PROTEINS: Structure, Function, and Genetics 35:321–331 (1999)

Refined Crystal Structure (2.3 Å) of a Double-Headed Winged Bean α-Chymotrypsin Inhibitor and Location of Its Second Reactive Site

Jiban K. Dattagupta,[1]* Aloka Podder,[1] Chandana Chakrabarti,[1] Udayaditya Sen,[1] Debashis Mukhopadhyay,[1] Samir K. Dutta,[2] and Manoranjan Singh[2]

[1]*Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, Calcutta, India*

[2]*Indian Institute of Chemical Biology, Jadavpur, Calcutta, India*

ABSTRACT **The crystal structure of a double-headed α-chymotrypsin inhibitor, WCI, from winged bean seeds has now been refined at 2.3 Å resolution to an R-factor of 18.7% for 9,897 reflections. The crystals belong to the hexagonal space group $P6_122$ with cell parameters a = b = 61.8 Å and c = 212.8 Å. The final model has a good stereochemistry and a root mean square deviation of 0.011 Å and 1.14° from ideality for bond length and bond angles, respectively. A total of 109 ordered solvent molecules were localized in the structure. This improved structure at 2.3 Å led to an understanding of the mechanism of inhibition of the protein against α-chymotrypsin. An analysis of this higher resolution structure also helped us to predict the location of the second reactive site of the protein, about which no previous biochemical information was available. The inhibitor structure is spherical and has twelve antiparallel β-strands with connecting loops arranged in a characteristic β-trefoil fold common to other homologous serine protease inhibitors in the Kunitz (STI) family as well as to some non homologous functionally unrelated proteins. A wide variation in the surface loop regions is seen in the latter ones. Proteins 1999;35:321–331.** © 1999 Wiley-Liss, Inc.

Key words: winged bean seeds; chymotrypsin inhibitor; Kunitz (STI) family; second reactive site; serine protease inhibitor

INTRODUCTION

The serine protease family is one of the most thoroughly studied classes of proteolytic enzymes that are involved in numerous physiological processes. As is well known, the active site of a serine protease is characterized by a catalytic triad comprising of Ser195, His57, and Asp102 according to chymotrypsin numbering. In case of chymotrypsin there is a serine residue (Ser189) at the bottom of the binding pocket that is specific for hydrophobic amino acids.[1] The inhibitor proteins to these proteases outclass them in variety, source, diversity, and number.[2,3] The almost universal presence of these inhibitors in *Leguminosae* plants suggests their possible role in protecting the plant tissue at the colonized site of the symbiotic bacteria against the action of the bacterial proteases.[4] These serine protease inhibitors are also found in plenty in legume seeds.

One of the important families[2,5] of the serine protease inhibitors is the Kunitz (STI) family. Albeit a more or less thorough analysis of the biochemistry of this family of proteins, the three-dimensional X-ray structures are available for only five of them: soybean trypsin inhibitor (STI) and its complex with porcine trypsin,[6,7] *Erythrina caffra* trypsin inhibitor (ETI),[8] proteinase K inhibitor (PKI3) from wheat germ,[9] winged bean chymotrypsin inhibitor (WCI),[10] and a seed storage protein, winged bean albumin (WBA-1).[11] WCI is the only member in this family which inhibits α-chymotrypsin.[12]

The winged bean (*Psophocarpus tetragonolobus*) is a legume crop of the tropical regions and contains as much protein and oil in its seed as the soybean.[13] These legume seeds of winged bean have been extensively studied not only to develop it as a food source but also to consider these seeds of high nutritive value as a possible substitute for soybean in the humid tropical countries. WCI was isolated from the seeds of winged bean and it was found to have the property of inhibiting α-chymotrypsin in an unusual 1:2 ratio[14] indicating that this inhibitor protein is of double-headed nature. The inhibitor, WCI, is a single-chain polypeptide of non-glycoprotein nature with a pI value of 5.5. It contains 183 amino acid residues and has a molecular mass of 20,244 daltons.[15] The amino acid sequence of the protein was also fully determined and the scissile bond Leu65–Ser66 at the first reactive site was indicated biochemically by Shibata et al. in 1988.[15] However, no information was available about the location of the second reactive site, a unique feature for an inhibitor in the Kunitz (STI) family. Further, no cleavage was reported at the second site and a lower binding constant of 7.75×10^5 M^{-1} (compared to 4.2×10^7 M^{-1} for the first site) was observed for this site from biochemical studies. We have earlier reported the X-ray structure of WCI at 2.95 Å resolution.[10] The β-trefoil fold[16] type structure of WCI, comprising of

Grant sponsor: Department of Biotechnology, Government of India; Grant Number: BT/R&D/15/15/92.

*Correspondence to: Prof. J.K. Dattagupta, Saha Institute of Nuclear Physics, 1/AF Bidhan Nagar, Calcutta 700 064, India. E-mail: jiban@cmb2.saha.ernet.in

Received 25 August 1998; Accepted 14 December 1998

twelve anti-parallel β-strands, resembles other structures in this family of proteins. The first reactive site was located on an external, protruding loop (Gln63–Phe68). An interesting feature observed in WCI and in proteins of similar structure, is the presence of a pseudo three-fold symmetry axis parallel to the barrel axis which passes through the center of the molecule.[16,17]

In our published 2.95 Å structure, some of the amino acid residues could not be fitted properly in the electron density map and were instead kept as alanines and two residues (Thr179, Ala180) of the flexible C-terminal region could not be located. The second reactive site of this double-headed protein also could not be identified in the structure due to ill-defined electron densities at some of the surface loop regions. Moreover, only 56 water molecules could be located in that structure. These deficiencies necessitated a higher resolution structure.

MATERIALS AND METHODS

Crystallization and Data Collection

Standard column chromatographic techniques have been used to isolate the protease inhibitor.[18,19] The best crystals were obtained by the hanging drop vapor diffusion method[20] using a droplet of 10 µl of 10 mg/ml protein in addition to 5 µl 25% ammonium sulphate in Tris-HCl buffer containing 400 mM NaCl and equilibrating it against a reservoir of 1 ml 25% ammonium sulphate solution. We have now modified the crystallization condition using sodium acetate as an additive in the precipitant solution. To 5 µl of 12 mg/ml protein solution in 100 mM Tris-HCl buffer containing 400 mM NaCl, 2 µl of 5% ammonium sulphate solution containing 10 mM sodium acetate at pH 5.4, was added and equilibrated against 25% ammonium sulphate solution containing 60 mM sodium acetate, at 4°C. This yielded more stable and reproducible single crystals of hexagonal bipyramidal shape in about seven days. These crystals, growing up to a maximum dimension of 0.6 × 0.3 × 0.3 mm, belong to the hexagonal space group $P6_122$ with cell parameters a = b = 61.8 Å and c = 212.8 Å. Intensity data up to 2.3 Å resolution were collected on a Rigaku R-axis Imaging Plate scanner with a rotating anode X-ray generator (50 kV, 100 mA) at Bhaba Atomic Research Centre, Mumbai, India. The crystal was mounted with the c axis perpendicular to the X-ray beam. A total of 72 images were recorded at an interval of 0.5°. The integrated intensities were measured and scaled using the

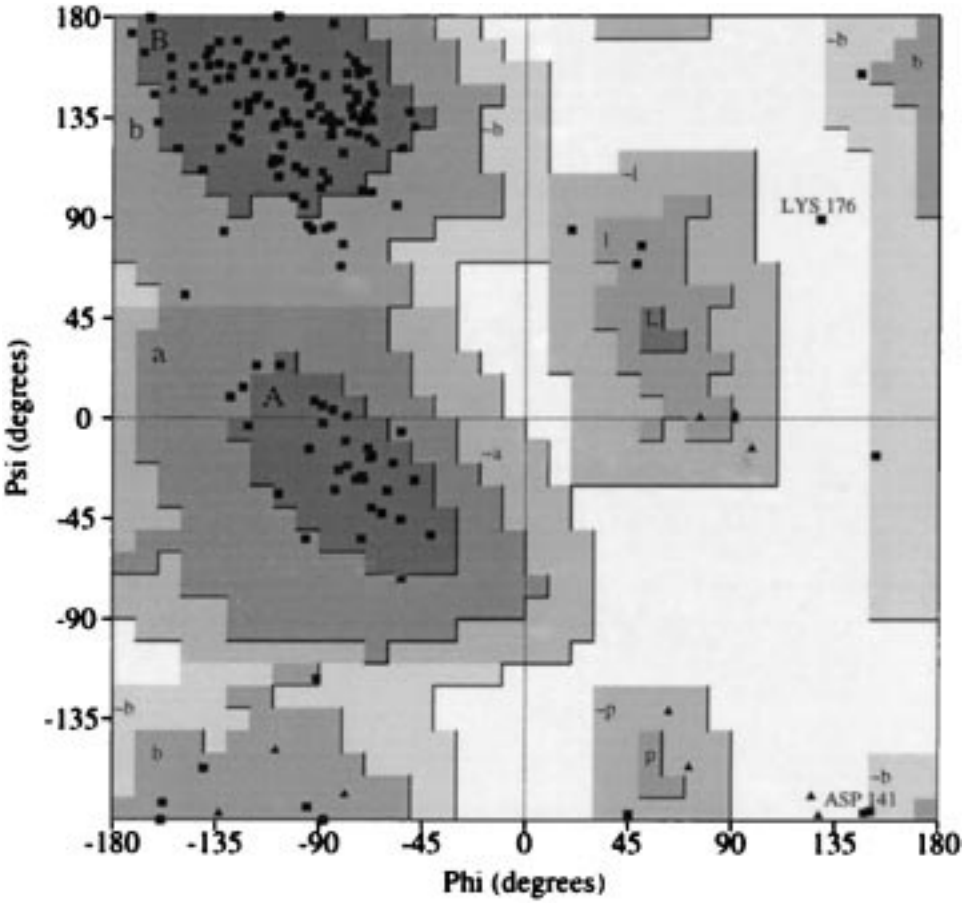

Fig. 1. A Ramachandran Plot (PROCHECK) of all 183 residues of WCI. Allowed regions are shaded according to convention. All non-glycine residues are indicated as squares and glycines are indicated as triangles.

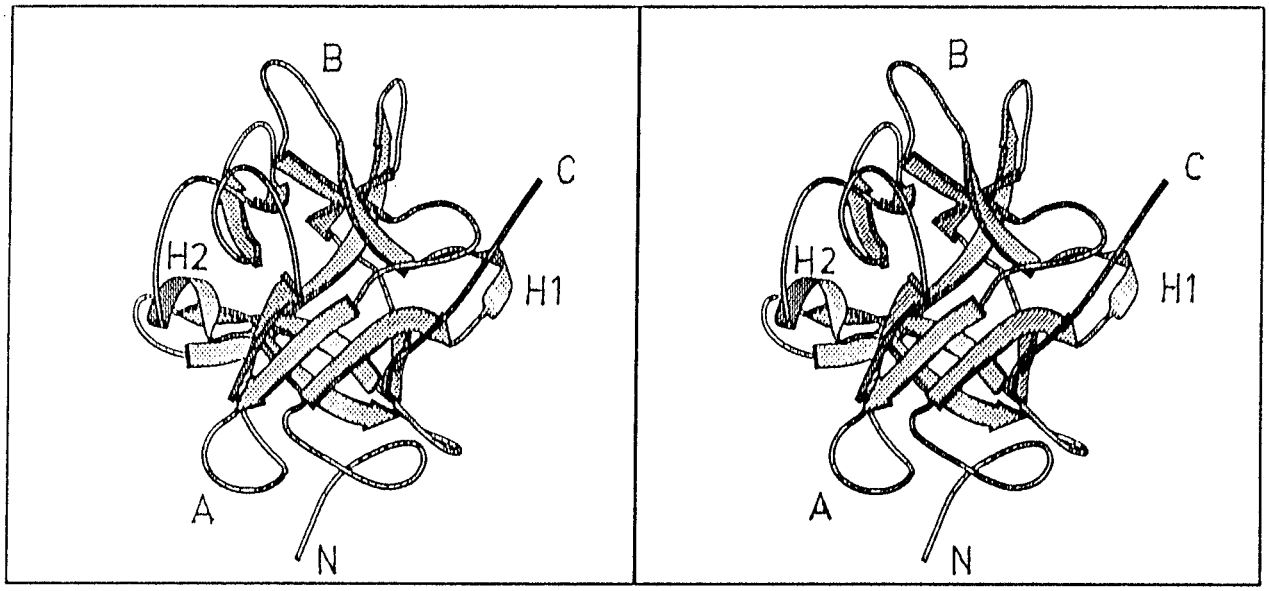

Fig. 2. A stereoscopic view of the schematic representation (Ribbon diagram)[70] of WCI molecule showing the β-trefoil fold with two small helical regions (H1, H2). First (A) and second (B) reactive sites and two termini of the protein are also marked.

TABLE I. Salt Bridges

Acceptor	Donor	Dist (Å)
LYS174 NZ	GLU13 OE2	2.89
GLN136 NE2	ASP145 OD2	2.98
ARG159 NE	ASP153 OD2	3.01[a]
ARG159 NH2	ASP153 OD2	2.78[a]
LYS103 NZ	GLU39 OE2	3.43[a]

[a]Marked salt bridges are conserved in the structures of ETI, STI, WBA in the Kunitz (STI) family.

TABLE II. Structural Similarity of Different Proteins with WCI[†]

Protein	Percentage sequence identity	Total number of residues	Number of Cα superposed	R.m.s.d (Å)
Erythrina caffra trypsin inhibitor	61	166	146	1.2
Winged Bean albumin	36	171	153	1.8
Interleukin-1α	12	145	109	2.5
Interleukin-1β	14	153	101	2.8
Basic fibroblast growth factor	10	126	105	2.7
Abrin-a sugar complex (B chain)	7	267	96	2.2
Interleukin-1 receptor antagonist	10	144	101	2.7
Agglutinin (Amaranthin)	10	299	115	2.6

[†]Co-ordinates of STI and PKI3 are not currently available in PDB.

program DENZO.[21] A total of 9,897 unique reflections out of 10,551 observations were obtained with a merging R-factor [$R_{merge} = \Sigma \mid - \langle I \rangle \mid \Sigma I$, where I = observed intensity, $\langle I \rangle$ = average intensity obtained from multiple observations of symmetry related reflections] of 9.5% which accounts for 80.1% of the expected number of reflections at 2.3 Å resolution. The intensities were converted to structure-factor amplitudes and a correction was applied to weak or negative measurements based on the a priori distribution. The overall temperature factor estimated from the Wilson plot[22] is 34.2 Å^2.

Refinement

The starting model for the refinement was the coordinate set of the previously determined structure[10] of WCI at 2.95 Å excluding the solvent water molecules. The initial R-factor calculated with the present data (taken at this stage in the 8.0 Å–2.95Å resolution shell but without any σ cutoff) was 36.2%. The structure was refined in several steps by a combination of rigid space refinement, positional refinement, and simulated annealing with the program X-PLOR[23] implemented on a SGI Indigo^2 workstation. Standard X-PLOR topology and parameter files (tophcsdx.pro and parhcsdx.pro) were employed in the process. The initial model after rigid space refinement was subjected to 100 cycles of positional refinement followed by 20 cycles of restrained B-factor refinement, which reduced the crystallographic R-factor from 36.2% to 24.1%. The

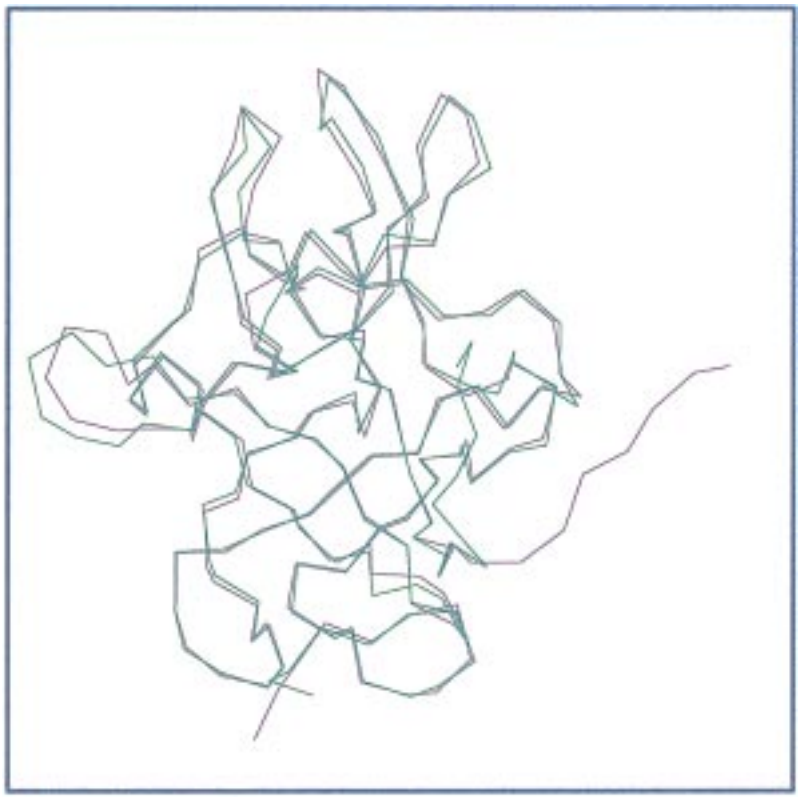

Fig. 3. Å superposition of C^{α} traces of co-ordinates from 2.95 Å (green) and 2.3Å models (magenta).

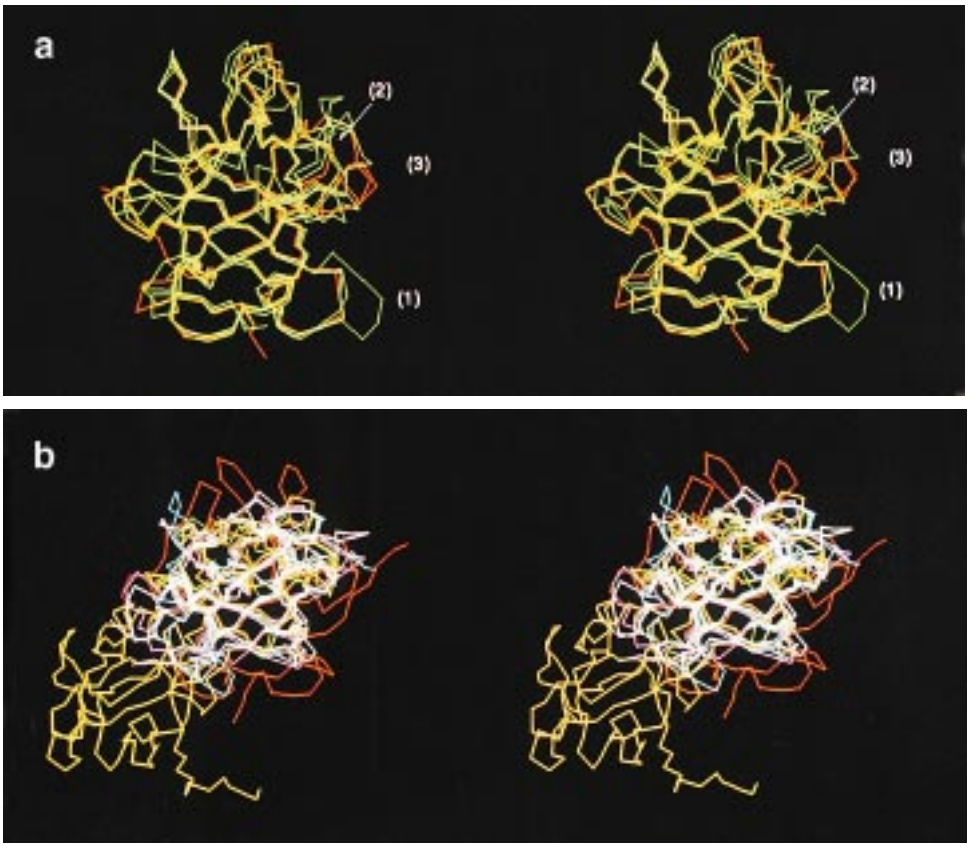

Fig. 4. Structural comparison of WCI with other proteins. **a**: C^{α}-chain superposition of structurally similar proteins of the Kunitz (STI) family showing the conserved region. WCI is shown in red. It shows maximum deviations with all the four structures near the loops marked 1, 2 (on the rear side of the molecule), and 3. **b**: C^{α} chain superposition of functionally unrelated proteins [Abrin-a (yellow), Interleukin 1α (magenta), Interleukin 1β (cyan), Interleukin receptor antagonist (white), Basic fibroblast growth factor (green), and WCI (red)] having structural similarity with WCI.

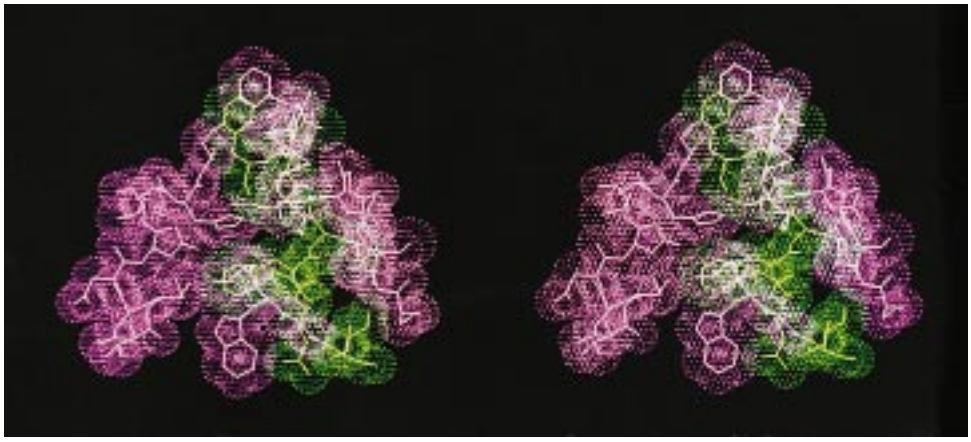

Fig. 5. The first reactive site of WCI (green) is docked at the enzyme (magenta) active site (van der Waals surface is shown). Figure shows His57 at the center having a hydrophobic environment (partially shown for clarity in viewing).

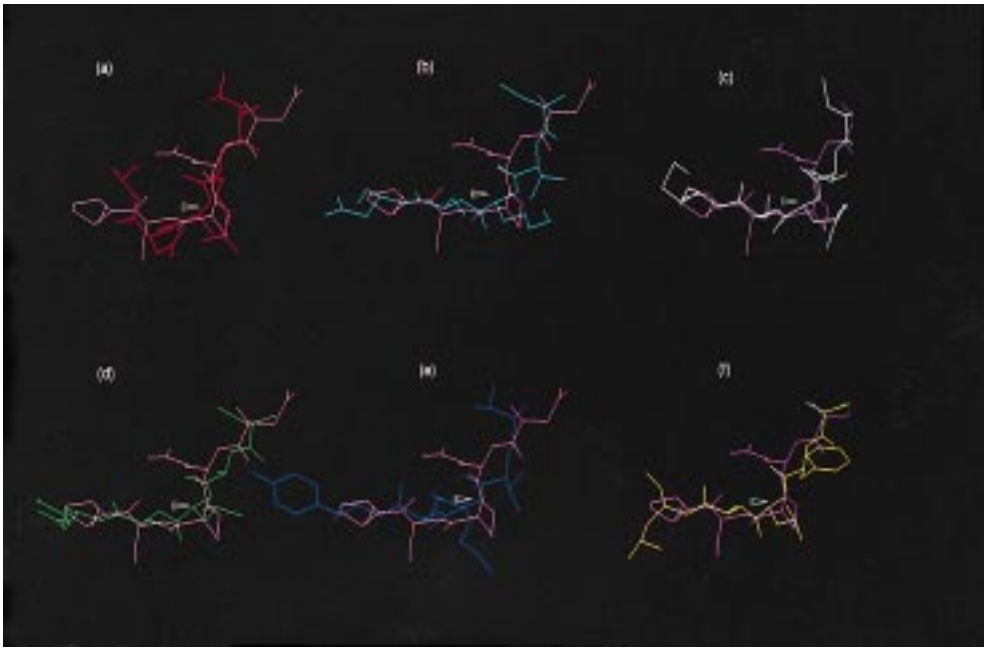

Fig. 6. A comparison of the canonical conformation (P3–P2') of the proposed second reactive site (Asn38–Pro42) of WCI with those of other reactive sites of serine protease inhibitor proteins. Figure shows main-chain superposition of WCI (P3–P2') (magenta) with **(a)** PTI, **(b)** PI-II, **(c)** BBIC, **(d)** BBIT, **(e)** WCI first site, and **(f)** CI-2. Arrows indicate P1 residues.

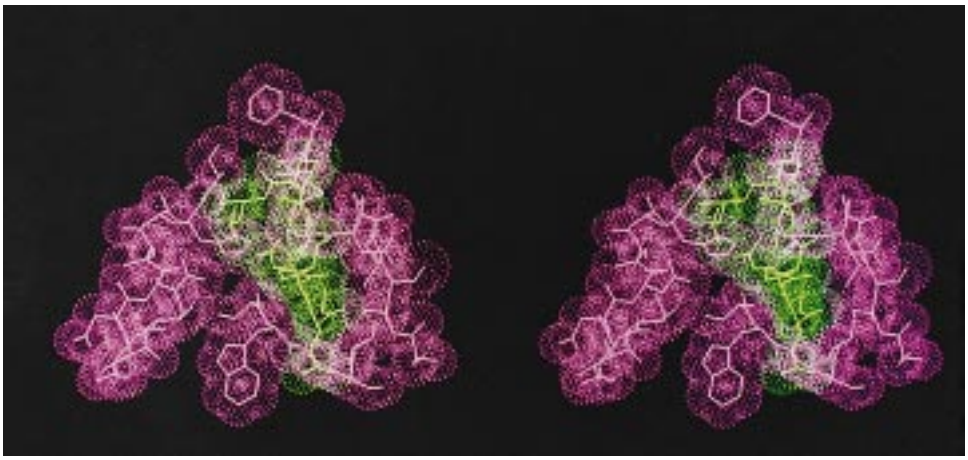

Fig. 7. The second reactive site of WCI (green) docked at the active site groove of the enzyme (magenta) with van der Waals surface. Anti parallel β-sheet (bottom) and hydrophobic Leu43 of WCI above His57 (center) of the enzyme is seen.

data set was then extended to 2.3Å in successive steps for a further 8 cycles, each cycle comprising of a round of X-PLOR refinement followed by manual intervention using TOM.[24,25] The R-value gradually fell from 24.1% to 22.3% and the electron density maps calculated during the course of refinement improved steadily and most of the missing side-chains of the earlier model could now be located in the $2F_o$-F_c and F_o-F_c maps. In the later stage of refinement we also used the program "O".[26] New potential solvent sites were located by visual inspection and with the program PEAKMAX.[27] Unidentified and distinct electron density peaks which appeared in greater than 2.5σ level of the F_o-F_c map were considered as water molecules in the model if they were within hydrogen bonding distances from a protein donor or acceptor molecule or from another water molecule. They were deleted from the model if they failed to reappear in 1.0σ level of the $2F_o$-F_c map. Fairly generous and modified[28,29] criteria for hydrogen bonds were taken, to take all genuine hydrogen bonds into consideration for this higher resolution structure. For N-H....O bonds the maximum H...O distance was taken as 2.60Å, and the minimum values for the angles N-H..O and C-O...H as 90°. For O-H...O and N-H...O bonds for which hydrogen positions were not assigned, the maximum distance O...O(N) was 3.6Å, and the minimum angle C-O...O(N) was 90°. The occupancy and isotropic temperature factors for all the currently located water molecules were assigned the values of 1.0 and 20 Å^2 respectively. Further, all main- or side-chain atoms at this stage with poor or no density at the 1.0σ level of the $2F_O$-F_C map were assigned zero occupancy.

The model was then subjected to the slow cooling protocol of simulated annealing refinement at an optimized temperature of 4,273 K in the resolution range 8–2.3 Å. The R-factor after slow-cooling, a few cycles of regularization through positional refinement and restrained B-factor refinement was 20.4%. At this stage further model building did not improve the R-factor and a model overfitting was indicated by the calculated free R-factor.[30,31] Calculation of the omit maps[32] omitting the relatively ambiguous loop regions (Asp1-Asp3, His137-Val142, Ser177-His183), followed by cautious stepwise incorporation of these missing residues, improved the R-factor to 19.8%. At the 2.3 Å resolution level we did not find it useful to consider the possible multiple conformation of some residues in spite of the presence of bifurcated electron density maps, as in the case of Lys143. Finally, the model was subjected to 120 cycles of positional and 20 cycles of individual isotropic B-factor refinement, with a weighting factor (W_A) of twice that of Brünger's suggested value[23] — this gave a better convergence of the refinement, without leading to serious deterioration of the model geometry.[33] The final R-factor for 1,572 non-hydrogen atoms is 18.7% for all 183 amino acid residues and 109 solvent molecules. During the entire process of refinement the free R-factor and its difference with the crystallographic R-factor was monitored. The structure factor data and the refined coordinates (Identity codes R2WBCSF and 2WBC respectively) have been deposited with the Protein Data Bank.[34]

RESULTS AND DISCUSSION

Quality of the Model and Comparison With Similar Structures

The current model of WCI contains all the 183 amino acid residues together with 109 water molecules. The electron density map of the model was of good quality with well- resolved positions for the majority of the main- and side-chain atoms. However, the maps for the residues His137–Asp141 and residues Ser177–His183 in the C-terminal region were of somewhat poor quality. Corresponding regions were also found to be ill defined in other structurally similar proteins like ETI,[8] Interleukin-1α,[35] Interleukin-1β[36,37,38] etc. The B-values for the regions His137–Asp141 and Ser177–His183 are relatively higher which may be indicative of disorder in these atoms. The high B-values in the regions could be due to dynamic flexibility or to static disorder within the crystal.[37] Moreover, the average B-values in the corresponding regions of all the above structures were seen to be relatively high, which may have some functional implications[38] also. The average B-factor for all protein atoms of WCI is 30.8 Å^2. The large solvent content and fewer contacts with symmetry related molecules in the crystal may be responsible for this high overall B-factor. In general the B-factor amplitudes correlate well with the location of the residues in the structure – the higher values are observed for the atoms on the surface loops and turns, while the lower ones correspond to β-rich core region atoms.

The estimated average error in coordinates of the new model is 0.28 Å as obtained from the Luzzati plot,[39] which corresponds well to the ideal value for protein structures. Almost all values of the main-chain dihedral angles (ϕ/ψ) lie within the stereochemically allowed regions[40] of the Ramachandran plot (Fig. 1).[41] The mean rmsd of bond lengths calculated for the backbone and side-chain atoms are 0.009 and 0.013 Å, respectively. The average rmsd for the bond angles, dihedral angles, and improper angles are 1.136°, 22.86°, 0.81°, respectively. The WCI protein molecule (Fig. 2) is approximately spherical (radius = 16.5 Å) and its structural architecture consists of 12 antiparallel β-strands connected by long loops which dominate the surface of the molecule. Six of the 12 β-strands form a β-barrel and the other six form a cover on one hollow end of the barrel—forming the characteristic β-trefoil fold structure.[16] The folding of the WCI molecule is such that the N-terminus comes close to the first reactive site loop (Gln63-Phe68). The side-chain of an asparagine residue (Asn14) near the N-terminus intrudes inside the reactive site loop and forms a network of hydrogen bonds with side-chain and main-chain atoms in the region, which stabilizes the loop.[10] One water molecule, located inside the loop, forms hydrogen bonds with Asn14 $O_{\delta 1}$ and Ser61 Oγ, which help to maintain the conformational rigidity of the loop. Altogether sixteen β-turns are identified in the WCI structure according to the hydrogen-bonding pattern Oi – Ni+3.[42] Of the five β-bulges identified, one antiparal-

lel classic β-bulge involving residues Ala34, Leu43 and Thr44 can be identified in the structure according to the hydrogen bonding pattern and the values of dihedral angles.[43] Another wide G1 type pseudo β-bulge is seen near Asn157 where H_2O205 and H_2O216 make the necessary contacts through hydrogen bonding. Four inverse γ-turns are located in the new model. Two 3_{10} helices (Trp25–His27 and Glu111–Asp113) were also evident in this predominantly β-sheet protein (Fig. 2). The second helix region actually takes the conformation of an αII helix with a bifurcated hydrogen bond at Glu111 O. Five salt-bridges seen in WCI are listed in Table I. Figure 3 shows a least-squares C^{α} chain superposition of the present model of WCI on the earlier one which shows some deviations, the largest being in the chain direction at the C-terminus (overall root mean square deviation (rmsd) value 0.7 Å). It is to be noted that the structure of WCI is similar to those of other legume seed serine proteinase inhibitors like ETI[8] and STI,[6,7] proteinase K inhibitor from wheat germ PKI3,[9] winged bean albumin WBA-1,[11] all belonging to the Kunitz (STI) family. Figure 4(a) shows a C^{α} superposition of all these five structures and overlapping is particularly good in the conserved β-sheet regions. Some variations are seen in the surface loop regions and interestingly, WCI shows consistent deviations in superposition with all the other four structures in the three loop regions (marked 1,2,3 in Figure 4(a)) in common. In conformation with the close sequence homology and bioactivity, WCI has very little variation in structure[10] in comparison to STI and ETI, whereas PKI3 shows considerable deviation in the loops 1, 2, and 3. Particularly loop 1 in PKI3, which corresponds to the first reactive site of WCI, lacking standard canonical conformation for serine protease reactive sites, cannot inhibit trypsin/chymotrypsin at this loop (instead, loop 2 of PKI3 acts as a reactive site against Proteinase K). In case of WBA, a four-amino acid insertion at loop 1 makes the loop much more protruding at the P1 equivalent side.[11] Structural similarity is also observed in functionally unrelated proteins like Interleukin-1α,[35] Interleukin-1β,[36,37,38] Human fibroblast growth factor,[44,45,46] Abrin-a sugar complex,[47] Interleukin-1 receptor antagonist[48] (Figure 4(b)) and in Amaranthin[49] whose structure has been recently reported. Interestingly, even for these functionally different proteins, the β-sheet regions appear to be somewhat conserved, but there is a wide variation in the superposition of surface loop regions in these proteins which may be thought to be responsible for their functional diversity. An analysis of the structural similarity of these proteins based on the protocol of Holm and Sander, 1993[50] is shown in Table II.

The Solvent Interactions and Crystal Packing

The crystals of WCI have a solvent content of ~68% (Vm = 3.85 $Å^3$/dalton). A total of 109 solvent molecules are included in the final refined model and their reliability was verified with the "Qualwat" parameters.[51] The average B-factor is 46.3 $Å^2$ for all solvent molecules. Most water molecules are located on the surface of the protein and form hydrogen bonds with polar side-chain atoms and/or another water molecule. Of the 109 water molecules, 15 are part of the second hydration shell which have at least one hydrogen bond with other water oxygen atoms within a cutoff distance of 3.6 Å. Only 3 water molecules (217, 255 and 261) have no potential hydrogen bonding partners and have contacts just outside the 3.6 Å limit. Very few water molecules are present in internal cavities and they are mostly clustered around β-strand termini. One extensive water channel[52] runs alongside residues Thr36–Val45 and makes contacts with the highly conserved Lys103, Arg159, and Leu160. Some of these multi-bonded water molecules are located near the barrel axis and roughly make contacts with the 3 sub-domains of the protein, which are related by the pseudo-three fold axis of symmetry. The water molecule 214 is one such solvent molecule which makes contact with main-chain oxygens of Val5 (3.04Å), His128 (2.78Å) and the main-chain nitrogen of Leu172 (2.89Å), thus conferring extra stability to the protein.

Within the unit cell, each WCI molecule makes intermolecular contacts with five symmetry ($P6_122$) related molecules. Maximum number of crystal packing interactions are seen between the surface regions His23–Trp25 (x,y,z) with His124–Asn126 (x,x-y,-z+⅙) and Asp95–Pro97 (x-y,-y,-z), Phe64–Leu65 (x,y,z) with Glu111–Leu115 (x-y,-y,-z) and Asp1–Asp3 (x,y,z) with Gln107–Pro110 (x,x-y,-z+⅙). An interesting revelation is the close packing of the C-terminus (imidazole ring of His183) into a hydrophilic pocket made by Glu139 (x,y,z), His23 (x-y,-y,-z) and His124 (x-y,x,z+⅙).

Mechanism of Enzyme Inhibition

In order to delineate the inhibitory action of WCI, we tried to model the WCI: α-Chymotrypsin complex by docking the first reactive site loop (Gln63–Phe68) of the inhibitor at the active groove of its cognate enzyme α-chymotrypsin guided by the criteria seen in STI:Trypsin complex.[6] We have used the PDB coordinates of chymotrypsin A-chain[53] in our model. While modelling the complex, some clashes were seen between moieties in the loop regions of the inhibitor and the enzyme, some of which do not belong to the eight recognized enzyme loops that control protease–inhibitor interaction in the chymotrypsin family.[54] Although there is a plausibility of structural readjustments of the residues in the reactive site loop, we did not consider it, as Bode and Huber in 1992[3] pointed out that the canonical conformation of this loop is the principal criterion behind enzyme–inhibitor recognition. So, to avoid the clashes, very small "hinge bending"[55] was introduced by changing the ϕ/φ angles around the P5-P4 and P4'-P5' residues of WCI reactive site loop, with an rmsd within 2.3 Å, keeping the P4-P4' region unchanged. In the overall process, the catalytic triad was kept within hydrogen bonding distances of the scissile bond (Leu65 – Ser66) residues and the N-terminal side of the scissile bond, in particular, was kept as an anti-parallel β-strand[3] with the corresponding sites of the enzyme (Fig. 5). After energy minimization,[23] the refined model gave a surface comple

mentarity value of 0.65 at the active site interface.[56] This model was further verified through the program GRAMM[57] employing a global range (six dimensional) molecular matching algorithm for complexation studies. The complex model reveals that the active site residues of the enzyme like Ser195, His57 and the scissile bond flanking hydrophobic residues (Leu69, Pro70, and Trp64) of the inhibitor are buried in the complex core (Fig. 5). Especially, Ser195 Oγ and His57 Nϵ atoms are screened by their immediate hydrophobic environment which makes them inaccessible to a solvent probe of 1.4 Å radius. It therefore follows that complete inaccessibility of the solvent to the active site residues in the acylenzyme intermediate of the enzyme would prevent deacylation and therefore even after the P1-P1' peptide bond is cleaved, the fragments of the inhibitor will be held in position in the complex. A similar mechanism of inhibition involving this reactive site loop has been reported for STI:Trypsin complex[7] and the model structure of ETI.[8] No enzyme inhibition is seen for WBA.[58] Recently, a comparison of the structure of WBA with our previous model of WCI has been made by McCoy and Kortt (1997).[11] From this structural comparison it is clear that the lack of inhibitory activity of WBA is due to the distorted structure of the expected inhibitory loop from the canonical conformation. Structural comparison with WCI shows that the change in structure of the loop is due to insertion of four amino acids in WBA between residues equivalent to P2 and P1. In case of PKI3, however, the mechanism of action against Proteinase K is different.[59]

TABLE III(a). Superposition of the Main-Chain Trace (P3-P3′) of Prospective Second Reactive Site Regions on That of the First Site of WCI

Region	R.m.s.d (Å)
Pro22-His27	1.59
Asn38-Leu43	1.11
Thr163-Leu16	2.41

TABLE III(b). C^{α} Trace (P3-P1′) Superposition of the Second Reactive Site of WCI on Other Known Sites of Serine Protease Inhibitors

Inhibitors	R.m.s.d (Å)
PTI	0.30
BBI (trypsin)	0.65
PI-II	0.40
BBI (chymotrypsin)	0.17
CI-2	0.47
WCI first site	0.43

The Second Reactive Site

One of our prime reasons for solving this inhibitor structure was also to make an attempt to locate its second reactive site, particularly because WCI is the only known double-headed inhibitor in the Kunitz (STI) family. It is to be noted that no previous biochemical information about the position of the second reactive site in WCI was available. In our current model, we could locate all the amino acid residues on the surface loops. Also, the electron density in the loop regions is better developed and the model fitting is of improved quality. The average B-factor for atoms in the loop regions in this model is also lower than that in our earlier model. Therefore we thought of exploring the possibility of locating the second reactive site of WCI from the structural details of this model.

To start with, a thorough search on the surface loops was made for a similar amino acid sequence pattern as that in the first reactive site (Gln63–Phe68). Additional emphasis was given to those regions that are related to the first site by the pseudo three-fold symmetry. As these efforts gave no substantial indications, we then tried to identify the second reactive site on the basis of structural similarity.[60] We chose three surface loops (Pro22–His27, Asn38–Leu43, Thr163–Leu168) that were not within a distance of 20 Å from the first reactive site, so as to avoid steric clash between the two independently and simultaneously bound chymotrypsin (each having approximately 40 Å diameter) molecules at the first and second reactive sites. The choice of the P1 residue for these regions was done through an optimization of the hydrophobic nature (for chymotrypsin specificity) of the residue concerned, where the rmsd of the region around it with that of the first reactive site is the lowest.

At this stage we wanted to know the influence of scissile bond-flanking residues in such enzyme–inhibitor interactions and with that view we wanted to analyze the hydrogen bonding pattern of these residues with the respective ones in the enzyme. So, all available coordinates of serine protease complexes with inhibitors at high resolution were collected from the PDB and contacts/hydrogen bonds in the interface region were calculated with the program CONTACT.[27] The analysis shows that more than 90% of the contact area (buried surface) between the two macromolecules is in the P3–P3' region of the reactive site of the inhibitor. A detailed protease cleavage analysis by Keil (1993)[61] also reveals this fact.

Keeping this observation in view, we superposed the main-chain traces of each of the three potential surface loops on that of the first reactive site (Gln63–Phe68) region of WCI in turn and the rmsd were calculated, using Insight II (1995)[62] molecular modelling system, for the P3–P3' regions. Two loop regions (Pro22–His27 and Asn38–Leu43) with lower rmsd values (Table III (a)) were selected for further investigation as a possible second reactive site. Next, we tried to dock each of these two potential loop regions at the chymotrypsin active site groove individually and we observed that docking of the loop Pro22–His27 was not possible without major changes in the conformation of either protease or inhibitor. This loop was thus eliminated leaving only one loop (Asn38–Leu43) for further consideration. Then we superposed (Fig. 6) the C^{α} trace of the

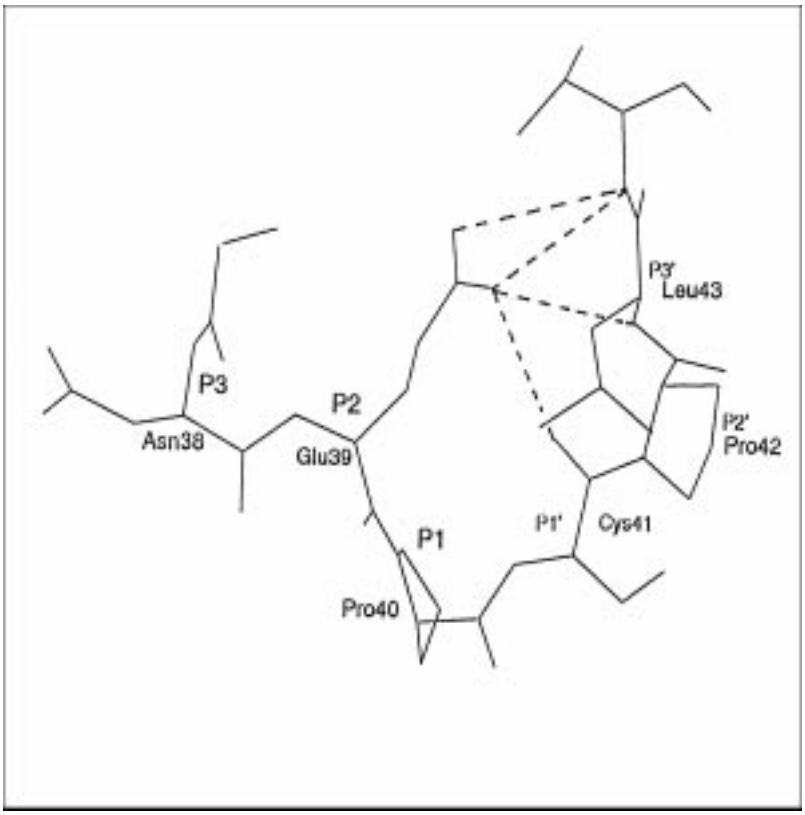

Fig. 8. The canonical conformation of the second reactive site loop. The side-chain of Glu39 intrudes inside the loop and stabilizes it through hydrogen bonding.

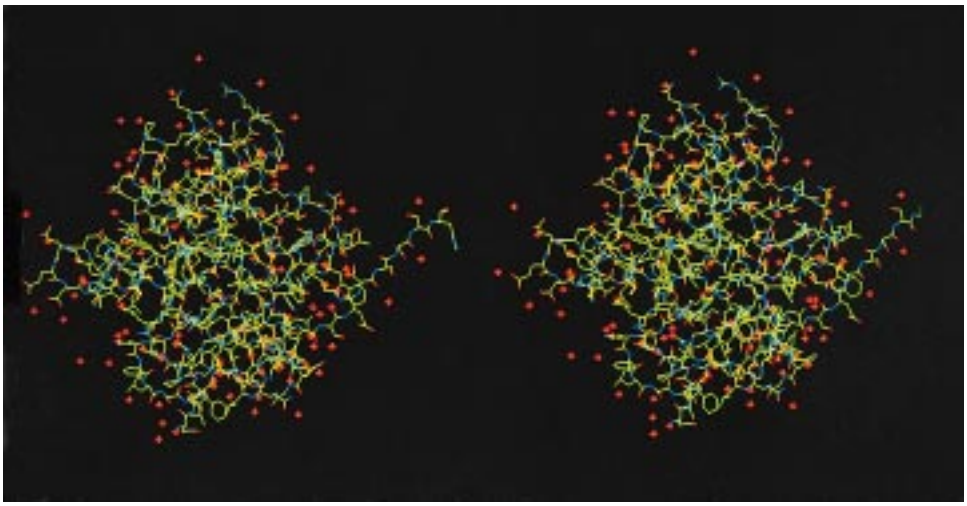

Fig. 9. Final model of WCI at 2.3 Å with red stars as water molecules.

proposed second reactive site loop on those of other known reactive site loops of serine protease inhibitors like the pancreatic trypsin inhibitor (PTI),[63] the Bowman-Birk type proteinase inhibitor (PI-II),[60] the Bowman-Birk inhibitor of chymotrypsin (BBIC)/trypsin (BBIT) from soybean,[64] the first site of WCI[10] and the chymotrypsin inhibitor from barley seeds (CI-2)[65] and rmsd were calculated in each case separately (Table III (b)). Consistently low rmsd values in each case was observed and the region comprising of residues Asn38–Leu43 was therefore considered a suitable candidate for the second reactive site of WCI.

Bode and Huber (1992)[3] showed that the serine protease inhibitor reactive sites must have a common canonical conformation determined by its φ/ψ distribution in P3–P3' region in order to bind appropriately at the enzyme active site. They also showed that to form a stable complex, the

inhibitor must present its reactive site flanking residues as an anti parallel β-sheet to the active groove of chymotrypsin, forming an intricate hydrogen bonding network. These criteria are reasonably satisfied by the proposed loop and this site can also be docked well at the binding site of the enzyme (Fig. 7). Figure 7 shows that the immediate vicinity of the P1–P1' (Pro 40–Cys 41) bond for the second site also is residing in a hydrophobic environment. Further, the presence of two Proline residues (P1 and P2' respectively) at the second site reduces the number of hydrogen bonding possibilities, which explains the lower binding affinity value (7.75×10^5 M^{-1} for the second site compared to 4.2×10^7 M^{-1} for the first site) of the inhibitor second site against the enzyme.[55,56] This also conforms to the biochemical observation that the 2:1 enzyme-inhibitor complex is formed only after supersaturation, showing that the distorted Michaelis type complex is formed through a competitive inhibition. The empirical surface complementarity value at this interface after docking is 0.48. Two other structural features of the region give stability to the loop, Asn38–Leu43. First, the presence of a disulfide bridge (Cys 41–Cys 85) in the vicinity of the P1-P1' bond and second, the extended hydrophilic side-chain of the P2 residue Glu39, which forms hydrogen bonds with other residues of the loop (Fig. 8), representing itself as a skeletal support element playing a similar role as that of Asn14 at the first site of the inhibitor.

It was reported[14] that there was no cleavage at the WCI second reactive site even after prolonged incubation with chymotrypsin. In this connection it is to be noted that proline is present at P1 and P2' sites of the proposed loop–perhaps the most important evidence in support of the absence of cleavage. This is because it is an observed fact[66] which is analytically supported by Keil (1993),[61] that there is virtually no cleavage against serine protease interaction when proline is present at the P1 position of the reactive site of substrate/inhibitor and presence of proline as one of the nearby flanking residues enhances this negative effect. Proline, being an imino rather than amino acid, does not posses the free hydrogen atom that is essential to the catalysis by chymotrypsin by forming an P1-S1 hydrogen bond with a backbone carbonyl oxygen and this may be the reason why serine proteases do not hydrolyze at prolyl bonds [67,68,69]. All these supporting evidences suggest that the surface loop Asn38-Glu39-Pro40-Cys41-Pro42-Leu43, may be considered as the second reactive site of WCI (Fig. 9 and also marked as B in Fig. 2).

CONCLUSION

In this paper, we have reported the improved structure of WCI at 2.3 Å resolution which leads to an understanding of the mechanism of inhibition of the protein against α-chymotrypsin. Analysis of this higher-resolution structure also helped us to predict the location of the second reactive site of the protein that was identified mainly on the basis of a best-fit structural study of the available canonical conformations of the serine protease inhibitor reactive sites.

ACKNOWLEDGMENTS

This work is supported by a grant (BT/R&D/15/15/92) from the Department of Biotechnology, Government of India. D.M. acknowledges the University Grants Commission, Government of India, for a Research Fellowship.

REFERENCES

1. Kraut J. Serine proteases: structure and mechanism of catalysis. Ann Rev Biochem 1977;46:331–358,
2. Laskowski, Jr M, Kato I. Protein Inhibitors of proteinases. Annu Rev Biochem 1980;49:593–626.
3. Bode W, Huber R. Natural proteinase inhibitors and their interaction with proteinases. Eur J Biochem 1992;204:433–451.
4. Ryan CA. Protease inhibitors in plants: genes for improving defenses against insects and pathogens. Annu Rev Phytopathol 1990;28:425–449.
5. Richardson M. The proteinase inhibitors of plant and microorganism. Phytochem 1977;16:159–169.
6. Sweet RM, Wright HT, Janin J, Chothia CH, Blow DM. Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6 Å resolution. Biochemistry 1974;13:4212–4228.
7. Song HK, Suh SW. Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. J Mol Biol 1998;275:347–363.
8. Onesti S, Brick P, Blow DM. Crystal structure of a Kunitz-type trypsin inhibitor from Erythrina caffra seeds. J Mol Biol 1991;217: 153–176.
9. Zemke KJ, Müller-Fahrnow A, Jany K-D, Pal GP, Saenger W. The three-dimensional structure of the bifunctional proteinase K/α-amylase inhibitor form wheat (PKI3) at 25 Å resolution. FEBS Lett 1991;279:240–242.
10. Dattagupta JK, Podder A, Chakrabarti C, Sen U, Dutta SK, Singh M. Structure of a Kunitz-type chymotrypsin inhibitor from winged bean seeds at 2.95 Å Resolution. Acta Crystallogr D 1996;52:521–528.
11. McCoy AJ, Kortt A. The 1.8 Å crystal structure of winged bean albumin 1, the major albumin from Psophocarpus tetragonolobus (L.) DC. J Mol Biol 1997;269:881–891.
12. Roy A, Singh M. Psophocarpin B1, a storage protein of Psophocarpus tetragonolobus, has chymotrypsin inhibitory activity. Phytochemistry (Oxford) 1988;27:31–34.
13. Gillespie JM, Blagrove RJ. Isolation and composition of the seed globulins of winged bean, Psophocarpus tetragonolobus (L.) DC. Aust J Plant Physiol 1978;5:257–369.
14. Kortt AA. Isolation and properties of a chymotrypsin inhibitor from winged bean seed (Psophocarpus tetragonolobus (L) DC). Biochim Biophys Acta 1980;624:237–248.
15. Shibata H, Hara S, Ikenaka T. Amino acid sequence of winged bean (Psophocarpus tetragonolobus (L.) DC.) chymotrypsin inhibitor, WCI-3. J Biochem 1988;104:537–543.
16. Murzin AG, Lesk AM, Chothia C. β-Trefoil fold: patterns of structure and sequence in the Kunitz inhibitors, interleukins - 1 β and 1α- and fibroblast growth factors. J Mol Biol 1992;223:531–543.
17. McLachlan AD. Three-fold structural pattern in the soybean trypsin inhibitor (Kunitz). J Mol Biol 1979;133:557–563.
18. Roy A, Singh M. Purification of a storage protein of Psophocarpus tetragonolobus. Phytochemistry (Oxford) 1986;25:595–600.
19. Shibata H, Hara S, Ikenaka T, Abe J. Purification and characterization of proteinase inhibitors from winged bean (Psophocarpus tetragonolobus (L.) DC.) seeds. J Biochem (Tokyo) 1986;99:1147–1155.
20. Dattagupta JK, Chakrabarti C, Podder A, Dutta SK, Singh M. Crystallization and preliminary X-ray studies of Psophocarpin B1, a chymotrypsin inhibitor from winged bean seeds. J. Mol. Biol. 1990;216:229–231.
21. Otwinowski Z. Oscillation data reduction program. In: Sawyer L, Isaacs N, Bailey S, editors. Proceedings of the CCP4 study weekend: data collection & processing. Warrington, England: SERC Daresbury Laboratory; 1993. p 56–62.

22. Wilson AJC. Determination of absolute from relative X-ray intensity data. Nature 1942;150:152.
23. Brünger AT. 1992a. X-PLOR manual version 3.1. New Haven: Yale University. 382 p.
24. Jones TA. A graphics model building and refinement system for macromolecules. J Appl Crystallogr 1978;11:268–272.
25. Cambillau C, Harjales E. TOM: a Frodo sub-package. J Mol Graph 1987;5:175–177.
26. Jones TA, Bergdoli M, Kjeldgaars MO. A macro-molecule modeling environment. In: Bugg CE, Ealick SE, editors. Crystallographic and modeling methods in molecular design. New York: Springer-Verlag; 1989. p 189–199.
27. CCP4-collaborative computational project, number 4. Acta Crystallogr D 1994; 50:760–763.
28. Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 1984;44:97–179.
29. Madhusudan, Kodandapani R, Vijayan M. Protein hydration of water structure: X-ray analysis of a closely packed protein crystal with very low solvent content. Acta Crystallog D 1993;49:234–245.
30. Brünger AT. 1992b. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature (London) 355:472–475.
31. Brünger AT. Assessment of phase accuracy by cross validation; the free R value. methods and applications. Acta Crystallogr D 1993;49:24–36.
32. Bhat TN, Cohen GH. OMITMAP: an electron density map suitable for the examination of errors in a macromolecular model. J Appl Crystallogr 1984;17:244–224.
33. Rypniewski WR, Dambmann C, von der Osten C, Dauter M, Wilson KS. Structure of inhibited trypsin from Fusarium oxysporum at 1.5 Å. Acta Crystallogr D 1995;51:71–84.
34. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein data bank: a computer-based archival file for the macromolecular structures. J Mol Biol 1977;112:535–542.
35. Graves BJ, Hatada MH, Hendrickson WA, Miller JK, Mydison VS, Satow Y. Structure of interleukin 1-α at 2.7 Å resolution. Biochemistry 1990;29:2679–2684.
36. Priestle JP, Schär H-P, Grütter MG. Crystal structure of the cytokine interleukin -1β. EMBO J 1988;7:339–343.
37. Priestle JP, Schär H-P, Grütter MG. Crystallographic refinement of interleukin -1β at 2.0 Å resolution. Proc Nat Acad Sci USA 1989;86:9667–9671.
38. Finzel BC, Clancy LL, Holland DR, Muchmore SW, Watenpaugh KD, Einspahr HM. Crystal structure of recombinant human interleukin-1 β at 2.0 Å resolution. J Mol Biol 1989;209:779–791.
39. Luzzati V. Treatment of statistical errors in the determination of crystal structures. Acta Crystallogr 1952;5:802–810.
40. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. Proteins 1992; 12:345–364.
41. Ramakrishnan C, Ramachandran GN. Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformation for a pair of peptide units. Biophys J 1965;5:909–933.
42. Venkatachalam CM. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. Biopolymers 1968;6:1425–1436.
43. Richardson JS. The anatomy and taxonomy of protein structure. Adv Protein Chem 1981;34:167–363.
44. Ago H, Kitagawa Y, Fujishima A, Matsuura Y, Katsube Y. Crystal structure of basic fibroblast growth factor at 1.6 Å resolution. J Biochem (Tokyo) 1991;110:360–363.
45. Zhang J, Cousens LS, Barr PJ, Sprang SR. Three-dimensional structure of human basic fibroblast growth factor, a structural homolog of interleukin 1β. Proc Natl Acad Sci USA 1991;88:3446–3450.
46. Zhu Z, Komiya H, Chirino A, et al. Three-dimensional structures of acidic and basic fibroblast growth factors. Science 1991;251: 90–93.
47. Tahirov TH, Lu TH, Liaw YC, Chen YL, Lin JY. Crystal structure of abrin-a at 2.14 Å. J Mol Biol 1995;250:354–367.
48. Schreuder HA, Rondeau JM, Tardif C, Soffientini A. Refined crystal structure of the interleukin-1 receptor antagonist. Presence of a disulfide link and a cis-proline. Eur J Biochem 1995;227: 838–847.
49. Transue TR, Smith AK, Mo H, Goldstein IJ, Saper MA. Structure of benzyl T-antigen disaccharide bound to amaranthus caudatus agglutinin. Nat Struct Biol 1977;4:779.
50. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–138.
51. Arnold E, Rossman MG. Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 Å. J Mol Biol 1990;211:763–801.
52. Kleywegt GJ, Jones TA. Detection, delineation, measurement and display of cavities in macromolecular structures. Acta Crystallogr D 1984;50:178–185.
53. Blevins RA, Tulinsky A. The refinement and the structure of the dimer of α-chymotrypsin at 1.67 Å resolution. J Biol Chem 1985;260:8865–8872.
54. Creighton TE, Darby NJ. Functional evolutionary divergence of proteolytic enzymes and their inhibitors. Trends Biochem Sci 1989;14:319–324.
55. Frigerio F, Coda A, Pugliese L, et al. Crystal and molecular structure of the bovine α-chymotrypsin-eglin c complex at 2.0 Å resolution. J Mol Biol 1992;225:107–123.
56. Lawrence MC, Colman PM. Shape complimentarity at protein/protein interfaces. J Mol Biol 1993;234:946–950.
57. Vakser A, Aflalo C. Hydrophobic docking: a proposed enhancement to molecular recognition techniques. Proteins 1994;20:320–329.
58. Kortt AA. Isolation and characterisation of a major seed albumin: a crystalline protein from winged bean [Psophocarpus tetragonolobus (L) DC]. Int J Pept Protein Res 1986;28:613– 619.
59. Pal GP, Kavounis CA, Jany KD, Tsernoglou D. The three-dimensional structure of the complex of proteinase K with its naturally occuring protein inhibitor, PKI3. FEBS Lett 1994;341: 167–170.
60. Chen P, Rose J, Love R, Wei CH, Wang BC. Reactive sites of an anticarcinogenic Bowman-Birk proteinase inhibitor are similar to other trypsin inhibitors. J Biol Chem 1992;267:1990–1992.
61. Keil B. Data treatment. In: Specificity of proteolysis. New York: Springer-Verlag; 1993. p 7–18.
62. Insight II User Guide. San Diego: Biosym/MSI; 1995.
63. Huber R, Kukla D, Rühlmann A, Epp O, Formanek H. The basic trypsin inhibitor of bovine pancreas. I. Structure analysis and conformation of the polypeptide chain. Naturwissenschaften 1970; 57:389–392.
64. Werner MH, Wemmer DE. Three-dimensional structure of soybean trypsin/chymotrypsin Bowman-Birk inhibitor in solution. Biochemistry 1992;31:999.
65. McPhalen CA, James MNG. Crystal and molecular structure of the serine proteinase inhibitor/CI-2 from barley seeds. Biochemistry 1987;26:261–269.
66. Laskowski M Jr, Sealock RW. Protein proteinase inhibitors - molecular aspects. In: Boyer PD, editor. The enzymes-hydrolysis: peptide bonds. Vol 3. New York: Academic Press; 1971. p 375–473.
67. Bizzozero SA, Zwiefel BO. The importance of the conformation of tetrahedral intermediate for the alpha-chymotrypsin catalyzed hydrolysis of peptide substrates. FEBS Lett 1975;59:105–108.
68. Dutler H, Bizzozero SA. Mechanism of the serine protease reaction. Stereoelectronic, structural and kinetic considerations as guidelines to deduce reaction paths. Acc Chem Res 1989;22:322–327.
69. Polgár L. Prolyl oligopeptidases. Meth Enzymol 1994;244:188–200.
70. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24: 946–950.