile, these distortions are very small. We note from the present example that there is no correlation between the carbonyl out-of-plane (pyramidalization) and torsional (ω planarity) distortions of the peptide group. It is noteworthy that in the four enzyme complexes of CMTI/CPTI (1ppe, 2sta, 2stb, 2btc) there is also no noticeable C-atom pyramidalization (−4 to +4°) at the scissile peptide. This is in contrast to the situation reported by Marquart *et al.* (1983) in complexes of bovine pancreatic and Kazal inhibitors, where the pyramidalization is large (−19 to −10°).

There is a significant variation in the main-chain region from the Pro4 C atom to the N atom of Ile6 in *A* and *B*. The maximum deviation (at Arg5 O) is 1.86 Å and the mean deviation between the corresponding main-chain atoms is 1.10 Å. The same segment in the CMTI/CPTI–trypsin complexes lies virtually halfway between the two extremes observed in *A* and *B*. This indicates that this region of the protein is quite flexible and that this plasticity may be required for optimal fit to the enzyme's binding cleft. The conformations of Pro4 and Leu7 are virtually identical in *A*, *B* and the CMTI-I–trypsin complexes, while the side chains of Ile6 have a similar general disposition but different rotamers. The major rotamer of Arg5 in molecule *A* forms two hydrogen bonds with two main-chain O atoms of neighbouring molecules *A* in the crystal [$N^{\eta 1}\cdots$O(14), 2.88 Å; $N^{\eta 2}\cdots$O(16), 2.93 Å], while both the terminal N atoms of the guanidinium group of the minor component point to $O^{\delta 2}$(15) of the same neighbour (Fig. 4*a*). A partial water, Wat68, which shares its site with the minor rotamer, is also hydrogen bonded to the major rotamer and Asp15 ($N^{\eta 2}\cdots$Wat68 3.29 Å; $O^{\delta 2}\cdots$Wat68 3.00 Å). Arg5 in molecule *B* forms four hydrogen-bonding interactions with neighbouring molecules of *A*, *B* and a water molecule (Fig. 4*b*), in which all the guanidinium N atoms are involved [$N^{\varepsilon}\cdots$O(7) (*A*), 2.84 Å; $N^{\eta 1}\cdots O^{\delta 2}$(15) (*B*), 2.98 Å; $N^{\eta 1}\cdots$Wat12, 2.97 Å; $N^{\eta 2}\cdots$O(7) (*A*), 2.98 Å].

3.5. The mutation site

The main-chain segment at the mutation site (Met8Leu) is virtually identical in *A*, *B* and the CMTI/CPTI–trypsin structures (Bode *et al.*, 1989; Helland *et al.*, 1999). To assess the extent of deviations, the structures have been compared using the *ALIGN* program (Cohen, 1997). In all the superpositions calculated in this study, the molecules of *A*, of CMTI-I and CPTI from the PDB files 1ppe, 2stb and 2btc or of 3cti and 1bxj from the NMR studies were overlaid on the current

molecule of *B* using all C^{α} atoms (see Table 3 for other alignments). The mean deviation of the main-chain segment at residue 8 in *A* is 0.34 Å and it is even less in the complex structures. The deviations are much higher, about 0.71 and 1.03 Å for the native (3cti) and Met8Leu (1bxj) structures determined by NMR. The orientation of the Leu8 side chain in *A* and *B* is identical and analogous to that seen for Met8 in the native protein complexes. These hydrophobic residues (Leu, Met), along with other hydrophobic residues (Ile6 and Leu17), serve the same purpose in the crystal structures. They close one end of an internal solvent channel within the CMTI-I molecule, the opposite end of which has an open mouth formed partly by Arg5.

3.6. His25 conformation

A proton NMR study by Krishnamoorthi *et al.* (1992) has suggested that at pH > 6 not only the protonation state (deprotonation) but also the conformation and interactions of His25 undergo profound changes from the situation observed at lower pH by NMR techniques (Holak *et al.*, 1989; Nilges *et al.*, 1991) as well as by X-ray crystallography (trypsin complex; Bode *et al.*, 1989). The model proposed by Krishnamoorthi and coworkers involved a hydrogen bond between His25 $N^{\varepsilon 2}$ and Tyr27 OH. The present high-resolution structure of a crystal grown at pH 6.5 does not support that model. Firstly, we note that the conformation of the His25 side chain is different in each of the present CMTI-I molecules and at the same time different from the low-pH crystallographic model deduced from the trypsin complex (Bode *et al.*, 1989). His25 in molecule *A* has its side chain entirely exposed to the solvent region, while the side chain of His25 in molecule *B* is rotated towards the core of the protein (Fig. 5). Secondly, the site of protonation of His25 in the two molecules is different. In molecule *A*, the proton resides at $N^{\delta 1}$ and is involved in a hydrogen bond to the carbonyl O atom of His25 from a different copy of molecule *B* ($N^{\delta 1}\cdots O$, 2.85 Å). In molecule *B*, $N^{\delta 1}$ cannot be protonated as it is evidently an acceptor of an intramolecular hydrogen bond from the amide NH group of Tyr27 ($N\cdots N^{\delta 1}$, 3.19 Å). Instead, $N^{\varepsilon 2}$ is protonated and involved in a strong hydrogen bond to a fully occupied water molecule ($N^{\varepsilon 2}\cdots O$, 2.75 Å) (Fig. 5). It has to be admitted, though, that the electron density of the latter histidine ring has an unusual feature consisting in a peak close to $N^{\delta 1}$ but not in the imidazole plane (see §2.5). A possible explanation that it might reflect an electronic effect caused by

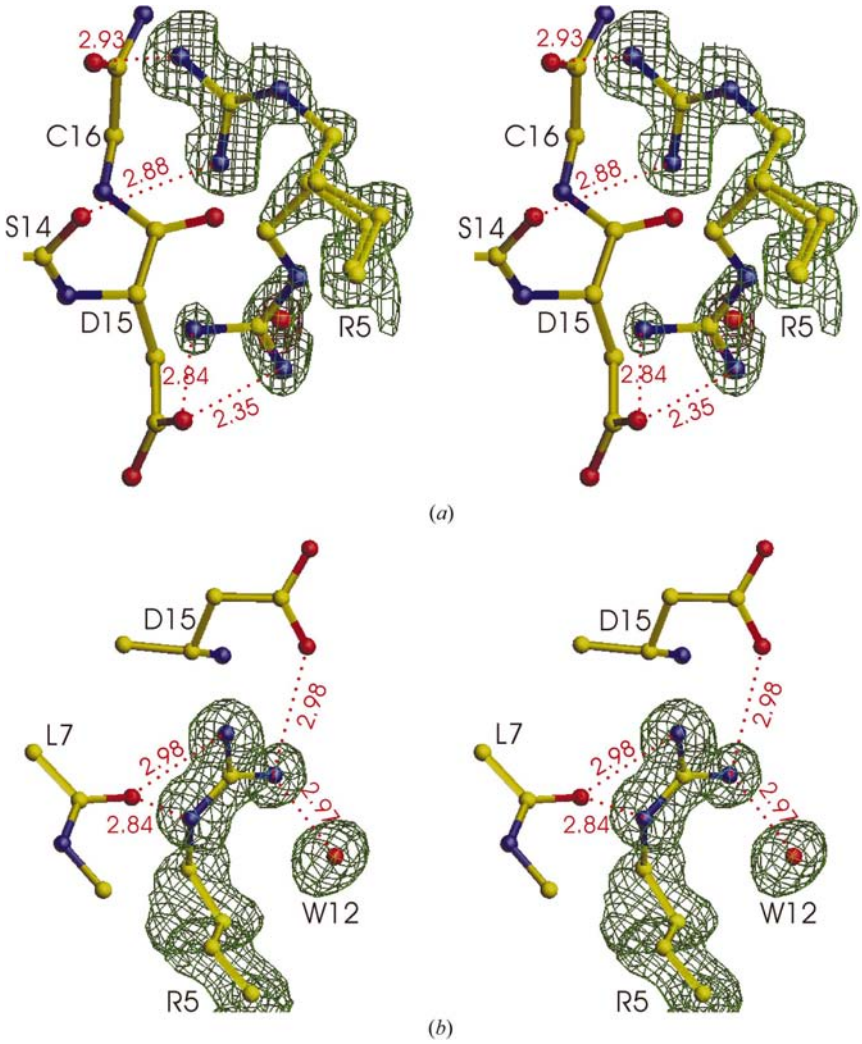

Figure 4
Stereoviews of the crystal environment and electron density around Arg5 with hydrogen-bonding distances in Å. (*a*) Arg5 in *A* is modelled in two conformations. Notice that the minor component (occupancy 0.31) shares the site with a partial water (occupancy 0.69). The $3F_o - 2F_c$ maps are contoured at two different levels, 0.85 e $Å^{-3}$ (1.5σ) for Arg5 (green) and at 1.7 e $Å^{-3}$ (3.0σ) for the partial water molecule (red). (*b*) Interactions of Arg5 in *B*. The $3F_o - 2F_c$ electron-density map for Arg5 and for the fully occupied Wat12 is contoured at 0.85 e $Å^{-3}$ (1.5σ).

Table 3
Coordinate superpositions (calculated in *ALIGN*; Cohen, 1997) using molecule *B* as the target.

The superpositions are calculated for the *A*, 1ppe, 2sta, 2stb, 2btc, 3cti and 1bxj models. The results, in terms of r.m.s. deviations (Å), are presented for C^{α} and all-main-chain superpositions.

Model	C^{α}	Main chain
A	0.467	0.426
1ppe	0.436	0.442
2sta	0.329	0.348
2stb	0.325	0.341
2btc	0.460	0.471
3cti†	0.600	0.617
1bxj‡	1.617	1.536

† The first of the six NMR models deposited in the PDB. ‡ The NMR model closest to the average structure (Zhukov, personal communication).

the N—H···$N^{\delta1}$ hydrogen bond is not convincing as this interaction is rather weak. Interestingly, the N—H group of Tyr27 in molecule *A* is hydrogen bonded to a partial water molecule that is located at the site corresponding to the residual difference peak near His25 in molecule *B*. The fractional occupancies of the tyrosine side chains in molecules *A* and *B* are of nearly identical proportions: 0.51:0.49 and 0.54:0.46, respectively. The minor components correspond to the orientation of Tyr27 in the CMTI/CPTI–trypsin complexes (Bode *et al.*, 1989; Helland *et al.*, 1999), while the major ones are oriented away from His25.

3.7. Residues in double conformation

One of the advantages of high-resolution structures is that they enable more accurate estimation of the extent and character of disorder. Typically, high-resolution models have more disordered fragments simply because it is possible to delineate alternate conformations that at lower resolution are blurred or invisible. For example, in the 0.86 Å resolution structure of BPTI (Addlagatta *et al.*, 2001), about 20% of the residues display disorder, but in all cases they could be modelled in distinct multiple conformations. In addition, it was possible to find concerted behaviour of the disordered fragments not only on the intramolecular but also on the intermolecular scale. The proportion of disordered residues is lower in the present 1.03 Å structure: about 11% in *A* and about 7% in *B*. However, the electron density, even for the 0.3 occupancy conformers of Arg5 in *A* and Val21 in *A* and *B*, was very clear at all stages of the refinement. The situation was even more obvious in the case of the disordered side chains of Tyr27. Identical disorder with equal proportions of the two alternate conformations is seen for the Val21 and Tyr27 side chains in both molecules. There is no obvious reason why these occupancies should be almost the same in the two crystallographic copies of the molecule. It may be noticed that Val21 is immediately flanked by right-handed disulfide bridges (Cys3–Cys20 and Cys10–Cys22). In both molecules Tyr27 is exposed to solvent. One of the alternatives in both *A* and *B* is bound to a partial water molecule which is further anchored to the protein. The alternative of Tyr27 in *B* is also bound to a free partial water (Fig. 6).

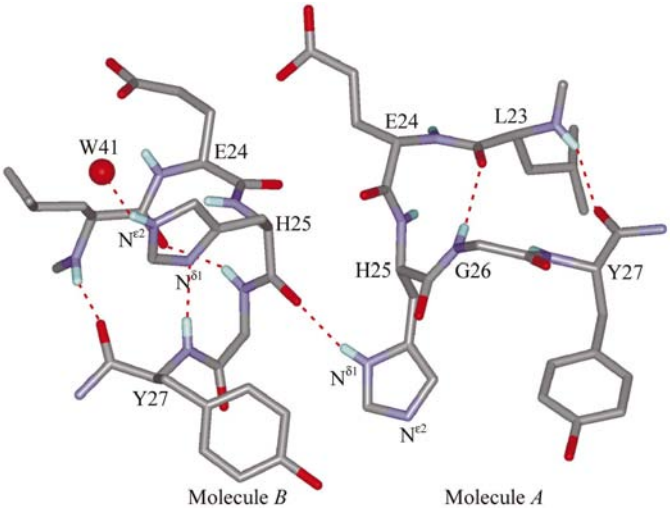

Figure 5
His25 in the two independent molecules. The imidazole ring of His25 in molecule *A* is in the solvent region and forms a hydrogen bond to the C═O acceptor of His25 in molecule *B*. The N—H donor in this interaction is at $N^{\delta1}$. The His25 side chain in molecule *B* has a completely different orientation and its $N^{\delta1}$ atom cannot be protonated as it is an acceptor in an intramolecular hydrogen bond from N—H of the Tyr27 main chain. The relative orientation of the His25 and Tyr27 side chains in both molecules precludes any direct hydrogen bonding between the aromatic rings. The intermolecular $N^{\delta1}(25A)\cdots O(25B)$ bond is to a symmetry-related copy of molecule *B*.

3.8. Residues deviating from Ramachandran geometry

In both molecules the highly strained Ala18 residue lies in a generously allowed Ramachandran region (see §3.1) with very unusual φ/ψ angles (55.0/−128.6 and 52.9/−131.0°). Ala18 is situated in a tight turn that is reinforced *via* an N(20)···O(17) hydrogen bond. Since it is located between two disulfide-forming cysteine residues (Cys16 involved in a left-handed bridge, Cys20 involved in a right-handed bridge), it could be argued that its conformation is forced by the stereochemical requirements of those two closely spaced S—S bridges. These bridges are topologically interesting because they form a covalent shortcut connecting the N- and C-termini (Fig. 7), which, as mentioned elsewhere, are additionally joined by an Arg1–OOC(29) salt bridge. This disulfide system is further stabilized by two antiparallel β-sheet-type hydrogen bonds involving the terminal Gly29 residue and Val21, which is downstream of the Ala18 bend (19–21). In turn, the N-terminal fragment, involving the Arg1 side chain and the main chain of residue 2, is connected to the 19–21 fragment with three N—H···O hydrogen bonds. Two of them, $N^{\eta1}(1)\cdots O(19)$ and $N(2)\cdots O^{\epsilon1}(19)$ interconnect the side chains and the main chain. The third bond, N(19)···O(2), clamps together two tight turns with roughly perpendicular

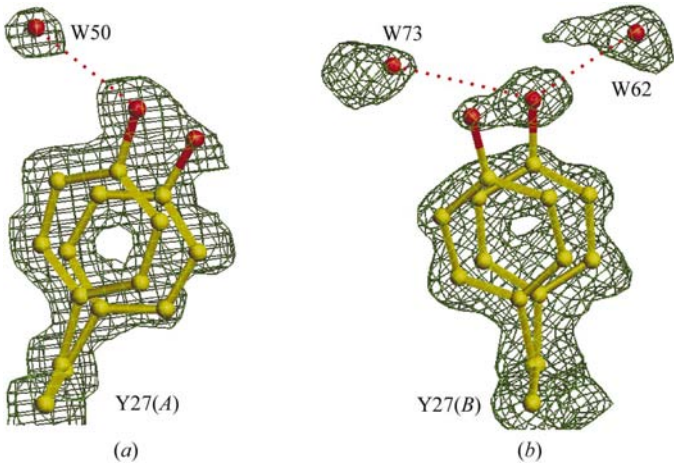

Figure 6
Double conformation of the Tyr27 side chain. (*a*) In molecule *A*, one of the conformers is hydrogen bonded to a partial water molecule. (*b*) In molecule *B*, one of the conformers is hydrogen bonded to two partial water molecules. The $3F_o - 2F_c$ electron density is contoured at 0.57 e Å^{-3} (1σ) in (*a*) and at 0.68 e Å^{-3} (1.2σ) in (*b*).

faces: the Ala18 turn and the turn formed *via* the right-handed disulfide bridge (Cys3–Cys20). In order for this reinforcement to be present, the Ala18 Ramachandran angles have to assume their unusual values. This is illustrated for molecule *B* in Fig. 7. For clarity, only the side chains of interest are shown. In short, the above strained system is stabilized by as many as eight hydrogen bonds, with the main-chain fragment Glu19–Val21 acting as a central template between the two above-mentioned disulfide bridges. The Glu19 residue, which is tightly coordinated to the zinc cation in the present structure, does not have any effect on this strained system, as a virtually identical situation is also found in the CMTI-I–trypsin complexes.

3.9. Disulfide bridges

The stable fold of this small protein is based on the presence of three disulfide bonds. Two of them have the less favoured right-handed configuration (Cys3–Cys20, Cys10–Cys22), while the third is left-handed (Cys16–Cys28). The electron density of all six Cys–Cys bridges is excellent. The C—S—S′—C′ torsion angles, given in Table 4, are close to idealized values and are nearly the same in the two present molecules, in the CMTI-I–trypsin complexes and even in the earlier NMR structure (3cti). In this light, it is rather surprising that in the NMR model of the Met8Leu mutant (1bxj) not only are the torsion angles far away from the preferred values, but the Cys3–Cys20 and Cys16–Cys28 bridges also have chiralities that are opposite to those in the other models.

3.10. The polypeptide fold

In addition to a short and irregular 3_{10}-helix of one and a half turns (Lys12–Cys16), the protein structure also contains a short intramolecular twisted antiparallel β-sheet in the region from Val21 to the C-terminus, including a hairpin turn around Glu24–His25. At one of its ends, this hairpin is connected to Arg1 through the C-terminus–Arg1 salt bridge. Arg1 also interacts with the Glu19 extension at the other end of the hairpin, forming an elongated twisted ring with main chain–side chain hydrogen bonds resembling β-sheet interactions [$N^{\eta1}(1)\cdots O(19)$, $N(2)\cdots O^{\varepsilon1}(19)$] (Fig. 8). In addition, there is a main chain–main chain hydrogen bond in this area [$N(19)\cdots O(2)$].

Table 4
Disulfide C—S—S′—C′ torsion angles (°).

Structure	Cys3–Cys20	Cys10–Cys22	Cys16–Cys28	Reference
A	103	91	−89	This work
B	100	90	−88	This work
1ppe	98	91	−89	Bode *et al.* (1989)
2sta	97	81	−87	Helland *et al.* (1999)
2stb	101	86	−86	Helland *et al.* (1999)
2btc	104	89	−87	Helland *et al.* (1999)
3cti	115	90	−95	Nilges *et al.* (1991)
1bxj	−167	65	132	Zhukov *et al.* (2000)

3.11. N- and C-termini

The N- and C-termini of both molecules are clearly defined in electron-density maps. In both cases, the side chain of Arg1 forms an intramolecular salt bridge with the C-terminal COO^- group. The ammonium groups of the terminal Arg1 residues are not involved in direct interactions with the C-terminus, but form three well defined hydrogen bonds (Fig. 9). In the high-resolution structure of BPTI (Addlagatta *et al.*, 2001) the termini are also in a hydrogen-bonding contact and the first residue is also arginine. However, this interaction, which is unique among BPTI crystal structures in that it provides stabilization of the C-terminus analogously to the situation in cyclic BPTI (Botos *et al.*, 2001), involves direct interactions of the terminal ammonium and carboxylate groups (Czapinska *et al.*, 2000). In CMTI-I, the separation between the C^α atoms of the N- and C-terminal residues is about 10 Å and is preserved in molecules *A* and *B* and in the complexed forms of CMTI-I (1ppe, 2sta). The C-terminus–Arg1 salt bridge is symmetrical in both *A* and *B*, the N···O separations being 2.85, 2.85 and 2.87, 2.91 Å, respectively, whereas in the 1ppe CMTI-I–trypsin (bovine) complex (Bode *et al.*, 1989) it is unsymmetrical (2.74, 3.39 Å). The situation is different in the 2sta complex with salmon trypsin and in the NMR structures, where this salt bridge is absent.

The hydrogen-bond partners of the N-terminus are clearly seen in the electron-density maps in both molecules, in contrast to the situation in the 1ppe CMTI-I–trypsin complex (Bode *et al.*, 1989). In molecule *A*, three water molecules (one of them with

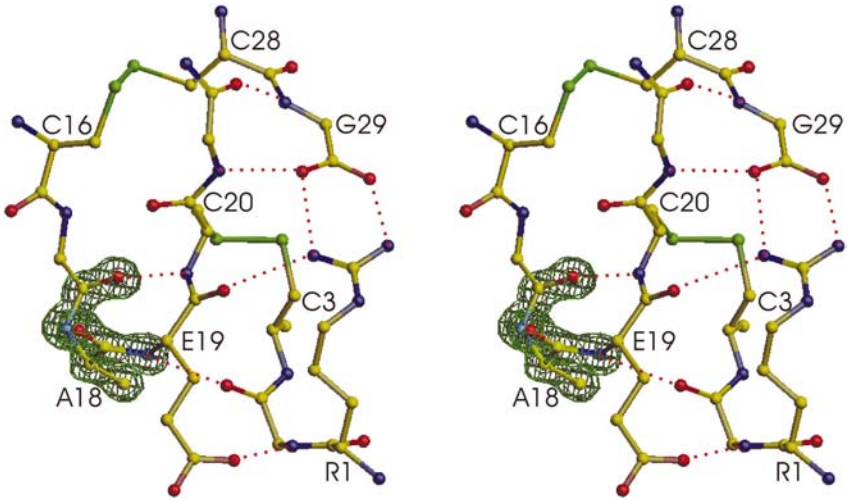

Figure 7
Unusual peptide geometry around Ala18 (stereoview). $3F_o - 2F_c$ electron density contoured at 1.71 e Å^{-3} (3σ) around Ala18 in *B*. Ala18 is situated in a tight turn that lies between the Cys3–Cys20 (right-handed) and Cys16–Cys28 (left-handed) disulfide bridges. These bridges form a covalent shortcut connecting the N- and C-termini. The hydrogen bonds (dotted lines) clamp together the Ala18 turn, the turns formed *via* the disulfide bridges and the 'turn' formed by the Arg1 side chain. For clarity, only those side chains that are involved in hydrogen bonding are shown.

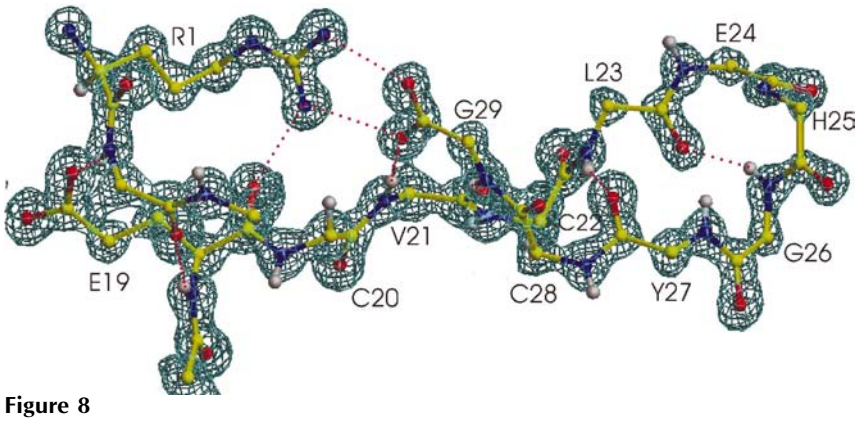

Figure 8
An antiparallel intramolecular twisted β-hairpin (in *B*) in the region Val21–Gly29. The $3F_o - 2F_c$ map is contoured at 1.42 e Å^{-3} (2.5σ). The hairpin is connected to Arg1 through the C-terminus–Arg1 salt bridge and through main chain–side chain hydrogen bonds that mimic β-sheet interactions.

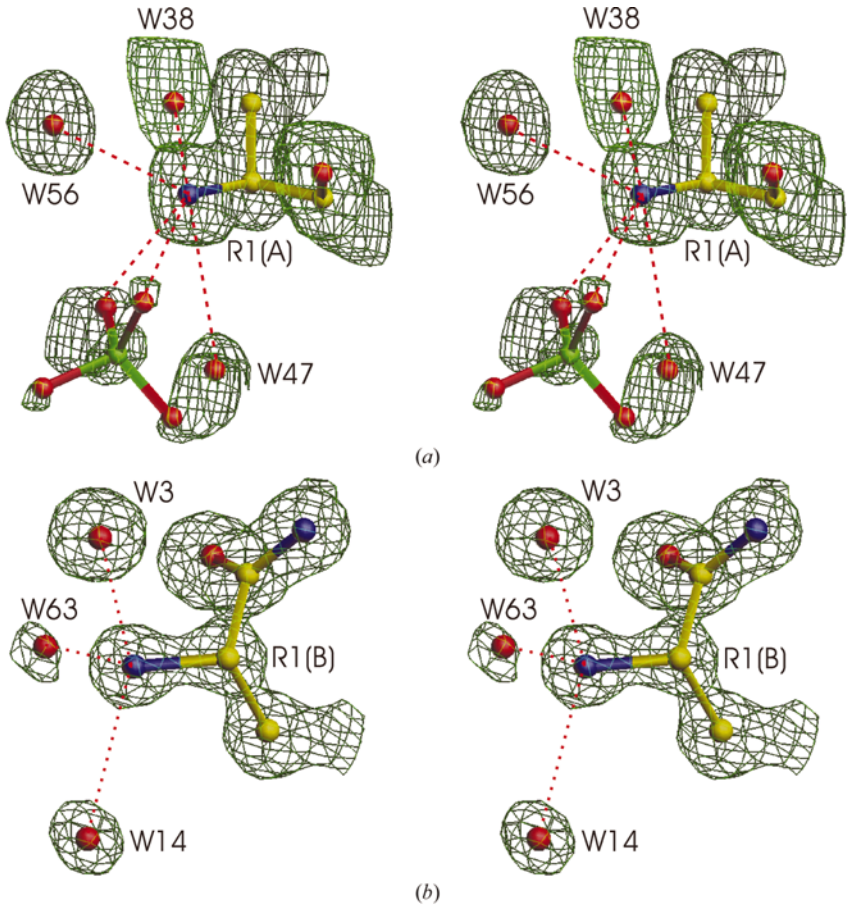

Figure 9
Stereoviews of the N-termini. (*a*) The N-terminus of molecule *A* interacting with three water molecules. One of them, Wat47, has partial occupancy (0.68), complemented by an O atom of the twofold-symmetric sulfate anion. The $3F_o - 2F_c$ map is contoured at 0.88 e Å^{-3} (1.5σ). (*b*) The N-terminus of molecule *B* is hydrogen bonded to two full and one partial (Wat63) waters. The $3F_o - 2F_c$ map is contoured at 0.97 e Å^{-3} (1.7σ).

fractional occupancy) approach the ammonium N atom at tetrahedral directions. There are additional electrostatic interactions with the partial and symmetric sulfate anion, two O atoms of which share the site with the partially occupied Wat47 (Fig. 9). In molecule *B*, two full and one partial water molecules approach the N-terminus (Fig. 9). The additional stabilization of the N-terminus in *A* by the electrostatic interactions with the sulfate anion is reflected in the lower B_{eq} and mean anisotropy of this N atom (Fig. 2).

3.12. Structural parameters

3.12.1. Planarity of the peptide bonds. It has been pointed out previously that the peptide bond can deviate from strict planarity by as much as ±20° (MacArthur & Thornton, 1996; Wilson *et al.*, 1998; Sandalova *et al.*, 1999; Addlagatta *et al.*, 2001). In the present structure the ω torsion angles deviate within +12 and −16° from 180°, while the standard errors of the individual ω angles in the well defined fragments of the protein are about 1°. The planarity of the peptide bond depends on the conformation of the neighbouring residues. For example, the highest deviation (−16°) is seen for Leu17 in molecule *B*. In molecule *A* it is also very large, about −13°. Residue 17 is flanked by Ala18, which has a highly strained conformation (see §3.8; Fig. 7). The peptide bond of the other neighbour of Ala18, Glu19, is also distorted to a significant degree, the ω deviations in molecules *A* and *B* being about −13 and −10°, respectively. As mentioned earlier, several hydrogen bonds stabilize the protein conformation in this area. This implies that under suitable conditions hydrogen bonds can influence, directly or indirectly, the peptide-bond planarity. However, it is not clear why the peptide bond of the highly deformed Ala18 is planar in both molecules. Another illustration of the influence of hydrogen bonding on peptide planarity is provided by the case of Arg5. This residue has different crystal environments in the two copies of the molecule (Fig. 4). The ω angle of Arg5 in molecule *A* is 179° and at the same time this residue is very weakly anchored in the structure and has a disordered side chain. In contrast, Arg5 in molecule *B* forms numerous stabilizing hydrogen bonds and its peptide bond deviates from planarity by 9°. The above observations confirm that strained geometries can be stabilized by both intramolecular and intermolecular hydrogen bonds. In assessing the influence of hydrogen bonding on the planarity of the peptide bond, one has to consider the interactions at the peptide's N—H and O centres (which may contribute to the distortion *via* purely

geometrical as well as electronic effects) and the interactions formed by the side chain (which may be transmitted as distortions of the main chain).

3.12.2. N—C$^{\alpha}$—C angles. Deviations of up to 5° on either side of the target value (111.2°) can be seen in the N—C$^{\alpha}$—C angles. The standard errors of the individual values in the best fragments of the structure are about 1°. The large spread of this angle is usually correlated with secondary structure (Ashida *et al.*, 1987; Jiang *et al.*, 1995; Karplus, 1996; Esposito, Vitagliano, Sica *et al.*, 2000). Such a correlation is not possible in the present case because of the lack of properly formed secondary-structure elements and because of the limited number of data. The highest N—C$^{\alpha}$—C deformations are observed at the C-terminal glycine residues from both *A* and *B*, and Arg1 of *B*. The perturbed geometries may be a consequence of the involvement of these residues in several hydrogen bonds. The C-termini form intramolecular hydrogen bonds with Arg1. Arg1, Pro4 of *B* and Cys22 of *A* also lie in the extreme regions of the N—C$^{\alpha}$—C histogram (not shown). All these residues have unambiguous electron density.

3.13. Solvent structure

3.13.1. Water structure. The high resolution of the present structure allows a clear definition of the solvent area. The final model includes 46 fully and 30 partially occupied water sites. Allowing for the presence of the Zn cation, organic solvent molecules and the sulfate anion, the Matthews volume (V_M = 1.95 Å^3 Da^{-1}) predicts 165 water molecules in the solvent volume of the asymmetric unit. The current structure therefore models 46% of the theoretical water sites. Naturally, this fraction of accounted-for water sites increases with resolution. For instance, this number is twice as high in the 0.86 Å BPTI structure (Addlagatta *et al.*, 2001). The quality of the present model of the solvent region is illustrated by the observation of a number of water molecules with tetrahedral coordination. The H atoms of water and other solvent molecules were not included in F_c calculations.

There are three tightly bound water molecules in both *A* and *B* with similar hydrogen-bond connectivities. The same water structure was also found in the CMTI/CPTI–trypsin complexes. These three water molecules are buried in an internal channel, which is closed at one end by three hydrophobic residues, Ile6, Leu8 and Leu17. This is shown for molecule *A* in Fig. 10. The water molecule located at the bottom of the channel is hydrogen bonded to N(17) and connected to a second water lying on top of it. The second water is further hydrogen bonded to O(4) and O(6). The water molecule at the mouth of the channel is tightly bound to the protein with three hydrogen bonds. In the CMTI/CPTI–trypsin complexes, the inhibitor's surface at the entrance to the water channel is covered by a complementary surface area of the enzyme. Interestingly, in the present structure this region is surrounded by organic solvent molecules. In *A*, the entrance to the cavity is capped by a partial glycerol molecule (Fig. 10), while in *B* a partial MPD molecule plays the same role.

3.13.2. 2-Methyl-2,4-pentanediol. The MPD used in the crystallization buffer was a commercial product (Fluka) that is a racemic mixture of the *R* and *S* stereoisomers. The difference electron-density maps at stage 11 showed a large and clear patch in the solvent region near the surface of the protein molecule *B* that could be best interpreted as corresponding to the *R* isomer of MPD. The geometry of the solvent molecules was restrained during the refinements using stereochemical parameters generated by a statistical analysis of the data deposited in the Cambridge Structural Database (Allen & Kennard, 1991). The occupancy of the MPD molecule was refined at the isotropic stage and converged at 0.5. Both hydroxyl groups of the MPD moiety are involved in hydrogen bonds. One of them forms one hydrogen bond with the side chain of Asp15 (*B*) and one with a full water molecule. The other hydroxyl group is linked to the main-chain O atom of Cys28 (*B*). The quality of this partial-occupancy molecule is reflected in the two O—C—C—C torsion angles along its backbone, which refined to fairly good values (75, −40°), although they have never been subjected to stereochemical restraints.

3.13.3. Glycerol molecules. There are two glycerol molecules included in the model. Their source can be linked to the cryoprotecting solution. One of them (Gol1) occupies a site near the surface of molecule *A* that is analogous to that occupied by the MPD molecule near molecule *B*. The other one (Gol2) is located on the surface formed by the side chains Glu9, Lys11 and Asp15 of molecule *B*. Gol1 has occupancy 0.60 and is

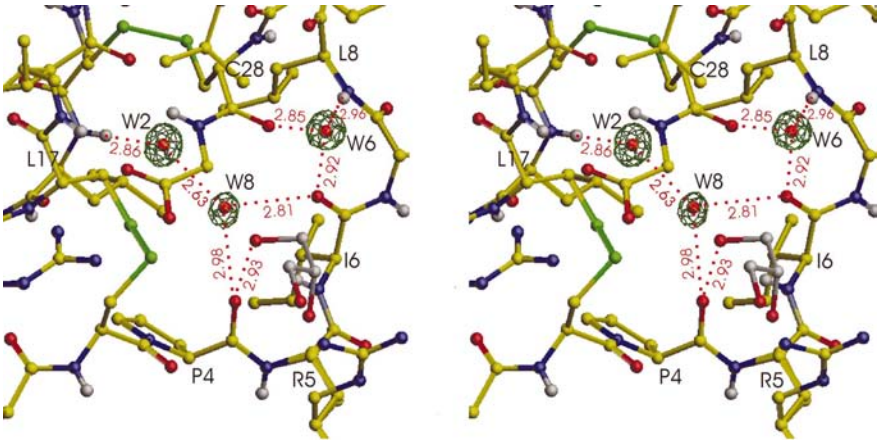

Figure 10
A stereoview of three tightly bound water molecules in *A*, shown in their $3F_o - 2F_c$ electron density contoured at 1.99 e Å^{-3} (3σ). Wat2 and Wat8 are placed deeply in an internal channel which is narrowed at one end by the Ile6, Leu8 and Leu17 hydrophobic side chains. Wat6, located near the surface, has three hydrogen-bond links, exclusively to the protein. A glycerol molecule (the C atoms of which are shaded grey for clarity) caps the channel at the other end. Notice that the minor rotamer of Arg5 is bent towards the glycerol molecule. Hydrogen-bond distances (dotted lines) are given in Å.

involved in five hydrogen bonds, three with molecule *A*, one with a fully occupied water molecule and one with a partial water molecule. In these contacts, all three OH groups of the solvent molecule are engaged. One of the hydrogen bonds to molecule *A* closely mimics that linking MPD and Asp15. The other two are hydrogen bonded to O^{γ}(14) and O(4). The two O—C—C—C torsion angles, again unrestrained during the refinement, have nearly perfect values of 176 and −58°. Gol2 has occupancy 0.45 and forms, through one of its OH centres, three rather long contacts with Asp15 (*B*) and a full water molecule. The O—C—C—C torsion angles are also very good, −67 and −164°. The two glycerol molecules have virtually identical conformation even though they have unrelated intermolecular contacts and were modelled independently in the electron-density maps.

3.13.4. Sulfate anion. Even at an early stage of model building and refinement, an outstanding $mF_o - DF_c$ peak on the twofold axis not far from the zinc site suggested a tetrahedral anion. Since zinc sulfate was used in the crystallization buffer, this density was modelled as a twofold-symmetric SO_4^{2-} anion. It was successfully refined even though the occupancy is as low as 0.32. The O atom (and its symmetry mate) facing the zinc-coordination centre overlaps with a partial water molecule (and its symmetry mate) of complementary occupancy and displays hydrogen bonding with one of the Glu19 carboxylates (*B*) in the metal-coordination sphere (Fig. 3*b*). Therefore, even though the sulfate/water site represents occupational disorder, the hydrogen-bond connection to the highly ordered metal-coordination site can be considered a stable structural element. The other (independent) sulfate O atom is hydrogen bonded to the N-terminal ammonium group of the other protein molecule (*A*). In this fashion, the sulfate ion provides a bridge between the two independent protein molecules in the coordination sphere.

3.14. Structural comparisons

It is fortuitous that there are two independent copies (*A* and *B*) of the protein molecule in the present structure because it significantly increases the number of available CMTI-I geometries and provides an 'internal standard' (of very high precision) for other structural comparisons. A multiple Ramachandran plot (Kleywegt & Jones, 1996) for molecules *A* and *B* has only a few noticeable residue pairs (not shown) in accord with the level of similarity between the two copies of the protein (see below). The largest conformational difference is observed for Ile6 (φ difference of 39.0°) at the scissile Arg5–Ile6 bond, which is in the area of the highest deviation between the two main chains (see §3.4).

The other CMTI-I models available for comparisons are 1ppe and 2sta, the crystallographic models from the trypsin complex (Bode *et al.*, 1989; Helland *et al.*, 1999), 3cti, an earlier NMR model (Nilges *et al.*, 1991), and 1bxj, a recent NMR model (Zhukov *et al.*, 2000) corresponding (in contrast to the above wild-type models) to the Met8Leu mutant, which is also the subject of the present study. In addition, one can also use the models of the related CPTI inhibitor from the trypsin complexes, 2stb and 2btc (Helland *et al.*, 1999). The coordinates of the first structure of the six 3cti NMR models were used for the comparisons. For meaningful comparisons, one of the six 1bxj structures, the closest to the average structure, was used. The overall structural features of *A*, *B* and 1ppe are virtually the same, except in the two mobile regions, Lys11–Lys12 and Glu24–His25, and Arg5. In *A* and *B*, significant deviations of both main chain and side chains can be seen only in these regions. The results of structural superpositions (including 2sta, 2stb and 2btc) with molecule *B* being the target are shown in Table 3. The situation is similar in 3cti, with somewhat higher deviations seen throughout other regions as well. A significant deviation of 1.7 Å at the N-terminal N atom can be seen in 3cti, while it is very large in the case of 1bxj (5.2 Å). The N-termini in *A*, *B* and 1ppe are virtually the same with deviations of about 0.6 Å (*A*) and 0.9 Å (1ppe). The situation is similar in the 2btc structure (0.8 Å) where the C-terminus–Arg1 salt bridge (see §3.11) is also present, but not in 2sta and 2stb (5.0 and 1.9 Å, respectively), where it is absent. Prominent deviations in the C-terminal regions of the C^{α} traces are seen in the NMR structures, where the C-terminus–Arg1 salt bridge is absent, while the C^{α} trace match is almost perfect in all crystallographic structures (A, *B* 1ppe, 2sta, 2stb, 2btc). In the 1ppe and 2btc trypsin complexes the trimethylene spacer between the C^{α} and N^{ε} atoms of the Arg1 side chain has a conformation that is different with respect to that observed in molecules *A* and *B*.

The solution structure of Met8Leu CMTI-I as determined by NMR (Zhukov *et al.*, 2000) is significantly different from the present crystallographic model. The most important differences include: (i) the conformation of the disulfide bridges, as discussed in §3.9, (ii) the much higher conformational flexibility of the first two N-terminal residues and of the protease-binding segment Pro4–Leu7 and (iii) the reorganization of the hydrophobic cluster formed by the side chains of Ile6, Leu8 and Leu17, leading to less effective packing of the hydrophobic groups and partial loss of hydrogen bonding between the Leu17 amide and water in the 1bxj model.

4. Conclusions

The crystal structure described here provides the first atomic resolution model of a trypsin inhibitor of plant origin. Another advantage is the existence of two independent copies of the molecule. Through their Glu19 carboxylic acids, the molecules coordinate a zinc cation in a tetrahedral and symmetric fashion. Metal coordination has been never described before in this class of proteins. This unique coordination pattern suggests a possibility of controlling protein aggregation through intermolecular coordination of metal ions. The results of these high-resolution refinements confirm the earlier conclusions about the deficiencies of some stereochemical restraints (peptide planarity, N—C^{α}—C angles). The structure suggests that the observed violations of these standards may have one of their sources in hydrogen-bonding interactions. The experience gained during the refinements and model building caution against hasty interpretations of residual

electron densities in the early stages of refinement, as they often tend to change appearance or even disappear as the refinement progresses. Finally, the success of the refinement protocol demonstrates that with high-quality diffraction data to 1 Å resolution it may be possible to relax or even remove some of the stereochemical restraints in the best parts of the model, such as the well ordered main-chain regions.

RT thanks the Mianowski Fund for a fellowship. Thanks are also due to Grzegorz Bujacz for skilled help with data collection. We acknowledge support from the European Community Access to Research Infrastructure Action of the Improving Human Potential Programme to the EMBL Hamburg Outstation, contract number HPRI-CT-1999-00017. Some of the calculations were carried out in the Poznan Metropolitan Supercomputing and Networking Center.

References

Addlagatta, A., Krzywda, S., Czapinska, H., Otlewski, J. & Jaskolski, M. (2001). *Acta Cryst.* D**57**, 649–663.
Allen, F. H. & Kennard, O. (1991). *Chem. Des. Autom. News.* **8**, 31–37.
Ashida, T., Tsunogae, Y., Tanaka, I. & Yamane, T. (1987). *Acta Cryst.* B**43**, 212–218.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bode, W., Greyling, H. J., Huber, R., Otlewski, J. & Wilusz, T. (1989). *FEBS Lett.* **242**, 285–292.
Bolewska, K., Krowarsch, D., Otlewski, J., Jaroszewski, L. & Bierzynski, A. (1995). *FEBS Lett.* **377**, 172–174.
Botos, I., Wu, Z., Lu, W. & Wlodawer, A. (2001). *FEBS Lett.* **509**, 90–94.
Brese, N. E. & O'Keeffe, M. (1991). *Acta Cryst.* B**47**, 192–197.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Cohen, G. H. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Czapinska, H., Otlewski, J., Krzywda, S., Sheldrick, G. M. & Jaskólski, M. (2000). *J. Mol. Biol.* **295**, 1237–1249.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.
Diskin-Posner, Y., Dahal, S. & Goldberg, I. (2000). *J. Chem. Soc. Chem. Commun.*, pp. 585–586.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A. & Mazzarella, L. (2000). *J. Mol. Biol.* **297**, 713–732.
Esposito, L., Vitagliano, L., Zagari, A. & Mazzarella, L. (2000). *Protein Sci.* **9**, 2038–2042.
Helland, R., Berglund, G. I., Otlewski, J., Apostoluk, W., Andersen, O. A., Willassen, N. P. & Smalas, A. O. (1999). *Acta Cryst.* D**55**, 139–148.
Holak, T. A., Gondol, D., Otlewski, J. & Wilusz, T. (1989). *J. Mol. Biol.* **210**, 635–648.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Lesme, V., Blessing, B. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.
Jiang, X., Cao, M., Teppen, B., Newton, S. & Schafer, L. (1995). *J. Phys. Chem.* **99**, 10521–10525.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Karplus, P. A. (1996). *Protein Sci.* **5**, 1406–1420.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Kleywegt, G. J. & Jones, A. T. (1996). *Structure*, **4**, 1395–1400.
Krishnamoorthi, R., Lin, C.-L. S., Gong, Y., VanderVelde, D. & Hahn, K. (1992). *Biochemistry*, **31**, 905–910.
Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.
Leiserowitz, L. (1976). *Acta Cryst.* B**32**, 775–802.
Longhi, S., Czjzek, M. & Cambillau, C. (1998). *Curr. Opin. Struct. Biol.* **8**, 730–737.
MacArthur, M. W. & Thornton, J. M. (1996). *J. Mol. Biol.* **246**, 1180–1195.
Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). *Acta Cryst.* B**39**, 480–490.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–225.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nilges, M., Habazettl, J., Brünger, A. T. & Holak, T. A. (1991). *J. Mol. Biol.* **219**, 499–510.
Otlewski, J., Whatley, H., Polanowski, A. & Wilusz, T. (1987). *Biol. Chem. Hoppe-Seyler*, **369**, 1505–1507.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 306–327.
Polanowski, A., Cieslar, E., Otlewski, J., Nienartowicz, B., Wilimowska-Pelc, A. & Wilusz, T. (1987). *Acta Biochim. Pol.* **34**, 395–406.
Polanowski, A., Otlewski, J., Leluk, J., Wilimowska-Pelc, A. & Wilusz, T. (1988). *Biol. Zentralbl.* **107**, 45–49.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Read, R. J. (1997). *Methods Enzymol.* **277**, 110–128.
Ridder, I. S., Rozeboom, H. J. & Dijkstra, B. W. (1999). *Acta Cryst.* D**55**, 1273–1290.
Sandalova, T., Schneider, G., Kack, H. & Lindqvist, Y. (1999). *Acta Cryst.* D**55**, 610–624.
Sheldrick, G. M. (1997). *The SHELX97 Manual.* University of Göttingen, Göttingen, Germany.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **227**, 319–343.
Teng, T. Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
Uson, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–546.
Wieczorek, M., Otlewski, J., Cook, J., Parks, K., Leluk, J., Wilimowska-Pelc, A., Polanowski, A., Wilusz, T & Laskowski, M. Jr (1985). *Biochem. Biophys. Res. Commun.* **126**, 646–653.
Wilson, K. S., Butterworth, S., Dauter, Z., Lamzin, V. S., Walsh, M., Wodak, S., Pontius, J., Richelle, J., Vaguine, A., Sander, C., Hooft, R. W. W., Vriend, G., Thornton, J. M., Laskowski, R. A. & MacArthur, M. W. (1998). *J. Mol. Biol.* **276**, 417–436.
Wilusz, T., Wieczorek, M., Polanowski, A., Denton, A., Cook, J. I. & Laskowski, M. Jr (1983). *Hoppe Seyler's Z. Physiol. Chem.* **364**, 93–95.
Winkler, F. K. & Dunitz, J. D. (1971). *J. Mol. Biol.* **59**, 159–182.
Zhukov, I., Jaroszewski, L. & Bierzynski, A. (2000). *Protein Sci.* **9**, 273–279.