Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Cryocrystallography of a Kunitz-type serine protease inhibitor: the 90 K structure of winged bean chymotrypsin inhibitor (WCI) at 2.13 Å resolution

S. Ravichandran, U. Sen, C. Chakrabarti and J. K. Dattagupta*

Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, 1/AF Bidhan Nagar, Calcutta 700 064, India

Correspondence e-mail: jiban@cmb2.saha.ernet.in

Received 19 April 1999
Accepted 19 July 1999

PDB Reference: winged bean chymotrypsin inhibitor, 4wbc.

The crystal structure of a Kunitz-type double-headed α-chymotrypsin inhibitor from winged bean seeds has been refined at 2.13 Å resolution using data collected from cryo-cooled (90 K) crystals which belong to the hexagonal space group $P6_122$ with unit-cell parameters $a = b = 60.84$, $c = 207.91$ Å. The volume of the unit cell is reduced by 5.3% on cooling. The refinement converged to an R value of 20.0% ($R_{\rm free}$ = 25.8%) for 11100 unique reflections and the model shows good stereochemistry, with r.m.s. deviations from ideal values for bond lengths and bond angles of 0.011 Å and 1.4°, respectively. The structural architecture of the protein consists of 12 antiparallel β-strands joined in the form of a characteristic β-trefoil fold, with the two reactive-site regions, Asn38–Leu43 and Gln63–Phe68, situated on two external loops. Although the overall protein fold is the same as that of the room-temperature model, some conformational changes are observed in the loop regions and in the side chains of a few surface residues. A total of 176 ordered water molecules and five sulfate ions are included in the model.

1. Introduction

The winged bean chymotrypsin inhibitor (WCI) is a 20 kDa double-headed inhibitor consisting of 183 amino-acid residues. The inhibitor, isolated from the seeds of winged bean (*Psophocarpus tetragonolobus*), belongs to the STI–Kunitz family of serine protease inhibitors and has the property of inhibiting α-chymotrypsin in an unusual 1:2 molar ratio (Kortt, 1980; Shibata *et al.*, 1986). WCI can also be hydrolyzed by α-chymotrypsin at one of its two recognition sites, called the first reactive site; the scissile peptide bond is located between Leu65 and Ser66 (Shibata *et al.*, 1988). Though the location of the other (second) reactive site was not shown biochemically, we have recently identified this site by best-fit structural studies after the three-dimensional structure of WCI was determined at 2.3 Å resolution. Furthermore, docking studies were performed using this structure in order to delineate the inhibitory action of WCI against its cognate enzyme α-chymotrypsin (Dattagupta *et al.*, 1999).

As a continuation of our work on WCI, we wanted to study the structure in more detail, concentrating on the surface-loop regions and on the solvent structure. These are of particular interest because it is known from the earlier three-dimensional model of WCI that more than 50% of the amino-acid residues are present in the loop regions and that the inhibitory function of WCI resides predominantly on the two external loops where the first (Gln63–Phe68) and second (Asn38–Leu43) reactive sites are located. Moreover, water molecules were found to play a major role in stabilizing these surface loops; for example, a water molecule maintained the

Table 1
Data statistics and refinement summary: a comparison between room-temperature (RT) and low-temperature (LT) data.

	LT (90 K)	RT (293 K)†
Resolution (Å)	2.13	2.3
Space group	$P6_122$	$P6_122$
Unit-cell dimensions (Å)	$a = b = 60.84$, $c = 207.91$	$a = b = 61.80$, $c = 212.80$
Unit-cell volume (10^5 Å^3)	6.66	7.03
Number of observations	63933	52764
Number of unique reflections (∞ to maximum resolution)	12963	9897
R_{merge} (%)‡	6.9	9.5
Completeness (%)	95.7	80.1
Wilson B (Å^2)	30.9	34.2
Number of protein atoms in refinement	1400	1427
Number of solvent molecules		
Waters	176	109
Sulfate ions	5	—
Number of reflections used in refinement	11100	8868
Resolution range (Å)	9.0–2.13	8.0–2.3
R factor (%)	20.0	18.7
R_{free} (%)	25.8	24.8
B factor (Å^2)		
All atoms	32.7	40.0
Protein atoms	30.8	39.3
Main-chain atoms	28.0	37.2
Side-chain atoms	33.6	41.4
Water molecules	41.4	46.3
Sulfate ions	70.5	—
R.m.s. deviations from ideality		
Bond lengths (Å)	0.011	0.011
Bond angles (°)	1.4	1.54

† From Dattagupta *et al.* (1999). ‡ $R_{merge} = \sum_{hkl} |I - \langle I \rangle| / \sum |I|$, where I is the observed intensity and $\langle I \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.

conformational rigidity of the first reactive-site loop through hydrogen bonding (Dattagupta *et al.*, 1996; 1999). However, most of the surface-loop residues and solvent water molecules in our earlier model at 2.3 Å resolution (Dattagupta *et al.*, 1999) had high atomic temperature-factor values ($\langle B \rangle$ for the residues in the loop regions was 43.9 Å^2 and for all water molecules was 46.3 Å^2). The data completeness at the higher resolution shell was also low (76% complete between 2.35–2.3 Å shell). An attempt is therefore made to improve the resolution and hence the quality of the WCI data using cryocrystallographic techniques. Here, we present the detailed structural features of WCI at 90 K compared with the earlier room-temperature model.

2. Materials and methods

2.1. Crystallization and data collection at 90 K

WCI was isolated and purified (Roy & Singh, 1986; Shibata *et al.*, 1986) and crystals were grown under the previously reported conditions (Dattagupta *et al.*, 1999). The reservoir solution contained 25%(*w*/*v*) ammonium sulfate in 10 m*M* sodium acetate buffer pH 5.4. For cryo-cooling, the crystals were soaked for 3 min in the same reservoir solution containing 25%(*v*/*v*) glycerol as cryoprotectant (Rodgers, 1994, 1997). A suitable crystal (0.6 × 0.3 × 0.3 mm) was then quickly suspended in a hair loop (Teng, 1990; Sauer & Ceska, 1997) and flash-frozen by rapid insertion in a dry nitrogen gas stream cooled at 90 K using a cryostat. Intensity data were collected using an imaging-plate detector mounted on a rotating-anode X-ray generator operating at 40 kV and 70 mA.

Flash-cooled WCI crystals showed diffraction spots to 2.0 Å resolution, although data were only collected to 2.13 Å resolution owing to spot overlaps and instrumental limitations. The long axis of the loop-mounted crystal was parallel to the rotation axis (along φ) and the image plate was placed at a distance of 165 mm from the crystal. A total of 46 images were recorded at an interval of 1°, with an exposure time of 900 s per frame. The data were indexed and integrated with *DENZO* (Otwinowski, 1993) and scaled using *SCALEPACK* (Otwinowski, 1993). The crystal data and the intensity data statistics are given in Table 1.

2.2. Structure refinement

The 2.3 Å coordinates of the WCI model at room temperature (PDB code 2wbc; Dattagupta *et al.*, 1999) were taken as the starting model for refinement. To avoid model bias, certain portions of this starting model such as (i) the two N-terminal residues Asp1 and Asp2, (ii) the flexible loop region at His137–Asp142, (iii) the C-terminal residues Lys176–His183 and (iv) the water molecules were not included during the initial stages of refinement. Also, all polar residues such as Lys, Arg, Glu, Gln and His were kept as Ala in the initial model. The initial calculations were carried out using the *X-PLOR* (version 3.851) program package (Brünger, 1992*a*). The position of the molecule was first adjusted by a few cycles of rigid-body refinement, using all data to 2.3 Å resolution, because of the slight differences between the unit-cell parameters of the room-temperature (RT) and the present low-temperature (LT) model. Next, the resolution was gradually increased to 2.13 Å; the R factor fell to 36.9%.

The model was restrained using standard *X-PLOR* topology files and Engh and Huber stereochemical parameters (Engh & Huber, 1991). It was then subjected to the simulated-annealing procedure using the slow-cooling protocol of *X-PLOR* (Brünger *et al.*, 1990) with a starting temperature of 3000 K, followed by several cycles of positional and individual B-factor refinement. In order to monitor the progress of the refinement, a test set of 10% of the reflections were selected randomly for the calculation of R_{free} (Brünger, 1992*b*). The R factor at this stage was 29.8% (R_{free} = 35.4%). After calculating ($2F_o - F_c$, α_c) and ($F_o - F_c$, α_c) maps, model building was performed using the graphics program *TOM* (Jones, 1985) implemented on a Silicon Graphics Indigo^2 workstation. During this step, the as yet unlocated long side chains and almost all the excluded portions of the initial model could be traced and rebuilt. Water molecules were gradually included in the model and a bulk-solvent correction was applied throughout the course of the refinement. However, no interpretable electron-density map was seen for the four C-term-

inal residues (Ala180–His183) and these were therefore excluded from the next model for refinement. At this stage, the R factor had a value of 23.1% ($R_{\rm free}$ = 28.8%).

In subsequent stages, the refinement was performed with maximum-likelihood functions using the program *REFMAC* (Murshudov *et al.*, 1997) in the *CCP*4 program package (Collaborative Computational Project, Number 4, 1994). During this step, the *X-PLOR*-corrected solvent amplitudes and phases (FPART and PHIPART) were used and the ($2m|F_o| - D|F_c|$, α_c) and ($m|F_o| - D|F_c|$, α_c) maps were calculated. Model building was continued and, at this stage, most of the ambiguous portions in the model such as the solvent-exposed side chains were rebuilt. More water molecules were then located using the program *ARP* (Lamzin & Wilson, 1993) from the *CCP*4 program suite. They were added to the coordinate list according to the following criteria: (i) the water molecule must make at least one stereochemically reasonable hydrogen bond, (ii) it should have well defined density above 1σ in a $2m|F_o| - D|F_c|$ map and (iii) it should have density above 3σ in a $m|F_o| - D|F_c|$ map. During this stage of refinement, the presence of sulfate anions was first indicated in the electron-density map and a total of five prominent peaks in the difference electron-density map were modelled as sulfate anions. They were identified on the basis of tetrahedral density shape and were confirmed by the presence of positive difference peaks within approximately 1.5 Å of the O atom if a water molecule was refined in place of the sulfate ion. The location of each of these sulfate ions was also supported by the surrounding chemical environment, in which basic amino-acid residues such as His (23, 27, 124#), Lys (131, 108#) and Arg (47, 158) were present.

A final OMIT map (Bhat, 1988) was calculated using the modified σ_A coefficients (Vellieux & Dijkstra, 1997) and the positions of all the solvent molecules were re-confirmed. Water molecules with temperature factors greater than 70 Å^2 were omitted from the structure. Even at this stage, the four missing residues (Ala180–His183) at the C-terminal region could not be seen in the OMIT electron-density maps and therefore could not be included in the final model. The final crystallographic R factor is 20.0% ($R_{\rm free}$ = 25.8%) for 11 100 unique reflections between 9.0 and 2.13 Å. Fig. 1 shows a representative portion of the electron-density map calculated using $2m|F_o| - D|F_c|$ coefficients and the fitting of the final model in the map. Refinement statistics are compared with those of the room-temperature model and are shown in Table 1. The atomic coordinates and the structure factors have been deposited with the Protein Data Bank (Bernstein *et al.*, 1977).

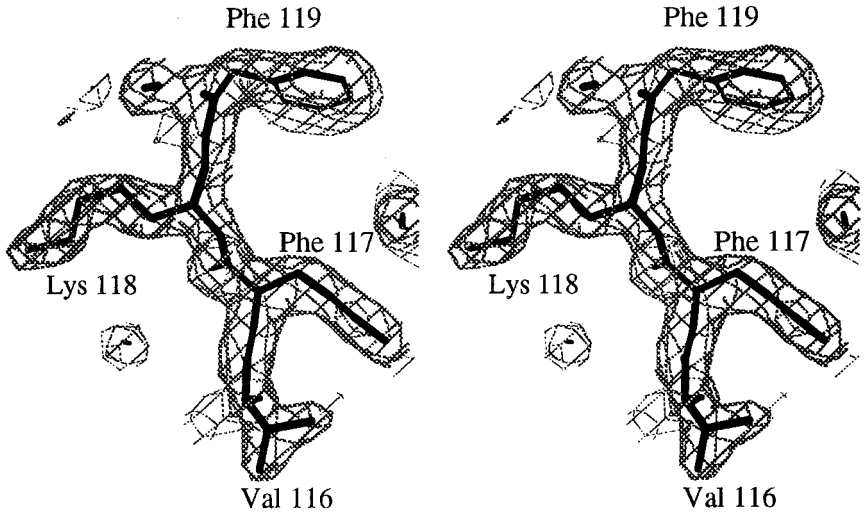

Figure 1
A region of the final $2m|F_{\rm obs}| - D|F_{\rm calc}|$ electron-density map (contoured at 1.5σ) and the corresponding refined model. The diagram was produced using the program *O* (Jones *et al.*, 1991).

3. Results and discussion

3.1. Quality of the model

The final model with 1400 non-H protein atoms, 176 solvent water molecules and five sulfate ions shows good stereochemistry. The main-chain dihedral angles (φ, ψ) are within the allowed limits (Morris *et al.*, 1992) in the Ramachandran plot (Ramachandran *et al.*, 1963; Fig. 2) and the results from *PROCHECK* (Laskowski *et al.*, 1993) show that around 134 residues (88.7%) are in the core region, 16 residues (10.6%) are in the allowed region and only one residue (0.6%) is in the generously allowed region of the plot.

The surface residues at His137–Asp141 and the C-terminal residues (Lys176–Thr179) are disordered and the corresponding regions were also reported as disordered in structurally similar proteins such as *Erythrina caffra* trypsin inhibitor (ETI) (Onesti *et al.*, 1989, 1991), soybean trypsin inhibitor (STI) (Meester *et al.*, 1998), STI–trypsin complex (Blow *et al.*, 1974; Sweet *et al.*, 1974; Song & Suh, 1998), interleukin-1α (Graves *et al.*, 1990) and interleukin-1β (Finzel *et al.*, 1989; Priestle *et al.*, 1989). The average B factors for main-chain, side-chain and all protein atoms are 28.0, 33.6 and 30.8 Å^2, respectively, and for the disordered regions $\langle B \rangle$ is above 80 Å^2. Fig. 3 shows the relation between the average B value for the main-chain atoms and the residue number, which is compared with the corresponding room-temperature values.

3.2. Overall structure and comparison with similar structures

The structure of WCI is composed of 12 antiparallel β-strands connected by irregular loops. Six of the strands form a 'β-barrel' type of arrangement and the remaining six cover one hollow end of the barrel in the form of a 'lid' (Fig. 4). Such a topology, termed the 'β-trefoil' fold (McLachlan, 1979; Murzin *et al.*, 1992), is not only observed in the STI–Kunitz family

of proteins but is also seen in some functionally unrelated proteins (Dattagupta *et al.*, 1999; Song & Suh, 1998; Umland *et al.*, 1997). The common functional features among these non-homologous proteins is that they are involved in recognition and binding. However, it has been suggested by phylogenetic studies that the cause for the structural conservation of the β-trefoil fold among these unrelated proteins may be a consequence of a possible double duplication of the gene in the superfamily of the protein (Mukhopadhyay, 1999).

3.3. Comparison with the WCI model at 293 K

As expected, the overall fold of WCI remains essentially the same as that of the RT model (Dattagupta *et al.*, 1999). Fig. 5 shows the C^{α} superposition of the two models. The r.m.s deviations for C^{α}, main-chain, side-chain and all atoms were calculated using the program *LSQKAB* from the *CCP*4 program suite. For the 171 residues superposed, the average shift in positions for the main-chain, side-chain and all atoms are 0.42, 1.06 and 0.73 Å, respectively. Major changes are seen only in the two loop regions, His137–Glu141 (average r.m.s. shift is 4.0 Å) and Lys176–Thr179 (average r.m.s. shift is 8.2 Å), where the model shows disorder. Also, it is interesting to note that the C-terminal region (Lys176–His183) in the RT model had a different chain direction, extending towards a hydrophilic pocket. Such differences in atomic positions may arise from either large mobility of the chain or static disorder in the crystal (Priestle *et al.*, 1989), which cannot be eliminated even at low temperatures. The functional role played by these terminal residues is still unknown and they also seem to participate little in crystal packing interactions. The unit-cell volume is reduced by 5.3% because of cooling; the contraction seems to arise from slight changes in the position of surface loops and side chains and the reduction in the thermal movement of water molecules on cooling when the molecule packs in the unit-cell (Hartmann *et al.*, 1982; Teeter *et al.*, 1993).

3.3.1. β-turns. The loop regions on the surface of the molecule are predominantly defined by the secondary structure such as β-turns (Richardson, 1981; Wilmot & Thornton, 1988). There are a total of 18 β-turns observed in the structure, compared with 16 found in the RT model, according to the hydrogen-bonding pattern $O_i\cdots N_{i+3}$ (Venkatachalam, 1968). The turns Ser125–His128 and Ala175–Glu178 were not located in the earlier RT model. Most of the residues in the β-turn have lower B values ($\langle B\rangle$ = 40.3 Å^2) compared with

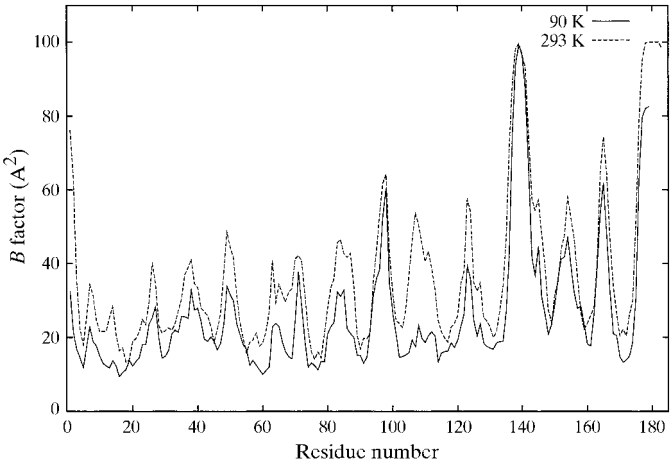

Figure 3
Plot of average temperature factors as a function of residue number, shown for main-chain atoms of the LT (solid line) and RT (dotted line) models.

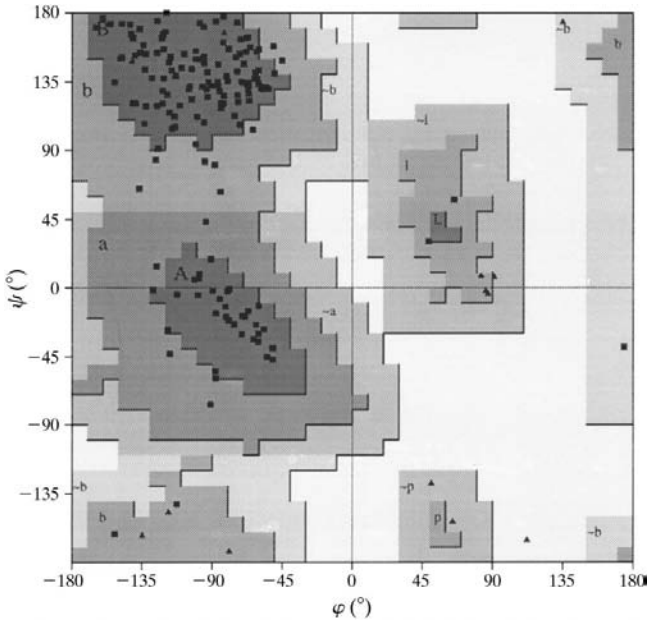

Figure 2
Ramachandran plot of the main-chain dihedral angles (φ, ψ) for the final model generated using the program *PROCHECK* (Laskowski *et al.*, 1993). Glycine residues are indicated by triangles and non-glycine residues by squares.

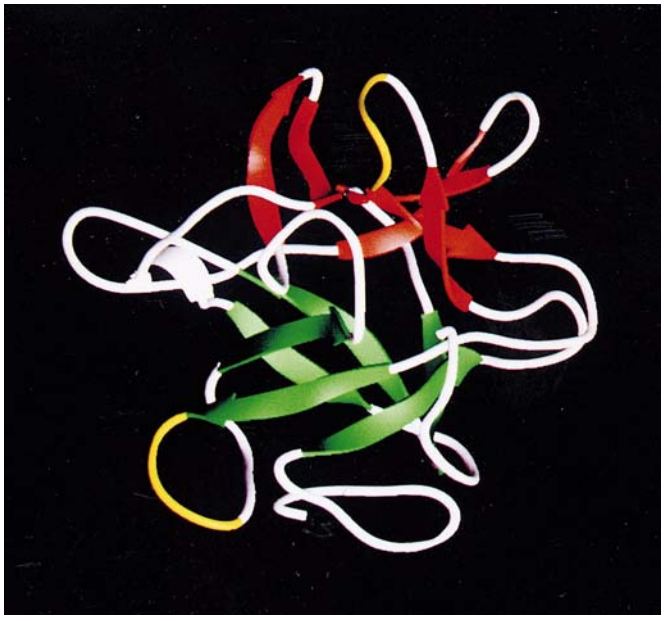

Figure 4
Ribbon representation of the WCI structure, showing the barrel in green and lid in red. The two reactive-site loops are shown in yellow.

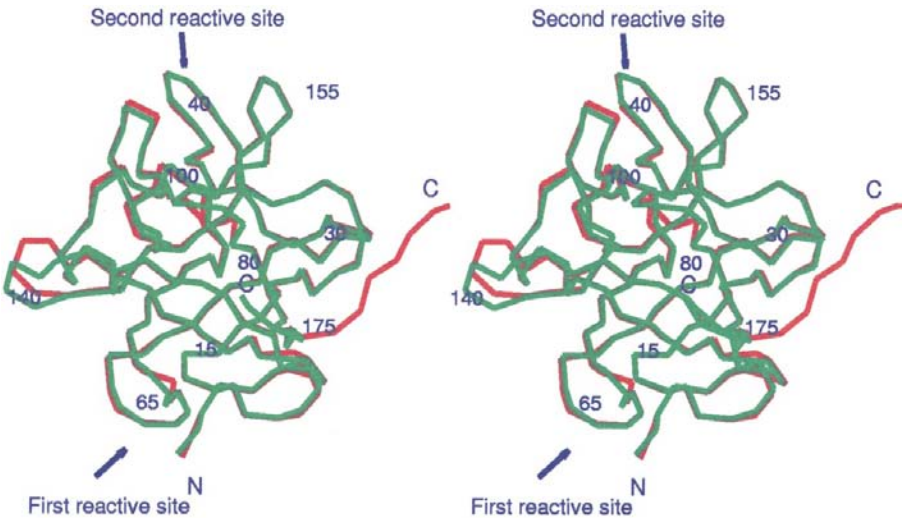

Figure 5
Stereoview of least-squares C^{α} superposition of the low-temperature (LT; shown in green) and room-temperature (RT; shown in red) models of WCI.

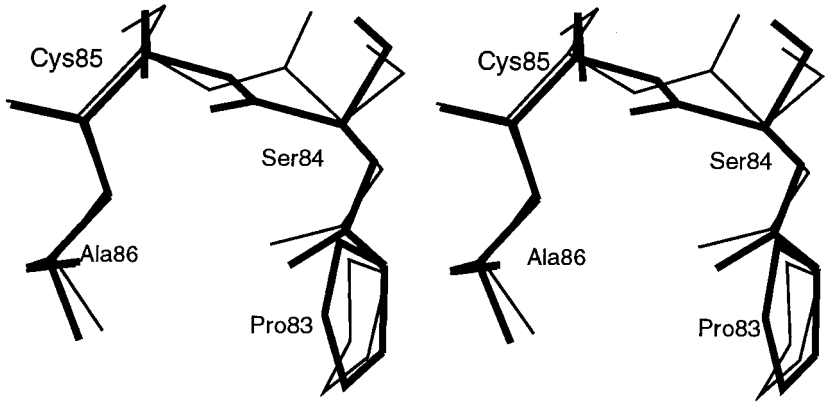

Figure 6
The β-turn at Pro83–Ala86 showing the 'peptide flip'. The conformation at residues 83–86 is shown at low temperature (LT; heavy lines) and room temperature (RT; light lines).

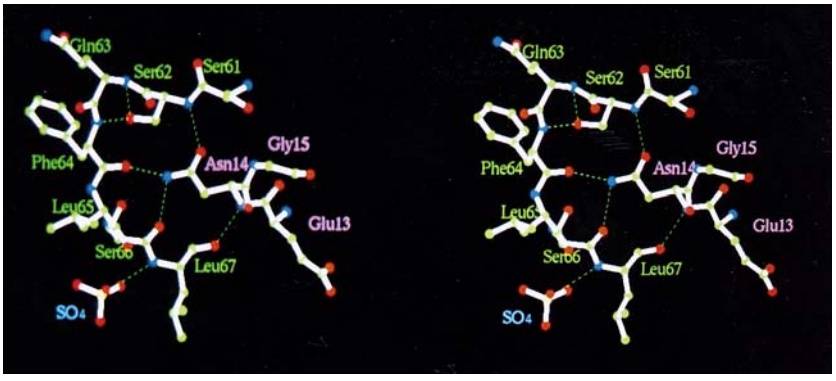

Figure 7
A stereoview of the first reactive-site loop (Ser61–Leu67; labelled in green) of WCI. The scissile bond is between Leu65 and Ser66. The strong hydrogen-bonding network between the main-chain atoms of the first reactive-site loop and the side-chain atoms of Asn14 (labelled in pink) is important in maintaining the correct conformation of the loop.

those of the RT model ($\langle B \rangle$ = 47.7 Å^2). Although the β-turns observed in the disordered region such as His137–Glu140 and Asp138–Asp141 show deviations in positions, the values of (φ_{i+1}, ψ_{i+1}) and (φ_{i+2}, ψ_{i+2}) corresponds to the same (type IV) conformation in both the models.

The most significant change in conformation is noticed in the type I β-turn at Pro83–Ala86 (φ_{84} = −63.5, ψ_{84} = −33.7° and φ_{85} = −70.8, ψ_{85} = −18.3°). In the RT structure, this was a type IV β-turn (φ_{84} = −71.6, ψ_{84} = 102.6° and φ_{85} = 152.6, ψ_{85} = −16.6°). Thus, the peptide bond between Ser84 and Cys85 has experienced a 'peptide flip' owing to the change in temperature. Fig. 6 shows the superposition of the conformations in the LT and RT models. The interconversion could be a consequence of the free-energy minima being close for the two conformations and the difference in molecular packing between the two models causing one of the conformations to freeze out (Fields *et al.*, 1994; Bartunik *et al.*, 1989). The peptide flip has occurred on cooling despite the presence of a disulfide bond (Cys41–Cys85) nearby, which would be expected to give supporting rigidity to the β-turn. The functional role played by the residues in this region is unknown for either WCI or ETI (Onesti *et al.*, 1989, 1991). However, it is interesting to note that in the structure of STI (Song & Suh, 1998), the corresponding region (Met84–Cys86), because of the three-residue insertions, has a 3_{10}-helix and the bond after Met84 is susceptible to cleavage by subtilisin (Laskowski *et al.*, 1974).

Apart from the β-turns, there are two 3_{10}-helices and one inverse γ-turn, which were also located in the previous RT model. Out of the five β-bulges, one antiparallel classic β-bulge involving residues Val75, Val116 and Phe117 has been located in this LT model.

3.3.2. Reactive-site loops. The two reactive sites of WCI are located sufficiently distant from each other on two external loops and these are defined by the characteristic canonical conformation (Bode & Huber, 1992).

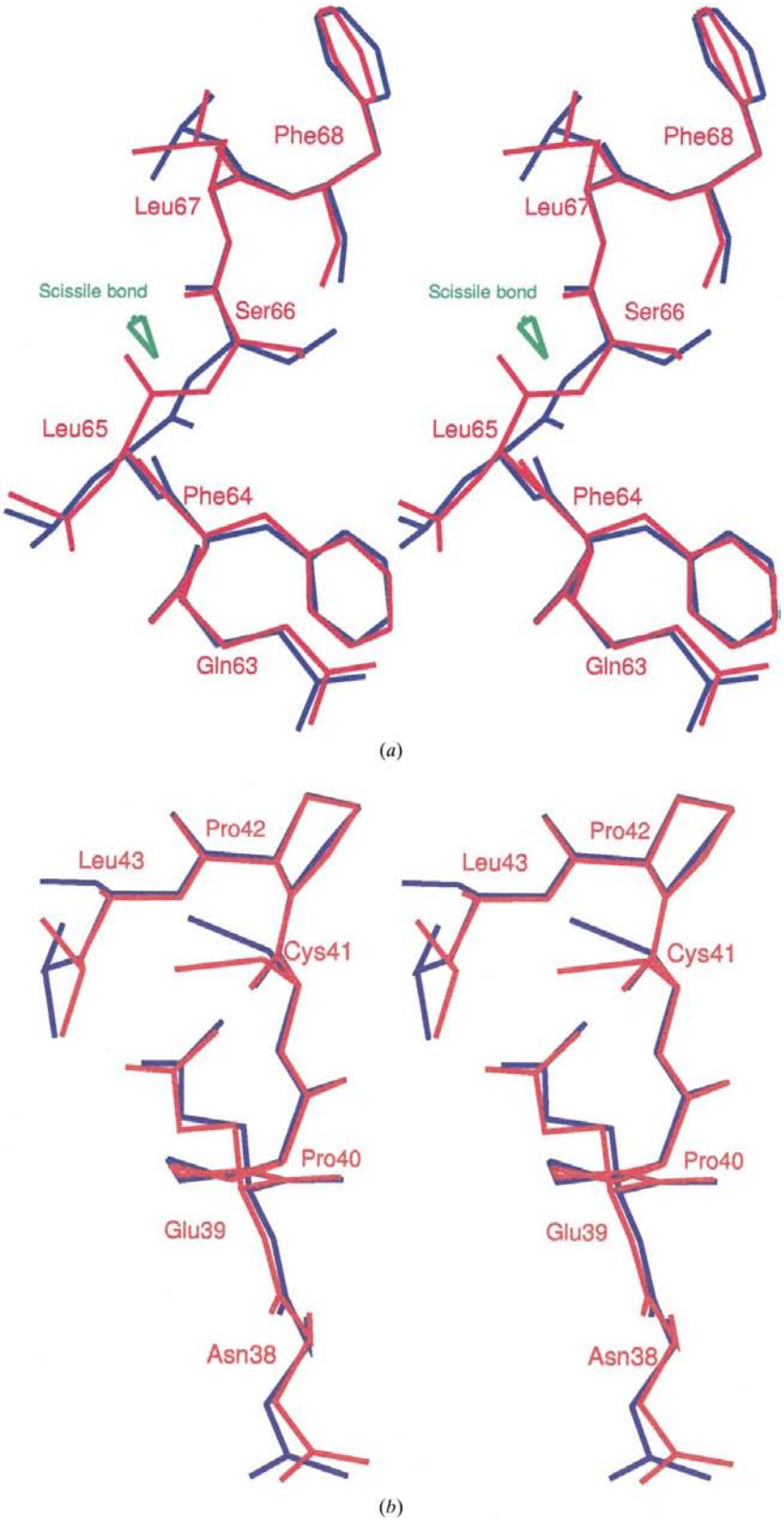

Figure 8
(*a*) Superposition of the first reactive-site loop in the LT model (shown in red) on that of the RT model (shown in blue). (*b*) Superposition of the second reactive-site loop (shown in red) on that of the RT model (shown in blue).

As in the RT model, the side chain of the highly conserved Asn14 residue intrudes inside the first reactive-site loop (Gln63–Phe68) and makes a strong hydrogen-bonding network with the main-chain atoms of the residues present in the loop (Fig. 7). However, these hydrogen-bonding interactions are stronger in the present LT model, as the bonds concerned have shorter lengths. These include bonding of Ser62 N with Asn14 $O^{\delta1}$, Gln63 N with Ser62 O^{γ}, Phe64 N with Ser62 O^{γ} and Leu65 N with Glu111# $O^{\varepsilon2}$; these have hydrogen-bond distances of 2.75, 2.73, 2.84 and 2.78 Å, respectively, in the LT model compared with 3.02, 3.02, 3.15 and 3.57 Å, respectively, in the RT model.

The P1 residue (Leu65) at the first reactive-site loop plays a vital role during the inhibitor–enzyme complex formation. This was clear from our earlier docking studies on the WCI–α-chymotrypsin complex model (Dattagupta *et al.*, 1999) and from the crystal structure of the STI–porcine pancreatic trypsin complex (Song & Suh, 1998), where the geometry of the carbonyl group of the P1 residue of the inhibitor was found to be important in understanding the interaction between the inhibitor and the protease during the catalytic mechanism. In this regard, when superposing the two models of WCI at LT and RT, the carbonyl O atom of Leu65 in the LT model shows significant deviation (∼2.5 Å) in its position (Fig. 8*a*). This is also reflected by a change in the (φ, ψ) value of the P1 residue; $\varphi_{65} = -93.3$, $\psi_{65} = 84.6^\circ$ in the RT model and $\varphi_{65} = -87.6$, $\psi_{65} = -16.6^\circ$ in the LT model. Moreover, two sulfate ions (Sul191 and Sul193, separated by 5.0 Å), which could not be identified in the earlier models, are located on the external surface of the reactive-site loop in the LT model. They are held within a distance of 4 Å by the reactive-site loop amide N atoms (Leu65 N, Ser66 N, Leu67 N and Phe68 N), the side-chain atoms Lys108# N^{δ} and Trp90# $N^{\varepsilon1}$. The location of the sulfate ions in the model is dealt with separately in §3.3.6.

In the case of the second reactive-site loop, Asn38–Leu43, there is no change in conformation with respect to the RT model (Fig. 8*b*). This may be because of the presence of the disulfide bridge between Cys41–Cys85, which limits the conformational freedom for the loop. Moreover, the presence of the two proline residues, Pro40 and Pro42, within the loop has reduced the possibility of having many solvent partners. However, the position

of the S^{γ} atom in Cys41 shows a shift (~1.1 Å), possibly owing to the peptide flip which has occurred near Cys85.

3.3.3. Side chains. At low temperature, most of the solvent-exposed surface side chains are well ordered and clearly defined in the electron-density map. In particular, the polar and long side chains are reoriented in a definite direction by hydrogen bonds to water molecules or with other surface residues. The side chains of Asp1, Asp2, Asn50, Arg59, Arg71, Gln98, Lys108, Glu165 and Lys174 (all solvent-exposed) show significant deviation in positions (r.m.s. shift between 2.0 and 7.0 Å) compared with the analogous regions in the RT model. Fig. 9(*a*) shows the reorientation of the Lys108 side chain in the LT model. In the RT model, the side chain extended outside towards the solvent region, whereas in the LT model the side chain curls back close to a surface turn at Ala86–Pro89 and forms a hydrogen bond with water W226 (d = 3.12 Å) and the carbonyl O atom of Ala87 (d = 2.81 Å).

Similarly, the side chain of Lys174 shows conformational differences (Fig. 9*b*). In the RT model, this side chain is involved in a salt bridge with Glu13, whereas the LT model clearly shows a different location of the side chain where it folds back and form hydrogen bonds with water W230 and the carbonyl O atom of Thr17, with hydrogen-bonding distances of 2.50 and 2.86 Å, respectively. Thus, the salt bridge (Glu13–Lys174) is no longer present and the $i \rightarrow i + 3$ hydrogen-bonding distance between the carbonyl O atom of Glu13 and the amide N atom of Gly16 is shortened (d = 3.23 Å in the RT and 2.85 Å in the LT models).

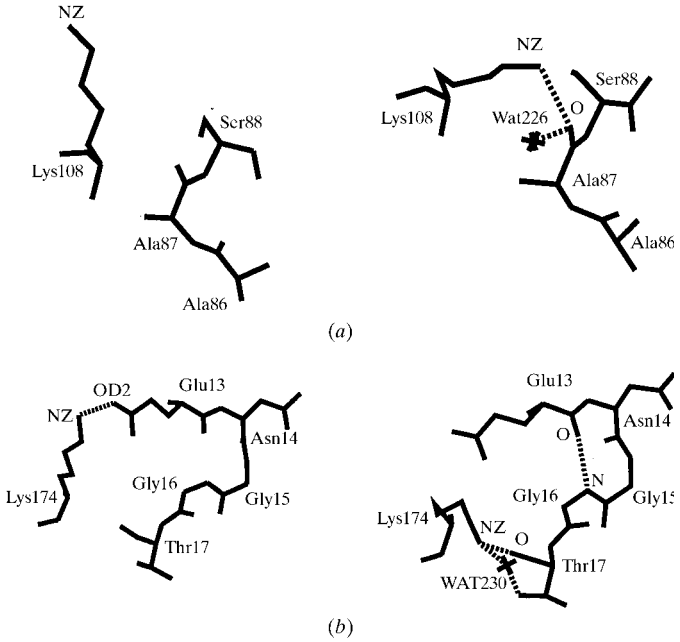

Figure 9
Conformational differences observed for the side chains of (*a*) Lys108 and (*b*) Lys174 at room temperature (RT; left) and low temperature (LT; right), respectively.

3.3.4. Salt bridges. There are 23 acidic residues (Asp and Glu) and 25 basic residues (Arg, Lys and His) in total and, as expected, a majority of them lie on the protein surface and interact with the solvent molecules. The number of salt bridges (ten) in the present model is greater than the number found (five) in the RT model. All residues involved in the salt bridges are well defined in the electron-density map, and the average hydrogen-bonding distance between the donor and the acceptor atoms is 3.0 Å. Only three of the salt bridges are found to be conserved in the structures of ETI, STI and winged bean albumin (WBA-1) (McCoy & Kortt, 1997) of the Kunitz–STI family.

3.3.5. Water structure. The solvent structure is well ordered in the current model. The solvent content calculated using the Matthews coefficient (Matthews, 1968) is 54%, and a total of 176 water molecules are identified in the LT model compared with 109 in the RT model. The average B factor of all the water molecules is 41.4 $Å^2$. Three water molecules (W226, W278 and W290) are found in special positions and each of these is assigned an occupancy factor of 0.5, as all the three lie on a crystallographic twofold axis. Using the program *CONTACT* of the *CCP*4 suite, the total number of water–protein and water–water contacts are identified in order to find the water distribution in different hydration shells. The hydrogen-bonding distance cutoff used is $2.3 < d < 3.6$ Å and the shortest water-to-water distance is 2.35 Å. Only a few water molecules are found buried in deep cavities; the majority are found on the surface of the molecule. A total of 138 water molecules have at least one contact with the protein and are therefore in the first hydration shell; 35 waters have only water–water contacts and are therefore in the second hydration shell. Only three water molecules (W287, W376 and W353) have no potential hydrogen-bonding contact partners within the 3.6 Å limit.

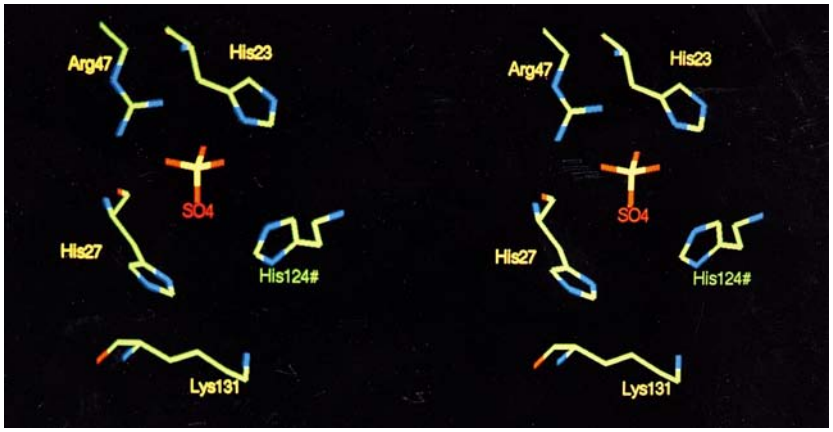

Figure 10
A stereoview of a sulfate ion (Sul195) identified on the protein surface where the side chains of basic residues are found.

When compared with the RT model, only 27 water molecules are found to be conserved in the present model. The selection of these conserved water molecules is made using the criterion that they lie within 1.0 Å of one another and form hydrogen bonds with the same protein or water molecule when the C^{α}

atoms of the two models are superposed by least-squares fitting. Many of these conserved water molecules are found in strategic positions where β-strands diverge away from one another or between or inside loops of the polypeptide on the protein surface. Thus, these water molecules probably help in defining the surface structure.

3.3.6. Sulfate ions. As the crystals of WCI were grown from ammonium sulfate solutions, the sulfate anions are likely to be found specifically bound to the exposed surface of the protein where the region has basic residues. All the five sulfate ions located in the structure are found on the surface of the protein. Of these, one lies on a special position, *i.e.* on a crystallographic twofold axis (and hence are half-occupied), and one is disordered and is hence modelled in two positions. The average B values for Sul191, Sul192 are 54 and 60 Å^2, respectively; for the remaining sulfates (Sul193, Sul194 and Sul195) B is above 70 Å^2.

Two sulfate ions (Sul191 and Sul193) are found close to the first reactive-site loop and together with water molecules (W224 and W264) are involved in a strong hydrogen-bonding network with the side chains of Asp1 and Asp3. Thus, the N-terminal residues are brought close to the first reactive-site loop through solvent interactions. Another sulfate ion, Sul195, makes hydrogen bonds with the N atoms of the guanidium group of Arg47, the amide N atom of Glu169, the water W213 and the N$^{\delta1}$ atom of the side chain of His27. Near this sulfate ion, side chains of a few basic residues (His23, His27, His124#, Arg47, Arg158 and Lys131) are found clustered on the surface (Fig. 10) within a sphere of 8 Å radius, thus making the region positively charged.

This work was supported by a grant (BT/PRO0139/R&D/15/11/96) from the Department of Biotechnology, Government of India. We are thankful to Dr M. Singh and his group at the Indian Institute of Chemical Biology, Calcutta, India for making WCI available to us and to Mr D. Mukhopadhyay and Dr A. Podder of our laboratory for many useful discussions. Thanks are also due to Dr F. M. D. Vellieux at the Institut de Biologie Structurale, CNRS, Grenoble, France for his advice on OMIT map calculations.

References

Bartunik, H. D., Summers, L. J. & Bartsch, H. H. (1989). *J. Mol. Biol.* **210**, 813–828.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Bhat, T. N. (1988). *J. Appl. Cryst.* **21**, 279–281.

Blow, D. M., Janin, J. & Sweet, R. M. (1974). *Nature (London)*, **249**, 54–57.

Bode, W. & Huber, R. (1992). *Eur. J. Biochem.* **204**, 433–451.

Brünger, A. T., Krukowski, A. & Erickson, J. (1990). *Acta Cryst.* A**46**, 585–593.

Brünger, A. T. (1992*a*). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, Connecticut, USA.

Brünger, A. T. (1992*b*). *Nature (London)*, **355**, 472–475.

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

Dattagupta, J. K., Podder, A., Chakrabarti, C., Sen, U., Dutta, S. K. & Singh, M. (1996). *Acta Cryst.* D**52**, 521–528.

Dattagupta, J. K., Podder, A., Chakrabarti, C., Sen, U., Mukhopadhyay, D., Dutta, S. K. & Singh, M. (1999). *Proteins Struct. Funct. Genet.* **35**, 321–331.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Fields, B. A., Bartsch, H. D., Bartunik, H. D., Cordes, F., Guss, J. M. & Freeman, H. C. (1994). *Acta Cryst.* D**50**, 709–730.

Finzel, B. C., Clany, L. L., Holland, D. R., Muchmore, S. W., Waterpaugh, K. D. & Einspahr, H. M. (1989). *J. Mol. Biol.* **209**, 779–791.

Graves, B. J., Hatada, M. H., Hendrickson, W. A., Miller, J. K., Madison, V. S. & Satow, Y. (1990). *Biochemistry*, **29**, 2679–2684.

Hartmann, H., Parak, F., Steigemann, W., Petsko, G., Ponzi, D. & Frauenfelder, H. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 4967–4971.

Jones, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

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

Kortt, A. A. (1980). *Biochim. Biophys. Acta*, **624**, 237–248.

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

Laskowski, M. Jr, Kato, I., Leary, T. R., Schrode, J. & Sealock, R. W. (1974). *Proteinase Inhibitors: Bayer Symposium V*, edited by H. Fritz, H. Tschesche, L. J. Greene & E. Truscheit, pp. 597–611. Berlin: Springer.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

McCoy, A. J. & Kortt, A. A. (1997). *J. Mol. Biol.* **269**, 881–891.

McLachlan, A. D. (1979). *J. Mol. Biol.* **133**, 557–563.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Meester, P. de, Brick, P., Lloyd, L. F., Blow, M. D. & Onesti, S. (1998). *Acta Cryst.* D**54**, 589–597.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., Thornton, J. M. (1992). *Proteins*, **12**, 345–364.

Mukhopadhyay, D. (1999). Personal communication.

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

Murzin, A. G., Lesk, A. M. & Chothia, C. H. (1992). *J. Mol. Biol.* **223**, 531–543.

Onesti, S., Brick, P. & Blow, D. M. (1991). *J. Mol. Biol.* **217**, 153–176.

Onesti, S., Lloyd, L. F., Brick, P. & Blow, D. M. (1989). *J. Mol. Biol.* **210**, 241–242.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Priestle, J. P., Schär, H. P. & Grutter, M. G. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 9667–9671.

Ramachandran, G. N., Ramakrishnan, G. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.

Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167–339.

Rodgers, D. W. (1994). *Structure*, **2**, 1135–1140.

Rodgers, D. W. (1997). *Methods Enzymol.* **276**, 183–203.

Roy, A. & Singh, M. (1986). *Phytochemistry*, **25**, 595–600.

Sauer, U. H. & Ceska, T. A. (1997). *J. Appl. Cryst.* **30**, 71–72.

Shibata, H., Hara, S. & Ikenaka, T. (1988). *J. Biochem.* **104**, 537–543.

Shibata, H., Hara, S., Ikenaka, T. & Abe, J. (1986). *J. Biochem.* **99**, 1147–1155.

Song, H. K. & Suh, S. W. (1998). *J. Mol. Biol.* **275**, 347–363.

Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974). *Biochemistry*, **13**, 4212–4228.

Teeter, M. M., Roe, S. M. & Heo, N. H. (1993). *J. Mol. Biol.* **230**, 292–311.

Teng, T. Y. (1990). *J. Appl. Cryst.* **23**, 387–391.

Umland, T. C., Wingert, L. M., Swaminathan, S., Furey, W. F., Schmidt, J. J. & Sax, M. (1997). *Nature Struct. Biol.* **4**, 788–792.

Vellieux, F. M. D. & Dijkstra, B. W. (1997). *J. Appl. Cryst.* **30**, 396–399.

Venkatachalam, C. M. (1968). *Biopolymers*, **6**, 1425–1436.

Wilmot, C. M. & Thornton, J. M. (1988). *J. Mol. Biol.* **203**, 221–232.