Three-Dimensional Solution Structure of *Cucurbita maxima* Trypsin Inhibitor-V Determined by NMR Spectroscopy[†,‡]

Mengli Cai,[§] YuXi Gong,[§] Jeff L.-F. Kao,[∥] and Ramaswamy Krishnamoorthi*[,§]

Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506, and Department of Chemistry, Washington University, St. Louis, Missouri 63130

Received November 17, 1994; Revised Manuscript Received February 3, 1995[⊗]

ABSTRACT: The solution structure of *Cucurbita maxima* trypsin inhibitor-V (CMTI-V), which is also a specific inhibitor of the blood coagulation protein, factor XII_a, was determined by ^1H NMR spectroscopy in combination with a distance-geometry and simulated annealing algorithm. Sequence-specific resonance assignments were made for all the main-chain and most of the side-chain hydrogens. Stereospecific assignments were also made for some of the β-, γ-, δ-, and ϵ-hydrogens and valine methyl hydrogens. The ring conformations of all six prolines in the inhibitor were determined on the basis of $^1H-^1H$ vicinal coupling constant patterns; most of the proline ring hydrogens were stereospecifically assigned on the basis of vicinal coupling constant and intraresidue nuclear Overhauser effect (NOE) patterns. Distance constraints were determined on the basis of NOEs between pairs of hydrogens. Dihedral angle constraints were determined from estimates of scalar coupling constants and intraresidue NOEs. On the basis of 727 interproton distance and 111 torsion angle constraints, which included backbone ϕ angles and side-chain χ_1, χ_2, χ_3, and χ_4 angles, 22 structures were calculated by a distance geometry algorithm and refined by energy minimization and simulated annealing methods. Both main-chain and side-chain atoms are well-defined, except for a loop region, two terminal residues, and some side-chain atoms located on the molecular surface. The average root mean squared deviation in the position for equivalent atoms between the 22 individual structures and the mean structure obtained by averaging their coordinates is 0.58 ± 0.06 Å for the main-chain atoms and 1.01 ± 0.07 Å for all the non-hydrogen atoms of residues 3−40 and 49−67. These structures were compared to the X-ray crystallographic structure of another protein of the same inhibitor family—chymotrypsin inhibitor-2 from barley seeds [CI-2; McPhalen, C. A., & James, M. N. G. (1987) *Biochemistry 26*, 261−269]. The main-chain folding patterns are highly similar for the two proteins, which possess 62% sequence differences. However, major differences are noted in the N- and C-terminal segments, which may be due to the presence of a disulfide bridge in CMTI-V, but not in CI-2.

Serine proteinase inhibitors play an important role in the regulation of proteinases, which control many important biological functions, such as digestion, blood clotting, etc. (Neurath, 1984). Structural knowledge provides insights into the mechanism of their functions and may lead to strategies in the rational designing of molecules of medicinal importance. Over the years, X-ray crystallographic structures have become available for many inhibitors and inhibitor−proteinase complexes; in addition, a few NMR solution structures are also available for some inhibitors (Bode & Huber, 1992).

Cucurbita maxima trypsin inhibitor-V (CMTI-V;[1] M_r 7K), a member of the potato I inhibitor family, consists of a single polypeptide chain of 68 amino acid residues, whose N-terminal is acetylated (Krishnamoorthi et al., 1990; Figure 1); the presence of a disulfide bridge between residues 3 and 48 is unique to CMTI-V, as no other member of the family possesses one. CMTI-V also specifically inhibits the human blood coagulation protein factor XII_a (activated Hageman factor) but not plasma kallikrein, another serine proteinase closely related to factor XII_a. The reactive-site of CMTI-V has been identified to be the the peptide bond between Lys44 and Asp45; CMTI-V loses its inhibitory activity toward factor XII_a, upon hydrolysis of the reactive-site peptide bond (Krishnamoorthi et al., 1990). Trypsin-catalyzed hydrolysis of CMTI-V occurs at a much slower rate, as compared to that of substrates, and, in this regard, CMTI-V may be described as a "slow" substrate. Similar observations have been made with smaller (M_r 3K) squash family inhibitors, CMTI-I and CMTI-III, also isolated from pumpkin seeds (Krishnamoorthi et al., 1990). There are no sequence similarities between CMTI-I or CMTI-III and CMTI-V.

[†] This work has been supported by a grant from the National Institutes of Health. R.K. is an NIH Research Career Development Awardee (1994−1999). The 11.75 T NMR instrument used for this project was purchased with an NSF-EPSCoR grant. This is publication 95-355-J from the Kansas Agricultural Experiment Station.

[‡] Atomic coordinates for the refined average structure of CMTI-V, along with the NMR constraints, have been deposited with Protein Data Bank, Brookhaven National Laboratories, Long Island, NY 11973, under the accession code 1TIN.

* Author to whom correspondence should be addressed. E-mail: krish@ksuvm.ksu.edu. Fax: (913) 532-7278.

[§] Kansas State University.

[∥] Washington University.

[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1995.

[1] Abbreviations: CMTI-V, *Cucurbita maxima* trypsin inhibitor-V; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; CI-2, chymotrypsin inhibitor-2 from barley seeds; DQF-COSY, double-quantum filtered correlated spectroscopy; P.E.COSY, primitive extended correlated spectroscopy; TOCSY, total correlated spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; RMSD, root mean squared deviation; ppm, parts per million.

0006-2960/95/0434-5201$09.00/0

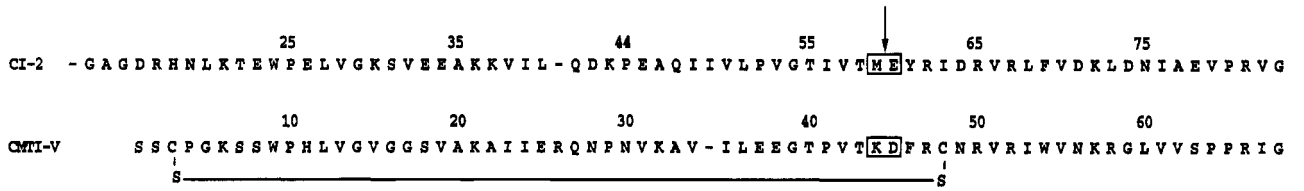

FIGURE 1: Alignment of amino acid sequences of CI-2 and CMTI-V. The arrow above the boxes indicates the reactive-site (P_1–P_1' peptide bond) in each inhibitor. The N-terminal of CMTI-V is acetylated (Krishnamoorthi et al., 1990).

We have been interested in characterizing structural and dynamic changes in CMTI-V that accompany the hydrolysis reaction and render the inhibitor nonfunctional toward factor XII_a. No single crystal structure of CMTI-V has yet been determined, although structural information is available for a couple of other members of the potato I inhibitor family: an X-ray crystal structure has been reported for barley chymotrypsin inhibitor 2 (CI-2), and NMR solution structures have been determined for CI-2 (Clore et al., 1987a,b) and eglin c, a chymotrypsin inhibitor isolated from the leech (Hyberts et al., 1992). The sequence from Asn19 to Gly83 in CI-2 can be aligned with the sequence from Pro4 to Gly68 in CMTI-V (Krishnamoorthi et al., 1990; Figure 1).

In a continuing effort to understand structure–function relationships in CMTI's, we have applied ^{1}H NMR spectroscopy to CMTI-V in solution, and, in this report, we present extensive proton resonance assignments (sequential and stereospecific), distance and torsional angle constraints, and the three-dimensional solution structure determination of CMTI-V on the basis of NMR experimental constraints and distance geometry and simulated annealing calculations. The average structure of CMTI-V thus obtained is compared to the X-ray crystallographic structure of CI-2 (McPhalen & James, 1987), and similarities and differences are discussed.

MATERIALS AND METHODS

Protein. Pure CMTI-V was isolated from pumpkin seeds, as previously described (Krishnamoorthi et al., 1990). A ~3 mM sample of CMTI-V in H_2O was prepared by dissolving the lyophilized protein in 0.5 mL of 90% H_2O/10% D_2O (v/v), which contained 50 mM KCl, pH 5.4. Another sample, also ~3 mM, was prepared in 99.996% D_2O, containing 50 mM KCl, pD ~5.4.

NMR Spectroscopy. 2D NMR experiments were performed with a 11.75 T Varian UNITY*plus* spectrometer (499.496 MHz for ^{1}H) or a 14.1 T Varian UNITY spectrometer (600 MHz for ^{1}H). Data sets were obtained at 10, 30, 40, and 50 °C and were processed, plotted, and analyzed using Varian NMR or FELIX software on a Silicon Graphics workstation. Spectra acquired at different temperatures helped resolve $C_\alpha H$–N*H* cross peaks that overlapped with the strong water signal. A total of 1024 increments of 4K data points were recorded for P.E.COSY and DQF-COSY experiments, and a total of 512 increments of 2K data points were recorded for all other experiments. All data sets were collected in hypercomplex phase sensitive mode. Before processing, the t_2 dimension of DQF-COSY data sets was zero-filled to 8K and the t_1 dimension of P.E.COSY and DQF-COSY data sets to 4K; the t_1 dimensions of data sets of all other experiments were zero-filled to 2K.

Backbone Torsional Angle Constraints. Backbone torsional angles, ϕ, were constrained on the basis of estimation of $^3J_{NH\alpha}$ coupling constants from peak-to-peak separations in the DQF-COSY map and from widths of amide to α-hydrogen cross peaks in the TOCSY map (Driscoll et al., 1989; Cai et al., 1992). Thus, 18 coupling constants larger than 8 Hz and 19 coupling constants less than 5 Hz were identified. The corresponding main chain ϕ torsional angles were constrained to be in the range −80° to −160° for the former and −40° to −90° for the latter. ϕ torsional angle constraints of negative values (−20 to −180°) were assigned to 13 additional residues whose amide hydrogens each gave a more intense NOE to the preceding residue's α-hydrogen, as compared with the NOE to its own α-hydrogen (Clubb et al., 1994). All the six prolines in the protein were identified to be *trans* in conformation, through NOEs from the α-hydrogen of the preceding residue to proline δ-hydrogens and through an NOE from the amide hydrogen of the following residue to the α-hydrogen of the proline (Wüthrich, 1986). All proline ω torsional angles were constrained to be 175–185°. No torsional angle constraints were applied to glycine residues.

Stereospecific Assignments and Side-Chain Torsional Angle Constraints. Stereospecific assignments of β-methylene hydrogens were made according to cross peak intensities of a short mixing time (10 ms) TOCSY map as well as intraresidue NOESY cross peak intensities (Wagner et al., 1987; Cai et al., 1995a). Methyl groups of valine residues were established according to the classification of Zuiderweg et al. (1985). From these stereospecific assignments, 12 χ_1 angles were constrained to be −60 ± 30°, four to be 180 ± 30°, and four to be 60 ± 30°. The large $^1J_{\alpha\beta}$ coupling constants (~10 Hz) for four isoleucine residues led us to constrain their χ_1 angles at −60 ± 30° (H_β at position *trans* to H_α). For valine residues, four of them were identified as g^- rotamer and five as g^+ rotamer. The corresponding χ_1 angles were constrained to be 60 ± 30 (g^-) and −60 ± 30° (g^+). γ-Hydrogens for four isoleucine residues were stereospecifically assigned (Cai et al., 1995a), and their χ_2 angles were constrained to be 180 ± 30°. Both χ_2 and χ_3 angles were constrained to be 180° for Arg52 and Lys21 during the process of stereospecific assignments of their γ- and δ-methylene hydrogens (Cai et al., 1995a). χ_1, χ_2, χ_3, and χ_4 angles of Lys6 were also determined, and stereospecific assignments were made for the side-chain hydrogens. Resolved resonances of all the six prolines were stereospecifically assigned (Cai et al., 1995b). Five proline ring conformations were identified to be in *Up* form (Pro4, 29, 41, 64, and 65) and one in *Down* from (Pro10). The χ_1, χ_2, and χ_3 angles were constrained to be −22 ± 15°, 33 ± 15°, and −31 ± 15° for *Up* form prolines and 26 ± 15°, −33 ± 15°, and 26 ± 15° for *Down* form proline.

Determination of Distance Constraints. Cross peaks were identified in the NOESY maps recorded at different mixing times (50, 80, 100, and 150 ms) and at

different temperatures (10, 30, and 50 °C). Distance constraints were determined from cross peak intensities in a NOESY map obtained with a mixing time of 100 ms, using the following guidelines: For cross peaks arising from only main-chain hydrogens, the upper limits of distances were set at 2.5, 3.0, 3.5, and 5.0 Å, respectively, for strong, medium, weak, and very weak NOEs. A cross peak was considered to be strong only if it appeared strong in a 50 ms mixing time NOESY experiment too. For cross peaks involving side-chain hydrogens, because of their likely higher flexibility, the upper limits were defined as 3.0, 4.0, and 5.0 Å, for strong, medium, and weak NOEs, respectively (Williamson et al., 1985). For methyl groups, 1.5 Å was added to the upper boundaries to correct for the center of average of the three methyl hydrogens, and to account for the higher intensities of the methyl resonances (Clore et al., 1987c; Wagner et al., 1987; Driscoll et al., 1989). For the stereospecifically unassigned methylene hydrogens, constraints were defined for the following two different situations: first, if NOEs were observed for both methylene hydrogens, upper boundaries were set according to the weak one; if, on the other hand, only one NOE was observed, or the two methylene hydrogens were degenerate, the constraint was applied to their carbon atoms, and 1 Å (C−H bond length) was added to the upper limits. In all cases, the lower boundaries were restricted to 2.0 Å as approximately the sum of van der Waals radii. Pseudostructures were used for phenylalanine residues, as described by Wüthrich et al. (1983). For the degenerate C2 and C6, and C3 and C5 hydrogens, constraints were applied to the carbon atoms at γ and 4 positions, respectively, and 2.0 Å was added to the upper limits.

Hydrogen-Bond Constraints. Twenty-five hydrogen-bond constraints were obtained on the basis of slow deuterium exchange kinetics of backbone amide hydrogens and interresidue NOEs (Wüthrich, 1986). Slowly exchanging backbone amide hydrogens were identified from a series of TOCSY maps, obtained at different time intervals after the protein was dissolved in D_2O. An amide hydrogen was considered to be slowly exchanging if it was still observed 24 h or longer after exposure to D_2O. A hydrogen-bond constraint was applied only for the regular secondary structures, α-helices, β-sheets, and β-turns, which were clearly identified on the basis of interresidue NOEs and $^3J_{HN\alpha}$ coupling constants. Two constraints, $r_{N-O} = 2.4-3.3$ Å and $r_{NH-O} = 1.7-2.3$ Å, were used for each hydrogen bond identified (Kraulis et al., 1989).

A total of 727 interproton distance constraints was determined, which included 216 intraresidue, 200 sequential (for residues i and j, $j - i = 1$), 65 short range ($j - i < 5$), 196 long range ($j - i \geq 5$), and 50 constraints on the basis of 25 hydrogen bonds. These distance constraints were obtained in two stages. In the first stage, only well-resolved and unambiguously assigned NOEs were converted into distance constraints. A total of 510 distance constraints was thus identified, including 216 intraresidue, 244 interresidue, and 50 distance constraints on the basis of hydrogen bonds. These constraints along with 111 dihedral angle constraints were used to compute the preliminary structure. Details of this are described in the structure computation subsection. With the preliminary structure available, additional NOEs could now be assigned with confidence in the second stage. NOEs from overlapping cross peaks with ambiguous intensities were assigned to the very weak class and the upper boundary set to 5 Å.

Structure Computation. Using the NMR constraints, solution structures were computed by means of a distance geometry (DG) and simulated annealing (SA) algorithm and were refined by means of a distance geometry/structure refinement program, X-PLOR (version 3.1; Brünger, 1992). Structures were displayed and analyzed, using the QUANTA program. The protocol used in the structure computation process includes three basic steps, i.e., substructure embedding and regulation, full-structure SA-regulation, and SA refinement. In the first step, X-PLOR translates covalent geometry from X-PLOR parameter files into interatomic distance constraints, adds experimentally derived constraints, and creates a matrix of upper and lower bounds on the distances between all atoms of the protein, and then a subset of atoms, including C, C_α, $C_\alpha H$, N, NH, C_β, and C_γ, is embedded. The resulting structures are then minimized against an effective energy term (DG) that represents all upper and lower bounds in the distance geometry matrix, with all the other energy terms (van der Waals, bond, angle, torsional angle, improper torsional angle, electrostatic and experimental constraint) turned off. A total of 30 structures were generated at this stage.

The structures generated at this stage only contained a subset of atoms and also required extensive regulation. The missing atoms were added to the molecule, one residue at a time, by best fitting to the model structure (Nilges et al., 1988). These structures were then regularized by a multistage minimization protocol with a variable target function and followed by a short simulated annealing (Brünger, 1992). Both images were regulated at this stage, and the one with a lower improper energy was subjected for further refinement. The regularized structures were further refined with the SA method (Nilges et al., 1988). Nonbonded terms, such as electrostatic, hydrogen bond, and van der Waals energy terms, were replaced by a simple van der Waals repulsion term (Nilges et al., 1988). The initial force constants were set to 50, 200, and 500 kcal/(mol rad^2) for improper, angle, and torsional angle constraints, respectively, and 1000, 0.002, and 50 kcal/(mol Å^2) for bond, repulsion, and distance constraints, respectively. For dynamics calculations, the Verlet integrations of motion were integrated in steps of 1 fs. After an initial 500 steps of Powell energy minimization, the initial velocities were randomly assigned at 1000 K according to a Maxwell distribution. After 4 ps, the van der Waals repulsion constant was increased by multiplying its value by 1.5 every 75 fs, until it reached 4.0 kcal/(mol Å^2). The force constants for angle and improper terms were then set at 500 kcal/(mol rad^2), and the system was integrated for another 4 ps at 1000 K. In the cooling stage, system temperature was reduced by 25 K until it reached 300 K. Finally, the structure was further minimized by 500 cycles of restrained energy minimization.

All 30 structures generated have the same global folding pattern. The RMSD of all 30 structures from their mean structure is only 0.69 Å for backbone atoms and 1.13 Å for all heavy atoms, for residues from Cys3 to Thr40 and from Asn49 to Ile67. Two N-terminal residues, a C-terminal residue and a loop region from Pro41 to Cys48, are flexible compared to other regions and have an RMSD of 2.7 Å. Among the 30 structures, four structures had distance constraint violations greater than 0.5 Å and torsional angle

constraint violations greater than 5°. Four structures had a slightly higher average RMSD and had slightly higher constraint and covalent energy violations compared to the other structures. These eight structures were regarded as possible local minima due to flexibility and were thus excluded from further consideration. The 22 structures which had no distance violations greater than 0.2 Å and torsional angle violations greater than 1.5° were selected for analysis.

A mean structure, designated ⟨SA⟩, was obtained by best-fitting the 22 best SA structures to each other and averaging their coordinates (Driscoll et al., 1989). This mean structure had covalent structure violations and bad van der Waals contacts resulting from averaging of coordinates and was thus refined with the simulated annealing calculations, as described above. The refined mean structure, together with the NMR constraints, has been deposited with Brookhaven Protein Data Bank (accession code 1TIN).

RESULTS

Sequential Resonance Assignments. The sequential proton resonance assignments were made, using standard 2D NMR methods, on the basis of through-bond and through-space connectivities (Wüthrich, 1986). The through-bond connectivities were made via DQF-COSY, P.E.COSY, and TOCSY experiments recorded at different mixing times; the through-space connectivities were established by means of NOESY experiments performed with different mixing times. For short side-chain residues, such as alanine, aspartic acid, valine, leucine, and isoleucine, through-bond connectivities were straightforward. For long side-chain residues, such as lysine, arginine, and proline, through-bond connectivities were made in a progressive manner. Sequence-specific assignments of CMTI-V are illustrated in a TOCSY map in Figure 2. For proline residues, relayed through-bond connectivities between α- and δ-hydrogens were made in a TOCSY map obtained with a long mixing time (70 ms; Figure 3A). Direct through-bond connectivities from α-hydrogen to β-hydrogens and from δ-hydrogens to γ-hydrogens were then made in a short mixing time TOCSY (10 ms; Figure 3B) or a DQF-COSY map (not shown). This assignment process is illustrated for some proline connectivities in Figure 3A: The relayed through-bond connectivities from α- to δ-hydrogens are seen for all the six prolines. The direct through-bond connectivities from α- to β-hydrogens and from δ- to γ-hydrogens are illustrated for Pro41 in Figure 3B. Lysine and arginine side-chains were assigned in a similar manner. Relayed cross peak from α- to δ-hydrogens for arginines and from α- to ϵ-hydrogen for lysines were first identified in a 70 ms TOCSY map. Direct through-bond connectivities—α−β, δ−γ (for arginine), and ϵ−δ (for lysine)—were then made in a 10 ms TOCSY map. These direct connectivities were confirmed in a DQF-COSY or P.E.COSY map. Resonances were assigned for all the main-chain hydrogens and almost all of the side-chain hydrogens. Sequence-specific and stereospecific assignments of CMTI-V are given as Supplementary Material. A summary of all the NOEs observed is shown in Figure 4.

Structure Statistics. Statistical details for the best 22 SA structures and the refined mean structure are given in Table 1. All 22 SA structures and the minimized mean structure display small deviations from idealized covalent residue geometry and NMR constraint violations and have good nonbonded contacts, as indicated by small values for the repulsive force and large negative values for van der Waals energies. All the residues lie within the energetically favorable regions, as indicated in a Ramachandran plot (Figure 5).

Superposition of main-chain atoms (C, C_α, and N atoms) of the 22 SA structures is shown in Figure 6. The conformation of all main chain atoms for the 22 SA structures is well-conserved, except for the first two N-terminal residues and the last C-terminal residue. A loop region from residues 41−48 displays higher RMSD compared with that of other regions. This is reflected by the small average RMSD of 0.58 Å of each SA structure from their average structure, $\langle SA \rangle^2$. The distribution of RMSD values as a function of residue number is shown in Figure 7. Average values and standard deviations of ϕ, ψ, and χ_1 angles are displayed in Figure 8. The segment containing residues 15−18 displays slightly higher RMSD values (Figure 7). The flexibility of this segment is due to the backbone rotation of Gly16 and Gly17. This is better reflected by the standard deviations of ϕ and ψ in Figure 8. The ϕ and ψ angles of Gly16 jump between 99 ± 12° and −57 ± 5° and between 120 ± 12° and −103 ± 6°, respectively; Gly 17 also lies in two conformational states with ϕ and ψ angles jumping between 60 ± 5° and −67 ± 5° and between −158 ± 17° and −46 ± 10°, respectively.

Most of the side-chain atoms are well-defined. The RMSD for all heavy atoms of residues 3−40 and 49−67 is only 1.01 Å. This is also indicated by the RMSDs of the χ_1 angles (Figure 8) and by the superposition of all heavy atoms of the only α-helix (Figure 9) and two strands of a parallel β-sheet along with two β-turns (Figure 10) present in the 22 SA structures of CMTI-V. Side-chains displaying higher RMSDs are located in the loop region or on the molecular surface. Most of these are charged residues, such as lysine, arginine, or aspartic acid.

DISCUSSION

Global Folding in Solution. The solution structure of CMTI-V consists of a rigid scaffold region and a flexible loop region, as do serine proteinase protein inhibitors in general (Bode & Huber, 1992). The rigid body region consists of a major α-helix and three β-sheets (two antiparallel and one parallel) connected by a flexible loop and four turns. The flexible loop, also called the binding loop, is made up of residues 41−48.

The α-helical region, extending from Ser18 to Asn28, is identified qualitatively by distinctive patterns of NOEs and quantitatively by their ϕ and ψ torsional angles (average values in the CMTI-V helix are −64 and −37°, respectively).

The β-sheets are qualitatively identified on the basis of typically strong NOEs from the amide NH of one residue to the α-hydrogen of the preceding one, large $^3J_{NH\alpha}$ coupling constant (>8 Hz) and NOEs between two strands of the β-sheet. The final computed structures are consistent with the experimental data. The average ϕ and ψ torsional angles for the β-sheet residues are −116 ± 25° and 137 ± 25°, respectively. The three pairs of β-sheet elements are residues 7−9 and residues 62−60 antiparallel to residues 67−65 and residues 54−56, respectively, and residues 32−38 parallel to residues 51−56.

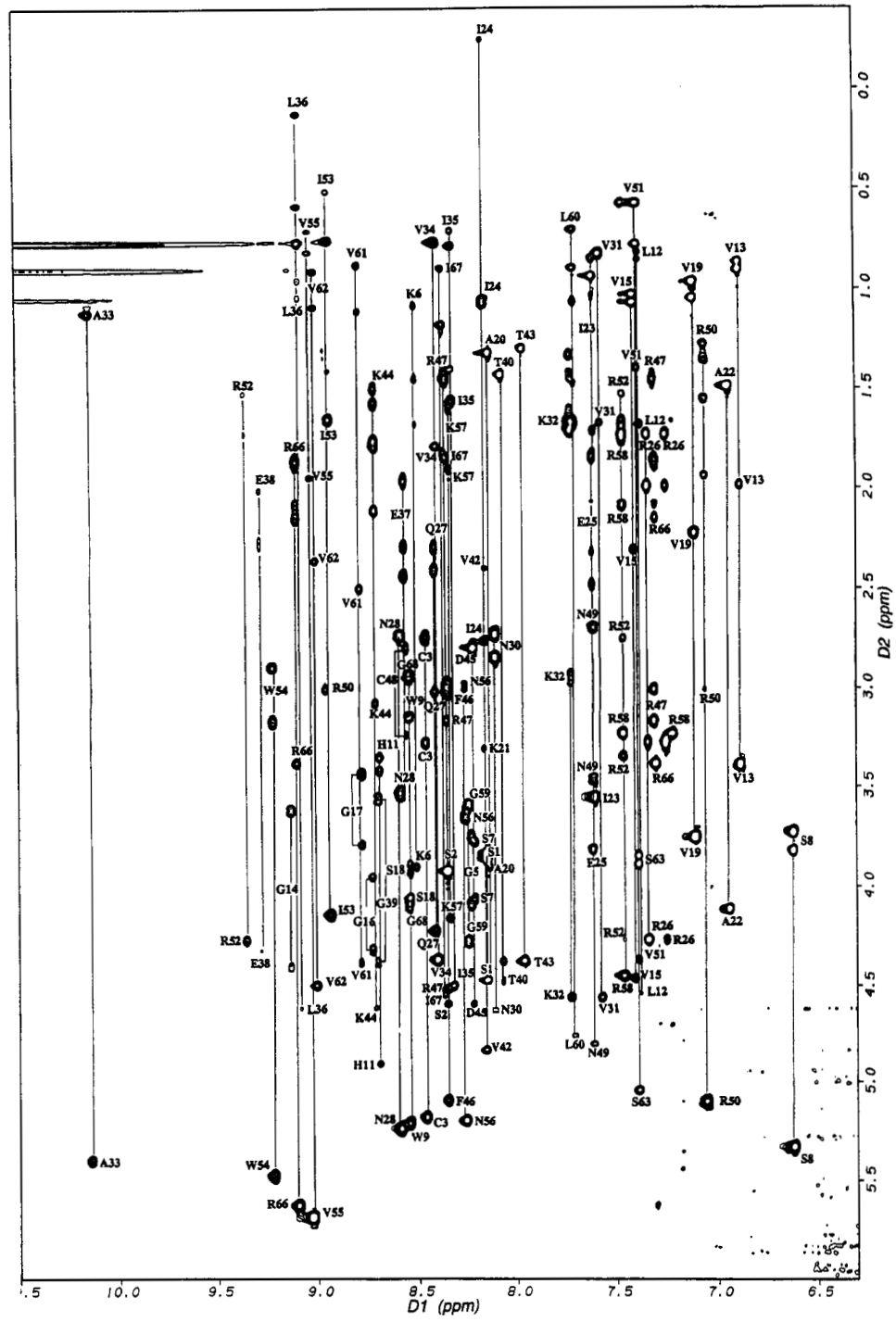

FIGURE 2: Portion of the 600 MHz TOCSY map of CMTI-V recorded at 30 °C, using a mixing time of 70 ms. Sequence-specific assignments are indicated. The through-bond connectivities from amide hydrogen to side-chain hydrogens are shown by lines connecting the cross peaks. The empty squares represent cross peaks observed in a 70 ms mixing time TOCSY experiment at 40 °C.

A type II turn is formed by residues 12−15, with average ϕ and ψ torsional angles of -57 ± 7 and $129 \pm 13°$ for Val13 and $90 \pm 26°$ and $3 \pm 13°$ for Gly14, respectively. Strong NOE_{NN} from Val15 to Gly14, $NOE_{N\alpha}$ from Gly14 to Val13, medium intensity $NOE_{N\alpha}$ from Val15 to Val13, and slowly exchangeable N*H* of Val15 are all consistent with a type II turn (Wüthrich, 1986).

Two type I turns from residues 28−31 and 47−50 are identified with average ϕ and ψ torsional angles, respectively, of $-59 \pm 1°$ and $-8 \pm 6°$ for Pro29, $-93 \pm 8°$ and $0 \pm 14°$ for Asn30, $-59 \pm 6°$ and $-21 \pm 13°$ for Cys48, and $-80 \pm 9°$ and $17 \pm 11°$ for Asn49. Strong NOE_{NN} from Val31 to Asn30, weak $NOE_{N\alpha}$ from Asn30 to Pro29, and slowly exchangeable N*H* of Val31 are all consistant with a type I turn for residues 28−31; strong NOE_{NN}s from Asn49 to Cys48 and from Arg50 to Asn49, slowly exchangeable N*H* of Arg50, and a small $^1J_{N\alpha}$ (<5.5 Hz) of Cys48 are together consistent with a type I turn for residues 47−50.

Another turn, consisting of residues 57−60, is identified as type III, with average ϕ and ψ torsional angles, respectively, of -47 ± 2 and $-24 \pm 6°$ for Lys57 and $-49 \pm 4°$ and $-15 \pm 4°$ for Arg58. Strong NOE_{NN}s from Gly59 to Arg58 and from Leu60 to Gly59 and a small $^3J_{Na}$

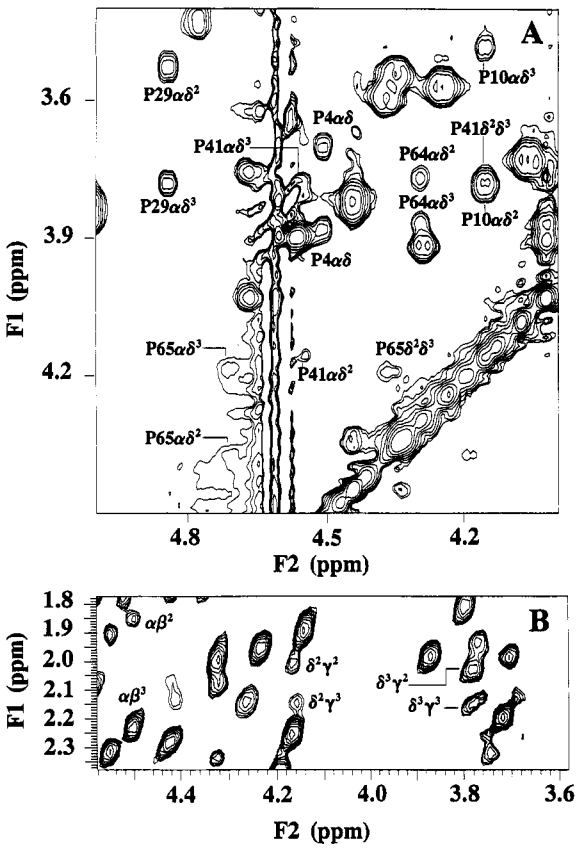

FIGURE 3: (A) Portion of the 500 MHz TOCSY map of CMTI-V recorded at 40 °C, using a mixing time of 70 ms. Relayed through-bond connectivities from α- to δ-hydrogens are shown for all the six proline residues. Cross peaks for Pro65 and Pro41 appear weak, as they occur proximal to the water signal. They do appear strong at other temperatures due to movement of the water signal. (B) Portion of the 500 MHz TOCSY map of CMTI-V recorded at 30 °C and a mixing time of 10 ms, showing the direct through-bond connectivities for Pro41.

(<5.5 Hz) of Arg58 are consistent with the type III turn computed.

Trp9. Trp9 appears to be an important residue, as it confers NOEs on every region of the inhibitor molecule, except the reactive-site loop (Figure 11): Strong NOEs from its ring hydrogens to residues in two of the three β-sheets (residues 67, 66, 65, and 51) and to residues in the α-helix (residues 24, 27, and 28) are observed. The average side-chain RMSD of the individual SA structures from their average structure for Trp9 is only 0.29 ± 0.11 Å, thus indicating an ordered side-chain conformation for this residue. The average distances from Trp9 N_π and NH_π to the backbone oxygen of Ile23 (a residue in the middle of the α-helix) are 3.24 ± 0.15 Å and 2.27 ± 0.15 Å, respectively, in the calculated 22 SA structures, which were obtained without using these hydrogen bond constraints. Furthermore, no hydrogen-bond term or electrostatic interaction term was included in the force field used. Thus, there is a strong indication of the presence of a hydrogen bond between Trp9 and Ile23, which likely holds together the N-terminal segment and the α-helix.

Side-Chains of Lys6, Lys21, Arg50, and Arg52. The side chains of lysine and arginine residues are positively charged under physiological conditions. They normally occur on the protein molecular surface for favorable interactions with polar solvent molecules and are usually freely rotating in solution.

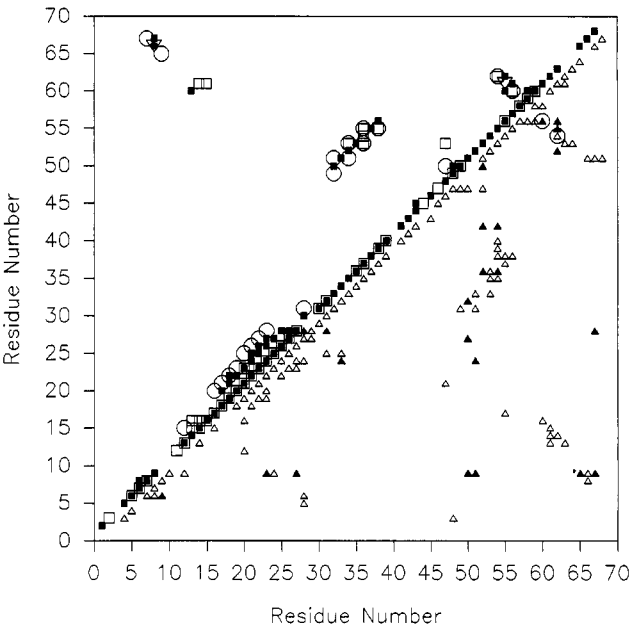

FIGURE 4: Summary of distance constraints used in the solution structure calculations of CMTI-V. Backbone N−N, N−α, and α−α through-space connectivities are represented by large empty squares, small filled squares, and large empty triangles (tip down), respectively, on the top half of the diagonal. N-side-chain and α-side-chain connectivities [empty triangles (tip up)] and side-chain−side-chain connectivities (filled triangles) are presented on the bottom half of the diagonal. Hydrogen bond connectivities are represented by large empty circles on the top half of the diagonal. Residues with slowly exchangeable amide hydrogens and $^3J_{NH\alpha}$ > 9 Hz are W9, V34, L36, R50, V51, I53, W54, V55, N56, and L60, and those with slowly exchangeable amide hydrogens and $^3J_{NH\alpha}$ < 5.5 Hz are A20, K21, A22, I23, I24, E25, R26, and Q27.

Lys6, Lys21, Arg50, and Arg52 side chains, however, appear to have rigid conformations in solution—stereospecific assignments of these residues (Cai et al., 1995a) indicate that their side-chains each have only a single rotamer at χ_1, χ_2, and χ_3, except that of Arg50, whose rotamers at χ_3 could not be determined due to chemical shift degeneracy of the δ-hydrogens. For Lys21, a single rotamer was identified at χ_4 and the ε-hydrogens were stereospecifically assigned (Cai et al., 1995b). For Arg50 and Arg52, side-chain NH_2 hydrogens were observed in NMR spectra, indicating that these hydrogens were slowly exchanging with the D_2O solvent and were likely hydrogen-bonded. Furthermore, stereospecific assignments and χ_4 torsional angle determination of Arg52 (Cai et al., 1995c) indicate a fixed side-chain conformation and hence its involvement in hydrogen-bonding interactions. For Arg50, there are not enough experimental constraints obtained so as to define well its side-chain conformation, and, therefore, no conclusion could be made as to its hydrogen-bonding partner. A model-building effort by rotating the side chains of Asp45 and Arg50 indicated that the side-chain oxygen of Asp45 and side-chain NH_2 of Arg50 could form a hydrogen bond without causing bad van der Waals contacts or any other experimental constraint violations.

In the calculated solution structures, the average distance from the two terminal side-chain nitrogens of Arg52 to the side-chain oxygen of Thr43 is 5.0 Å; from Lys6 terminal nitrogen to Gln27 side-chain oxygen is 3.9 Å; and, from Lys21 terminal nitrogen to Ala33 backbone oxygen is 4.0 Å. Considering the fact that electrostatic and hydrogen bond

Table 1: Structure Statistics for CMTI-V

Distance Constraints			
		rms deviations from experimental distance constraints (Å)	
total	distance	SA[a]	$\langle SA \rangle_r$[b]
727	all	0.0090 ± 0.0015	0.0094
216	intraresidue	0.0097 ± 0.0011	0.0117
200	sequential ($\|i-j\| = 1$)	0.0090 ± 0.0030	0.0076
65	short range ($\|i-j\| \leq 5$)	0.0046 ± 0.0027	0.0059
196	long range ($\|i-j\| > 5$)	0.0136 ± 0.0065	0.0086
50	hydrogen bond	0.0132 ± 0.0040	0.0116

Torsional Angle Constraints			
		rms deviations from experimental distance constraints (deg)	
total	angle	SA[a]	$\langle SA \rangle_r$[b]
111	all	0.220 ± 0.010	0.216
49	ϕ angles	0.015 ± 0.027	0.007
62	χ angles	0.297 ± 0.018	0.296

Energies (kcal/mol)[c]		
	SA[a]	$\langle SA \rangle_r$[b]
F_{noe}	3.00 ± 1.02	3.23
F_{tor}	0.84 ± 0.12	0.81
F_{repel}	4.13 ± 0.32	3.23
F_{L-J}	−346 ± 7	−338

RMS Deviations from Idealized Geometry		
	SA[a]	$\langle SA \rangle_r$[b]
bond (Å)	0.0075 ± 0.0002	0.0070
angle (deg)	2.0880 ± 0.0043	2.0920
impropers (deg)	0.1072 ± 0.0245	0.1270

RMS Deviations of 22 SA's and Their Mean Structure (Å)	
whole protein	
main chain	1.25 ± 0.25
all heavy atoms	1.62 ± 0.22
residues 3−40 and 49−67	
main chain	0.58 ± 0.06
all heavy atoms	1.01 ± 0.07
loop region (residues 41−48)	
main chain	1.88 ± 0.59
all heavy atoms	2.75 ± 0.61

[a] SA represents the 22 individual structures refined by the simulated annealing method, and $\langle SA \rangle_r$ represents the refined mean structure, as described in the text. [b] Root mean squared deviations in angstroms when the calculated final distance between hydrogens in the structure after simulated annealing refinements exceeds the boundaries of distance constraints and torsional angle constraints, averaged over all constraints of different constraints groups listed in the table. [c] The force constants used for these calculations are 50 kcal/(mol Å^2) for distance constraints, 500 kcal/(mol rad^2) for torsional angle constraints, and 4 kcal/(mol Å^2) for repulsion term. F_{noe}, F_{tor}, and F_{repel} are, respectively, the constraint violation energies associated with distance constraints and torsional angle constraints and the van der Waals repulsion energy with hard-sphere van der Waals radii set to 0.8 times the standard values used in the CHARMm empirical energy function (Brooks et al., 1983). E_{L-J} is the van der Waals energy recalculated with the same coordinates, but with the electrostatic terms included.

terms are not included in the force field used for structure calculation, these distances strongly favor hydrogen bonds between the mentioned atom pairs.

Another portion of experimental data strongly suggests the presence of a hydrogen bond between the side chains of Gln27 and Lys6. Generally, the two side-chain amide hydrogens of glutamine and asparagine residues show very strong NOEs to each other and to their own γ- or β-hydrogens; a cross peak between these two hydrogens appears in TOCSY maps of appropriate mixing times. This has been

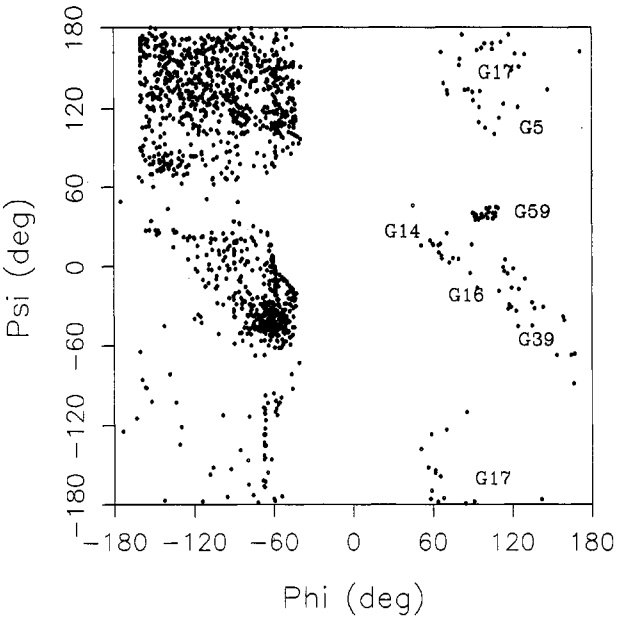

FIGURE 5: Ramachandran plot of 22 structures of CMTI-V determined by the dynamic simulated annealing method (SA structures).

found to be the case for all glutamine and asparagine residues in CMTI-V with the exception of Gln27, whose side-chain amide hydrogens were not observed in any of the data sets collected. This may be due to their fast exchange with solvent molecules. The fast exchange rate of Gln27 side-chain amide hydrogens may be explained by invoking the presence of a hydrogen bond between its side-chain oxygen atom and the NH_3^+ group of Lys6. This hydrogen bond would render Gln27 amide hydrogens labile, and hence their exchange with solvent deuterons is facilitated. Structures calculated with the constraint that terminal nitrogen of Lys6 and Lys21 lie in the range of 2.4−3.3 Å from the side-chain oxygen atom of Gln27 and the main-chain oxygen of Ala33, respectively, and energy-minimized in 500 steps show no constraint violations for the newly introduced constraints and cause no increase in violations for all other experimental constraints or van der Waals energy term. If, on the other hand, the distances between the side-chain nitrogens of Arg52 and the side-chain oxygen of Thr43 are constrained, distance and torsional angle constraint violations increase for some SA structures.

As found in other members of serine proteinase inhibitors, the binding loop appears to be anchored to the scaffold region by Arg50 and Arg52 in CMTI-V by hydrogen-bonding and electrostatic interactions. The side chain of Arg52 and also, most likely, that of Arg50 extend from the β-sheet toward the reactive-site loop, providing the stabilizing contacts between the flexible loop and the rigid body of the inhibitor, as found for CI-2 (McPhalen & James, 1987). The hydrogen-bonding interactions between the side chains of Lys6 and Glu27 help pull the N-terminal segment and the α-helix together; those between the side chain of Lys21 and the main-chain oxygen of Ala33 seem to provide a contact between the α-helix and β-sheet. As a whole, these hydrogen-bonding interactions appear to be important in stabilizing the whole inhibitor molecule.

Residues Showing Unusual Chemical Shifts. An inspection of chemical shifts of assigned peaks of CMTI-V (Supplementary Material) reveals that several residues display

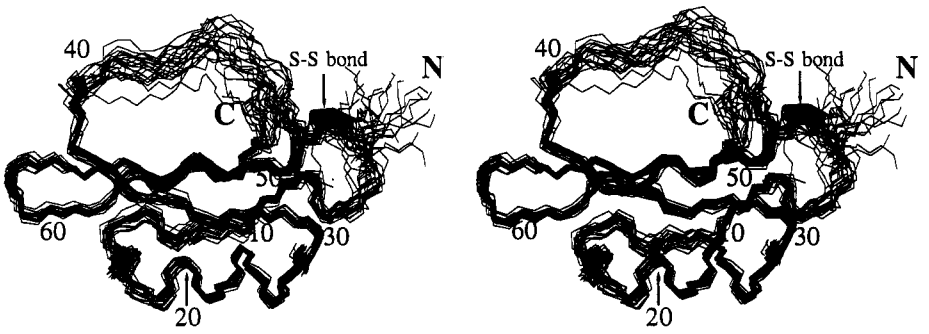

FIGURE 6: Stereoview of the best-fit superposition of 22 SA structures of CMTI-V. The main chain atoms (N, C_α, and $C_{C=O}$), including the Cys3−Cys48 disulfide bridge, are shown.

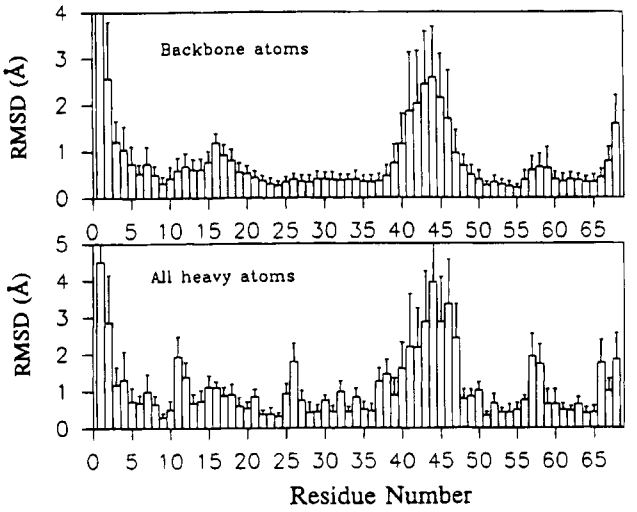

FIGURE 7: Atomic RMSD distribution of the individual SA structures about the mean SA structure computed for CMTI-V. The filled bars indicate the average RMSD values of 22 SA structures, and the standard deviation is indicated by the vertical error bar.

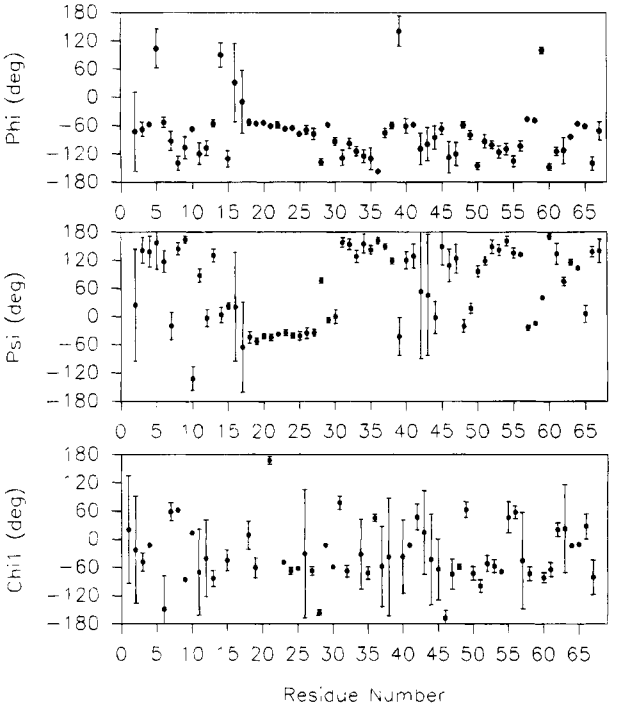

FIGURE 8: Angular RMSD distribution of backbone ϕ, ψ angles and side-chain χ_1 angles for CMTI-V. The average angle values of 22 SA structures are represented by filled circles, and the standard deviations are indicated by vertical error bars.

unusual chemical shifts. Residues that possess significantly upfield-shifted peaks are as follows: Lys6 (−0.03 ppm, $H_{\gamma 2}$; 1.70 ppm, H_ϵ); Val13, Lys21, and Ile23 (H_αs at 3.39, 3.31, and 3.54 ppm, respectively); Ile24 (2.76 ppm, H_α; 1.06 ppm, H_β; −1.70 and 0.61 ppm, H_γs; −0.25 and 0.03 ppm, γ- and δ-methyl groups, respectively); Ile35 (0.78 ppm, $H_{\gamma 3}$); Leu36 (0.78 ppm, $H_{\beta 3}$; 0.95 ppm, $H_{\beta 2}$); Arg50 (2.95 ppm, side-chain NH_2); and Ile53 (0.27 ppm, $H_{\gamma 3}$). Residues that possess significantly downfield-shifted peaks are as follows: Ala33 (10.14 ppm, NH; 5.41 ppm, H_α); Arg52 (9.57 ppm, side-chain NH_2); and Arg66 (5.63 ppm, H_α).

An examination of the three-dimensional solution structure of CMTI-V reveals that the unusual chemical shifts observed may be explained by invoking the ring-current effects of the neighboring aromatic residues: hydrogen atoms responsible for the upfield-shifted resonances of Lys6, Val13, Lys21, Ile23, Ile24, Ile35, and Ile53 are found to be located either below or above the ring-plane of Trp9, and those of Leu36 are located below the Trp54 ring plane. The side-chain NH_2 group of Arg50 is upfield-shifted, most likely, by the ring current of Phe46. As it is located in the flexible loop region, the Phe46 ring conformation is not experimentally well-constrained. The computed structures of CMTI-V indicate that the NH_2 group of Arg50 is proximal to the Phe46 ring but are ambiguous as to whether it is above or below the ring plane. The downfield-shifted side-chain NH_2 group of Arg52 is found to be located in the plane of Trp54 aromatic rings. Similarly, the downfield-shifted H_α of Arg66 is located in the ring plane of Trp9. In the case of Ala33, the observed downfield shifts of its NH and H_α resonances may be attributed to the likely hydrogen bond between its amide hydrogen atom and the side-chain oxygen of Glu25, as the distances from one of the side-chain oxygens of Glu25 to the amide nitrogen and amide hydrogen of Ala33 are determined to be 3.2 and 4.3 Å, respectively, in the energy-minimized average structure of CMTI-V.

Comparison between the Solution Structure of CMTI-V and the Crystal Structure of CI-2. The reported alignment of primary structures of CI-2 and CMTI-V (Krishnamoorthi et al., 1990; Figure 1) indicates about 62% differences in the sequences. Yet the average solution structure computed for CMTI-V shows a high degree of similarity to the high-resolution crystal structure determined for CI-2 (McPhalen & James, 1987; Brookhaven Protein Data Bank accession code 2ci2). Both proteins have very similar folding patterns, as illustrated by ribbon diagrams (Figure 12A) and main chain superposition of the two protein structures (Figure 12B) for the aligned sequences, Pro4−Gly68 in CMTI-V and Asn19−Gly83 in CI-2 (Krishnamoorthi et al., 1990; Figure 1).

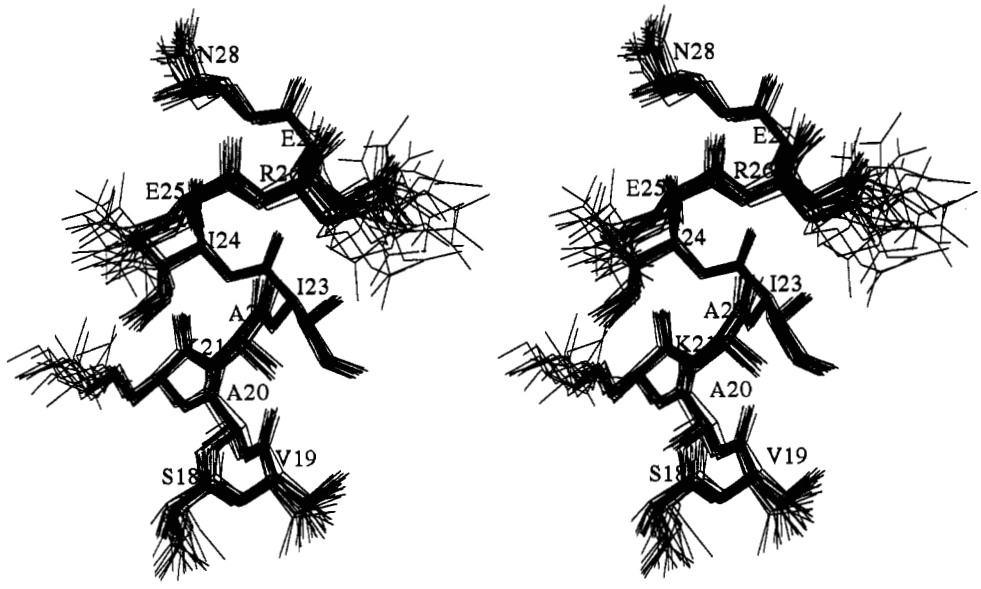

FIGURE 9: Stereoview of the α-helix, from Ser18 to Asn28, of CMTI-V. Only heavy atoms are drawn.

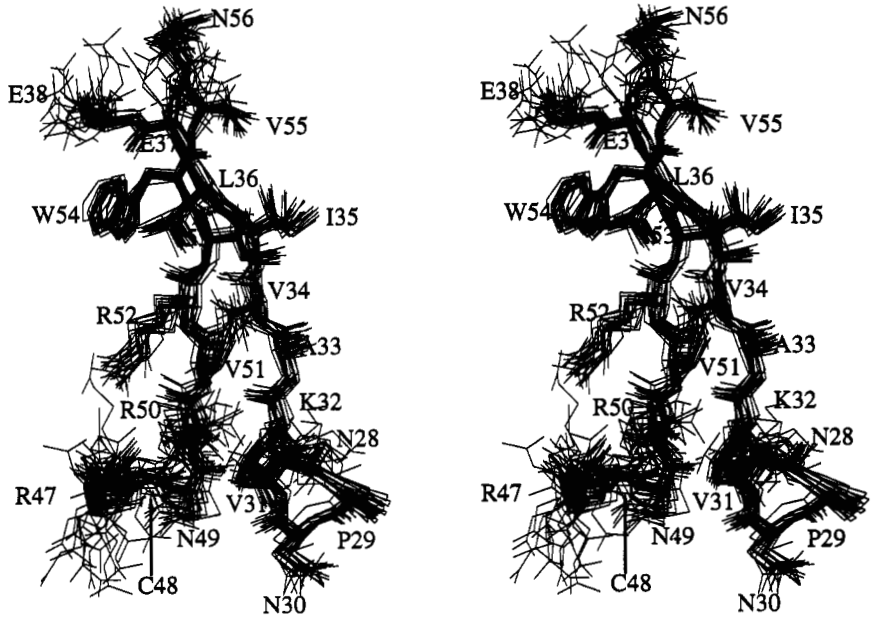

FIGURE 10: Stereoview of a parallel β-sheet, consisting of residues 50−56 and 32−38, and two type I turns, consisting of residues 47−50 and 28−31, of CMTI-V. Only heavy atoms are drawn.

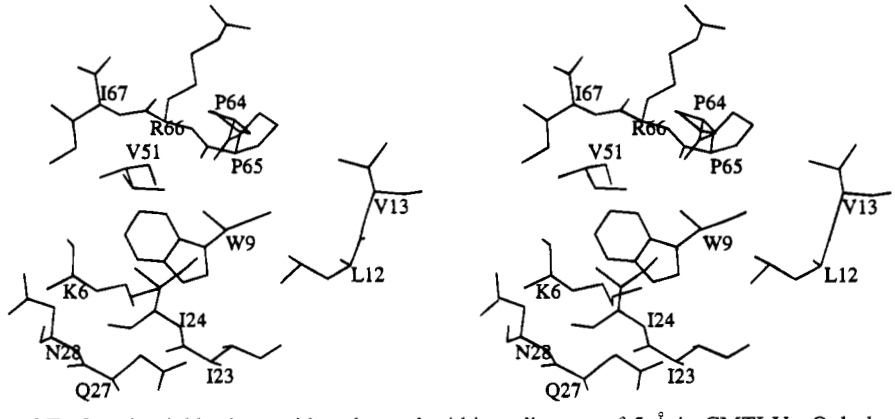

FIGURE 11: Stereoview of Trp9 and neighboring residues located within a distance of 5 Å in CMTI-V. Only heavy atoms are drawn.

The major secondary structures in both proteins are highly conserved: the major α-helix containing residues 31−43 in CI-2 (McPhalen & James, 1987) is equivalent to the α-helix in CMTI-V; the β-sheet structures in CI-2, a parallel one comprising residues 47−53 and 64−71, and two antiparallel ones—one consisting of residues 22−24 and 82−80 and the

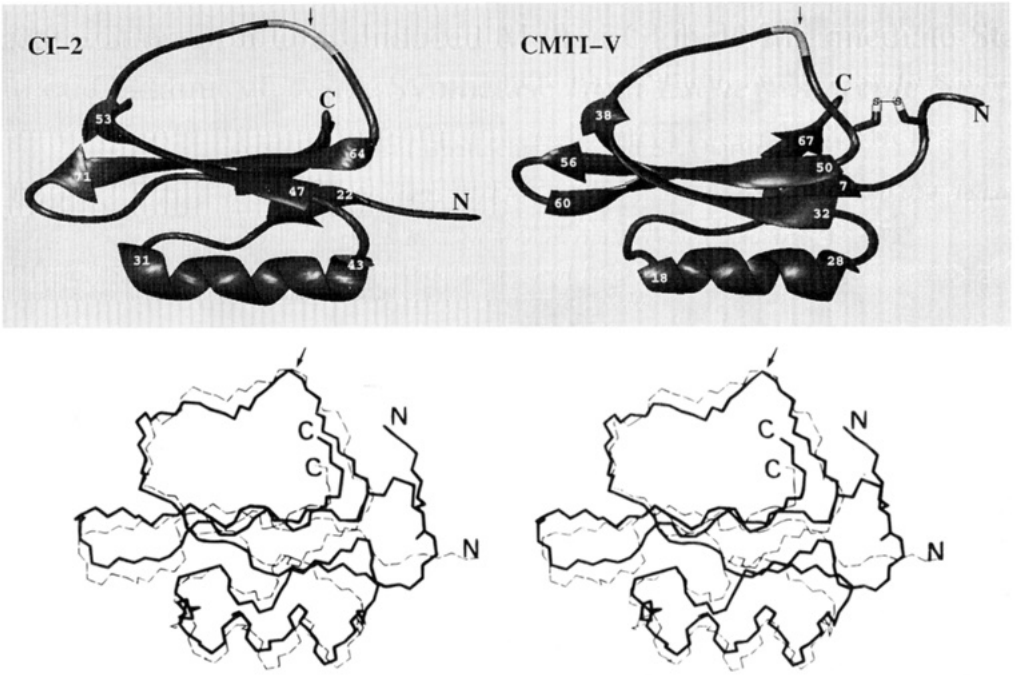

FIGURE 12: (A, top) Comparison of ribbon diagrams of the crystal structure of CI-2 (McPhalen & James, 1987) and the refined average solution structure of CMTI-V. (B, bottom) Stereoview of superposition of the crystal structure of CI-2 (broken thin line) and refined average solution structure of CMTI-V (solid thick line). Only main-chain atoms are drawn.

other consisting of residues 69−71 and 77−75—are also similar to the β-sheet structures characterized in CMTI-V.

Three β-turn structures in CI-2—a type II turn comprising residues 27−30 and two type I turns consisting of residues 62−65 and 71−74—remain unchanged in the equivalent positions in CMTI-V. The average RMSD of all main chain atoms between the energy-minimized average structure of CMTI-V from residues 12−56 and the crystal structure of CI-2 from residues 27−71 is 1.25 Å. This number includes differences imposed by amino acid substitutions between the two proteins. Relative packing of side-chains can be different due to substitutions, and this can lead to a change in the position of occurrence of a secondary structural element.

The major differences between the solution structure of CMTI-V and the crystal structure of CI-2 are due to the lone Cys3−Cys48 disulfide bridge in CMTI-V, which does not exist in CI-2. In the case of CI-2, a type III turn, made up of residues 24−27, orients the N-terminal segment away from the reactive-site loop. On the other hand, in the case of CMTI-V, the disulfide bond pulls the N-terminal segment closer to the binding loop (Figure 12A). This leads to drastic changes in dihedral angles for Pro10 and His11 in CMTI-V. The corresponding residues in CI-2 are Pro25 and Glu26, with the ϕ and ψ angles of −52° and −32°, respectively, for the former and −69° and −15°, respectively, for the latter. In the solution structure of CMTI-V, the corresponding dihedral angles change to −43° and 138°, respectively, for Pro10 and −100° and 133°, respectively, for His11. It is to be noted that three amino acid substitutions occur between the two proteins (Leu73, Asp74, and Asn75 in CI-2 corresponding to Arg58, Gly59, and Leu60, respectively, in CMTI-V) and the swap between aspartic acid and glycine is drastic (in both size and charge). As a result of the presence of a disulfide bridge in CMTI-V, the N-terminal segment consisting of residues 7−9 forms an antiparallel β-sheet with residues 67−65, a structural property unique to CMTI-V.

Another major difference between the crystal structure of CI-2 and the solution structure of CMTI-V occurs in the C-terminal segment: CI-2 has a type I turn consisting of residues 71−74, whereas CMTI-V has a type III turn consisting of residues 57−60.

The positively charged side chains (Lys21, Lys36, Arg65, and Arg67) and Trp24 in CI-2 show some similarities and differences, as compared to the equivalent residues in CMTI-V—Lys6, Lys21, Arg50, Arg52, and Trp9 (Figure 1). Arg65 and Arg67 in CI-2, which are equivalent to Arg50 and Arg52, respectively, in CMTI-V, extend toward the binding loop from the β-sheet, providing most of the electrostatic and hydrogen-bonding interactions between the loop and the β-sheet (McPhalen & James, 1987). These electrostatic and hydrogen-bonding interactions appear to be preserved in CMTI-V. In CI-2, the indole NH of Trp24 forms a hydrogen bond with the side chain of Asp42 (distances from one of the side-chain oxygen of Asp42 to N_π and NH_π of Trp24 are 2.7 and 1.7 Å, respectively), while, in the case of CMTI-V, Trp9 indole NH forms a hydrogen bond with the main chain oxygen of Ile23, as discussed earlier. In CI-2, Lys21 appears to have electrostatic interactions with both the side-chain Glu26 and Asp42, distances between side-chain N_ζ and one of the side-chain oxygens of Glu26 and Asp42 being 2.8 and 2.7 Å, respectively; in CMTI-V, the equivalent residue, Lys6, electrostatically interacts with the side-chain of Gln27 (equivalent to Gln41 in CI-2). These differences concerning electrostatic and hydrogen-bonding interactions of tryptophan and lysine side chains are most likely due to the presence of the Cys3−Cys48 disulfide bond present only in CMTI-V.

In contrast to Lys21 in CMTI-V, which has a fixed side-chain conformation and forms a hydrogen bond with the main-chain oxygen atom of Ala33, the equivalent Lys36 in CI-2 extends to the solvent and is thus freely rotating. This observation is supported by the chemical shift data: the β-hydrogens of Lys21 (CMTI-V) resonate at 1.80 and 1.45

ppm, while those of Lys36 (CI-2) show chemical shift values of 1.75 and 1.73 ppm (Kjaer et al., 1987).

The different specificities of inhibition of serine proteases exhibited by CI-2 and CMTI-V—chymotrypsin and trypsin, respectively—arise out of the different P_1 residues located in the binding loops—Met59 in CI-2 and Lys44 in CMTI-V, although the binding loops appear to have similar conformations. In the case of CMTI-V, hydrolysis of the Lys44—Asp45 peptide bond results in the loss of the protein's inhibitory property toward factor XII_a. This strongly suggests a drastic change in the binding affinity of the nicked inhibitor toward that blood coagulation protein. Work is in progress to determine the three-dimensional solution structure of the reactive-site hydrolyzed CMTI-V and to characterize the attendant structural changes.

ACKNOWLEDGMENT

We thank Drs. Jianhua Liu and Om Prakash and Mr. Dave Manning for their technical assistance. We also thank the anonymous reviewers for their helpful comments and suggestions.

SUPPLEMENTARY MATERIAL AVAILABLE

One table of ^1H NMR chemical shifts of CMTI-V (4 pages). Ordering information is given on any current masthead page.

REFERENCES

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

Brünger, A. T. (1992) X-PLOR version 3.1, Yale University Press, New Haven and London.

Brooks, B. R., Bruccoleri, R. E., Olafson, R. E., State, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.

Cai, M., Bradford, E. G., & Timkovich, R. (1992) *Biochemistry 31*, 8603−8612.

Cai, M., Liu, J., Gong, Y., & Krishnamoorthi, R. (1995a) *J. Magn. Reson., Ser. B* (in press).

Cai, M., Huang, Y., Liu, J., & Krishnamoorthi, R. (1995b) *J. Biomol. NMR* (in press).

Cai, M., Gong, Y., & Krishnamoorthi, R. (1995c) *J. Magn. Reson., Ser. B 106*, 297−299.

Clore, G. M., Gronenborn, A. M., James, M. N. G., Kjaer, M., McPhalen, C. A., & Poulsen, F. M. (1987a) *Protein Eng. 1*, 313−318.

Clore, G. M., Gronenborn, A. M., Kjaer, M., & Poulsen, F. M. (1987b) *Protein Eng. 1*, 305−311.

Clore, G. M., Gronenborn, A. M., Nilges, M., & Ryan, C. A. (1987c) *Biochemistry 26*, 8012−8023.

Clubb, R. T., Ferguson, S. B., Walsh, C. T., & Wagner, G. (1994) *Biochemistry 33*, 2761−2772.

Driscoll, P. C., Gronenborn, A. M., Beress, L., & Clore, G. M. (1989) *Biochemistry 28*, 2188−2198.

Hyberts, S. G., Goldberg, M. S., Havel, T. F., & Wagner, G. (1992) *Protein Sci. 1*, 736−751.

IUPAC-IUB Commission on Biochemical Nomenclature (1970) *J. Mol. Biol. 52*, 1−17.

Kjaer, M., Ludvigsen, S., Sørensen, O. W., Denys, L. A., Kindtler, J., & Poulsen, F. M. (1987) *Carlsberg Res. Commun. 52*, 327−354.

Kraulis, P. J., & Jones, T. A. (1989) *Proteins 2*, 188−201.

Krishnamoorthi, R., Gong, Y., & Richardson, M. (1990) *FEBS Lett. 273*, 163−167.

McPhalen, C. A., & James, M. N. G. (1987) *Biochemistry 26*, 261−269.

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

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.

Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Gō, N., & Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611−639.

Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295−312.

Wüthrich, K. (1986) *NMR of Protein and Nucleic Acids*, John Wiley and Sons, New York.

Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949−961.

Zuiderweg, E. R. P., Boelens R., & Kaptein, R. (1985) *Biopolymers 24*, 601−611.

BI942673Y