Solution Structure and Backbone Dynamics of Recombinant *Cucurbita maxima* Trypsin Inhibitor-V Determined by NMR Spectroscopy[†,‡]

Jianhua Liu,[§] Om Prakash,[§] Mengli Cai,[§] YuXi Gong,[§] Ying Huang,[§] Lisa Wen,[||] Joanna J. Wen,[||] Jeng-Kuen Huang,[||] and Ramaswamy Krishnamoorthi*[,§]

Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506, and Department of Chemistry, Western Illinois University, Macomb, Illinois 61455

Received October 17, 1995[⊗]

ABSTRACT: The solution structure of recombinant *Cucurbita maxima* trypsin inhibitor-V (rCMTI-V), whose N-terminal is unacetylated and carries an extra glycine residue, was determined by means of two-dimensional (2D) homo and 3D hetero NMR experiments in combination with a distance geometry and simulated annealing algorithm. A total of 927 interproton distances and 123 torsion angle constraints were utilized to generate 18 structures. The root mean squared deviation (RMSD) of the mean structure is 0.53 Å for main-chain atoms and 0.95 Å for all the non-hydrogen atoms of residues 3−40 and 49−67. The average structure of rCMTI-V is found to be almost the same as that of the native protein [Cai, M., Gong, Y., Kao, J.-L., & Krishnamoorthi, R. (1995) *Biochemistry 34*, 5201−5211]. The backbone dynamics of uniformly ^{15}N-labeled rCMTI-V were characterized by 2D ^{1}H−^{15}N NMR methods. ^{15}N spin−lattice and spin−spin relaxation rate constants (R_1 and R_2, respectively) and {^{1}H}−^{15}N steady-state heteronuclear Overhauser effect enhancements were measured for the peptide NH units and, using the model-free formalism [Lipari, G., & Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559, 4559−4570], the following parameters were determined: overall tumbling correlation time for the protein molecule (τ_m), generalized order parameters for the individual N−H vectors (S^2), effective correlation times for their internal motions (τ_e), and terms to account for motions on a slower time scale (second) due to chemical exchange and/or conformational averaging (R_{ex}). Most of the backbone NH groups of rCMTI-V are found to be highly constrained ($\langle S^2 \rangle = 0.83$) with the exception of those in the binding loop (residues 41−48, $\langle S^2 \rangle = 0.71$) and the N-terminal region ($\langle S^2 \rangle = 0.73$). Main-chain atoms in these regions show large RMSD values in the average NMR structure. Residues involved in turns also appear to have more mobility ($\langle S^2 \rangle = 0.80$). Dynamical properties of rCMTI-V were compared with those of two other inhibitors of the potato I family−eglin c [Peng, J. W., & Wagner, G. (1992) *Biochemistry 31*, 8571−8586] and barley chymotrypsin inhibitor 2 [CI-2; Shaw, G. L., Davis, B., Keeler, J., & Fersht, A. R. (1995) *Biochemistry 34*, 2225−2233]. The Cys^{3}−Cys^{48} linkage found only in rCMTI-V appears to somewhat reduce the N-terminal flexibility; likewise, the C-terminal of rCMTI-V, being part of a β-sheet, appears to be more rigid.

Cucurbita maxima trypsin inhibitors (CMTIs)[1] are functionally important serine proteinase inhibitors in that they are also highly spccific inhibitors of the human blood coagulation protein, factor XIIa. Inhibitors belonging to the squash family (Hojima et al., 1982; Wieczorek et al., 1985) and the potato I family (Krishnamoorthi et al., 1990) have been isolated and characterized. Structural studies of CMTI-I (∼3 kDa) and CMTI-III (a natural variant, E9K, of CMTI-I) of the squash family and their reactive-site (Arg^{5}−Ile^{6}) hydrolyzed forms have been reported (Bode et al., 1989; Holak et al., 1989a,b; Krishnamoorthi et al., 1992a,b). Functional properties of these and other variant inhibitors have also been described in detail (McWherter et al., 1989; Otlewski et al., 1990; Wynn & Laskowski, 1990; Otlewski & Zbyryt, 1994; Hayashi et al., 1994). CMTI-V (∼7 kDa) of the potato I family shows no sequence homology with any of the squash inhibitors, yet it shows functional similarities (Krishnamoorthi et al., 1990). Determination of structures and dynamics of inhibitors of these two families will aid in understanding structure−function relationships. That information may be useful in efforts to design and develop highly specific and physiologically acceptable inhibitors of factor XIIa.

[†] This work was supported by grants from the National Institutes of Health (HL-40789 to R.K., HL-52235 to L.W.). R.K. is an NIH Research Career Development Awardee (1994−1999). The 11.75 Tesla NMR instrument at KSU was purchased with funds from an NSF-EPSCoR grant. This is contribution 96-224-J from the Kansas Agricultural Experiment Station.

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

* To whom correspondence should be addressed. E-mail: krish@ksuvm.ksu.edu. Phone: (913) 532-6262. Fax: (913) 532-7278.

[§] Department of Biochemistry, Kansas State University, Manhattan, KS 66506.

[||] Department of Chemistry, Western Illinois University, Macomb, IL 61455.

[⊗] Abstract published in *Advance ACS Abstracts,* February 1, 1996.

[1] Abbreviations: CMTI, *Cucurbita maxima* trypsin inibitor; rCMTI-V, recombinant *Cucurbita maxima* trypsin inibitor-V; CMTI-V*, reactive-site hydrolyzed *Cucurbita maxima* trypsin inibitor-V; NMR, nuclear magenetic resonance; NOE, nuclear Overhauser effect; HSQC, heteronuclear single-quantum coherence; HMQC, heteronuclear multiple-quantum coherence; TOCSY, total correlated spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; CPMG, Carr-Purcell-Meiboom-Gill; RMSD, root mean squared deviation; SA, simulated annealing; ppm, parts per million; CI-2, chymotrypsin inhibitor-2 from barley seeds.

0006-2960/96/0435-1516$12.00/0

CMTI-V consists of 68 amino acid residues, including a Cys^3-Cys^{48} bridge. Its reactive site is the peptide bond between Lys^{44} and Asp^{45} (Krishnamoorthi et al., 1990). The three-dimensional solution structure of CMTI-V has been determined (Cai et al., 1995a). It is similar to those of two other members of the family for which three-dimensional structures are known—barley chymotrypsin inhibitor 2 (McPhalen et al., 1987; Ludvigsen et al., 1991; Prat Gay et al, 1994) and eglin c (Hyberts & Wagner, 1990). Recently, functional and thermodynamic properties and the three-dimensional solution structure of reactive-site hydrolyzed CMTI-V (CMTI-V*) have been described (Cai et al., 1995b). Structural changes resulting in significant loss in inhibitory function have been characterized. They are confined to the binding loop and the N-terminal region, which become more flexible with an increase in entropy.

An inhibitor protein acts on its cognate enzyme by binding to its active site tightly in a reversible way (Laskowski & Kato, 1980), and this process is expected to depend on structural and dynamic properties of the enzyme as well as the inhibitor. Internal motions of a protein on different time scales, extending from picosecond to second, have been suggested to play an important role in its functions such as ligand-binding, enzyme catalysis, antibody recognition, etc. (Karplus & McCammon, 1983). However, correlation between atomic motion and protein function is not completely understood. It is useful to determine structural factors that influence dynamics as a means to probe the relationship between protein structure and dynamics, on the one hand, and protein structure and function, on the other. Herein we report the determination of internal dynamics of rCMTI-V from ^{15}N longitudinal and transverse relaxation times and $\{^1H\}^{15}N$ heteronuclear Overhauser effect (NOE) enhancements. The ^{15}N-labeled recombinant CMTI-V (rCMTI-V) utilized for this purpose differs from the native protein in that it has an extra Gly at the N-terminal residue (Huang et al., to be published) which, unlike that of the native protein (Krishnamoorthi et al., 1990), is unacetylated. The three-dimensional solution structure determination of rCTMI-V is also reported so that the dynamic properties of the protein could be reliably interpreted in structural terms.

MATERIALS AND METHODS

Proteins. rCMTI-V was overexpressed in *Escherichia coli*, isolated, and purified, as described previously (Wen et al., 1993; Krishnamoorthi et al., 1990). Uniformly ^{15}N-labeled, and $^{15}N/^{13}C$-doubly labeled rCMTI-V were prepared by growing the cells in media containing $^{15}NH_4Cl$ and ^{13}C-labeled glucose as the sole nitrogen and carbon source, respectively. The original recombinant version of CMTI-V (Wen et al., 1993) contains seven extra N-terminal residues (ARIRARG). The protein used in the present study was prepared after trypsin-affinity chromatography and was found by sequence analysis to contain only an additional N-terminal Gly residue (Huang et al., to be published). The residue numbering was kept consistent with that of CMTI-V, with negative numbers denoting the extra N-terminal residues in rCMTI-V. A sample of ~5 mM ^{15}N-labeled rCMTI-V in H_2O, pH 5.4, was prepared by dissolving the lyophilized protein in 0.5 mL of 90% H_2O/10% D_2O (v/v). Two other samples, ~5 mM unlabeled and ~2 mM $^{15}N/^{13}C$ doubly labeled protein, were prepared in 99.996% D_2O, pD 5.4. All samples also contained 50 mM KCl.

NMR Spectroscopy. 2D and 3D NMR experiments were done at 30 °C with a 11.75 T (499.496 MHz for ^1H) Varian UNITY*plus* instrument. Data sets were processed, plotted, and analyzed using the Varian VNMR 4.3B software on a Silicon Graphics Indigo^2 XZ workstation. The ^{15}N-edited 3D HMQC-TOCSY (Wijmenga et al., 1989; Lerner & Bax, 1986) experiment with a mixing time of 70 ms and the NOESY-HMQC (Zuiderweg & Fesik, 1989) experiment with a mixing time of 100 ms were recorded with uniformly ^{15}N-labeled rCMTI-V with 1024 × 64 × 32 data points for the f_1, f_2, and f_3 dimension, respectively. The data sets were linear-predicted and/or zero-filled to 1024 × 512 × 256 and processed. Spectral widths were set to be 7000 Hz for the ^1H dimensions and 3500 Hz for the ^{15}N dimension. The ^{13}C-edited 3D HMQC-NOESY (Fesik & Zuiderweg, 1988) experiment with a mixing time of 150 ms was performed with the doubly labeled protein with 1024 × 64 × 64 data points for the f_1, f_2, and f_3 dimension, respectively. The data set was linear-predicted and/or zero-filled to 1024 × 512 × 512 points and processed. Spectral widths for these experiments were set to 6967, 6284, and 3419 Hz for the f_1, f_2, and f_3 dimension, respectively. 2D TOCSY (Bax & Davis, 1985) and NOESY (Anil Kumar et al., 1980) spectra with different mixing times or temperatures were recorded with a spectral width of 7000 Hz on 256 increments of 2K data points each and processed after zero-filling to 4K data points in each dimension. 2D TOCSY and NOESY experiments were performed with the ^{15}N-labeled protein; the 4K × 1K data matrix was zero-filled to 8K × 4K and processed to determine $^3J_{N\beta}$ coupling constants (Cai et al., 1995d). The 2D HSQC (Bodenhausen & Ruben, 1980; Kay et al., 1989a) spectrum was recorded with spectral widths of 7000 and 3500 Hz for the ^1H and ^{15}N dimension, respectively, with 4096 × 256 complex data points in the $f_1(^1H)$ × $f_2(^{15}N)$ dimensions. All data sets were collected in hypercomplex phase-sensitive mode. ^1H chemical shifts were referenced by assigning a value of 4.71 ppm to the H_2O peak. ^{15}N chemical shifts were referenced indirectly to liquid NH_3, using CH_3NO_2 as a secondary reference with an assigned ^{15}N chemical shift of 380.23 ppm (Live et al., 1984).

2D sensitivity-enhanced proton-detected heteronuclear NMR spectroscopy (Cavanagh et al., 1991; Palmer et al., 1991) was used to measure spin−lattice relaxation rate constants (R_1), spin−spin relaxation rate constants (R_2), and $\{^1H\}-^{15}N$ steady-state NOE. Relaxation measurements were performed using inversion−recovery (Vold et al., 1968), Carr-Purcell-Meiboom-Gill (CPMG; Meiboom & Gill, 1958), and steady-state NOE (Noggle & Shirmer, 1971) pulse sequences, as described by Kördel et al. (1992) and Skelton et al. (1993). For each measurement, 256 (for R_1 and R_2) and 192 (for NOE) increments of 4K data points each were acquired with 4, 8, and 16 transients per block for R_1 and R_2 and NOE experiments, respectively. A total recycling delay of 2 s was used for R_1 and R_2, and 4 s for NOE measurements. Spectral widths were set to 7000 and 3500 Hz for the ^1H and ^{15}N dimension, respectively. For the R_1 measurements, 11 spectra were recorded using relaxation delays of 0.008(×2), 0.05(×2), 0.12(×2), 0.25(×2), 0.40, 0.70(×2), 1.1, 2.5 s, whereas 13 spectra were recorded with relaxation delays of 0.008(×2), 0.02(×2), 0.035, 0.055(×2), 0.08, 0.11, 0.15(×2), 0.2, and 0.5 s during the CPMG period for R_2 experiments, duplicate measurements being denoted by ×2 within parentheses. NOE measurements were repeated three

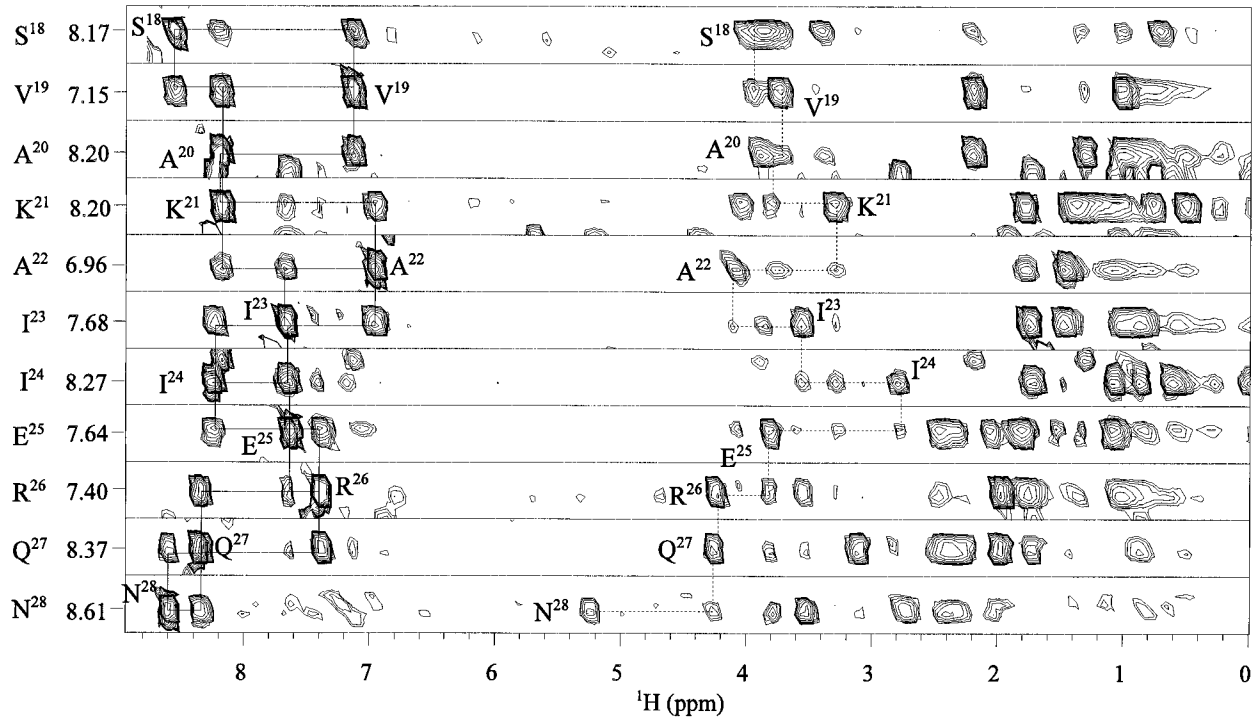

FIGURE 1: Small region of the "reduced" 3D $^{15}N-^{1}H$ NOESY-HMQC spectrum of ^{15}N-labeled rCMTI-V at pH 5.4 and 30 °C. This figure is composed of narrow strips taken from different f_1-f_3 planes of the 3D spectrum. Solid lines indicate NH−NH connectivities; dashed lines show NH(i+1)−C_αH(i) connectivities.

times. A line-broadening factor of 6 Hz was applied to the f_2 dimension, and a shifted Gaussian window function was applied to the f_1 dimension. A shifted Gaussian window function was applied to both dimensions to resolve the overlap of cross-peaks of Gly[5] and Gly[59], Leu[12] and Val[15], and Asn[30] and Val[42]. The solvent signal was not suppressed during data acquisition. It was removed by means of a digital low-pass filter prior to Fourier transformation (Marion et al., 1989). The preacquisition delay was adjusted so that first order phasing was zero and no baseline correction was necessary.

NMR Spectral Assignments and Solution Structure Determination. Spectral assignments—both sequential and stereospecific—of rCMTI-V were made, utilizing both 2D and 3D NMR data sets, following the standard procedure (Wüthrich, 1986). A small section, corresponding to residues 18−28 in the α-helix, of the "reduced" 3D NOESY-HMQC spectrum (Driscoll et al., 1990) of ^{15}N-labeled rCMTI-V is shown in Figure 1. Stereospecific assignments and side-chain dihedral angle, χ^i, determinations made for rCMTI-V were found to be the same as for CMTI-V (Cai et al., 1995a) for most of the residues. In the case of Arg[52] in rCMTI-V, χ^1 was determined to be 180° on the basis of homonuclear coupling constants ($^3J_{\alpha\beta}{}^3 < 5$ Hz; $^3J_{\alpha\beta}{}^2 > 10$ Hz) and heteronuclear coupling constants ($^3J_{N\beta}{}^3$ and $^3J_{N\beta}{}^2$ both less than 1 Hz), as opposed to −60° reported for native CMTI-V (Cai et al., 1995a), which was found to be in error upon reexamination due to ambiguity in intensities of NOE cross-peaks. Accordingly, χ^2, of Arg[52] was corrected from 180 to 60° in CMTI-V. For Val[34] in CMTI-V, χ^1 could not be determined because of degeneracy of its methyl groups. Using the ^{15}N-labeled protein, it was determined as −60° on the basis of the coupling constants, $^3J_{\alpha\beta}$ (<5 Hz) and $^3J_{N\beta}$ (<1 Hz). The Pro puckers in rCMTI-V were found to be the same as in native CMTI-V (Cai et al., 1995c).

The three-dimensional solution structure determination of rCMTI-V was carried out, as described previously for CMTI-V (Cai et al., 1995a). In the present case, more distance constraints were employed because of the availability of ^{15}N- and ^{13}C-edited 3D data. Thus, a total of 123 dihedral angle constraints and 927 interproton distances was determined: 256 intraresidue, 249 sequential (for residues i and j, $|i - j| = 1$), 124 short range ($|i - j| \leq 5$), 248 long range ($|i - j| > 5$), and 50 constraints on the basis of 25 hydrogen bonds. Using the simulated annealing (SA) protocol (Brünger, 1992; Nilges et al., 1988), an energy-minimized mean structure (Driscoll et al., 1989), designated ⟨SA⟩, based on a family of 18 NMR structures, was obtained. The atomic coordinates of the refined average structure of rCMTI-V, along with the NMR constraints, have been deposited with the Brookhaven Protein Data Bank under the accession code 1MIT.

Measurements of ^{15}N Relaxation Parameters. The relaxation rate constants were determined by nonlinear least-squares fitting of experimental peak heights to the following equations:

$$I(t) = I_\infty - (I_\infty - I_0)\exp(-R_1 t) \quad (1)$$

$$I(t) = I_0 \exp(-R_2 t) \quad (2)$$

where t is the parametical relaxation delay in each measurement, I_∞ is the long-time steady-state resonance intensity, R_1 is the spin−lattice relaxation rate constant, and R_2 is the spin−spin relaxation rate constant. Data for R_2 measurement with a delay of 0.11 s were not used because receiver overflow occurred during that experiment. Uncertainties in peak heights were assessed by RMS baseline noise level. These values are generally underestimated in comparison with those obtainable from duplicate data points; nonetheless, the procedure used here allows for comparison of dynamical

Table 1: Structure Statistics for rCMTI-V

Distance Constraints			
		rms deviations from experimental distance constraint (Å)	
total	distance	SA[a]	$\langle SA \rangle_r$[b]
927	all	0.0111 ± 0.0015	0.0091
256	intraresidue	0.0094 ± 0.0008	0.0091
249	sequential ($\|i-j\|=1$)	0.0105 ± 0.0034	0.0074
124	short range ($\|i-j\| \leq 5$)	0.0080 ± 0.0033	0.0017
248	long range ($\|i-j\| > 5$)	0.0127 ± 0.0025	0.0093
50	hydrogen bond	0.0166 ± 0.0037	0.0094

Torsional Angle Constraints			
		rms deviations from experimental distance constraints (deg)	
total	angle	SA[a]	$\langle SA \rangle_r$[b]
123	all	0.270 ± 0.029	0.262
54	ϕ angles	0.090 ± 0.030	0.100
69	χ angles	0.349 ± 0.035	0.336

Energies (kcal/mol)[c]		
	SA[a]	$\langle SA \rangle_r$[b]
F_{noe}	5.85 ± 1.57	3.84
F_{tor}	1.37 ± 0.30	1.27
F_{repel}	4.39 ± 0.27	4.18
$F_{L\text{-}J}$	−345 ± 11.0	−355

RMS Deviations from Idealized Geometry		
	SA[a]	$\langle SA \rangle_r$[b]
bond (Å)	0.0075 ± 0.0001	0.0075
angle (deg)	2.0973 ± 0.0079	2.0915
impropers (deg)	0.1992 ± 0.0359	0.1822

RMS Deviations of 18 SA's and Their Mean Structure (Å)	
whole protein	
main chain	1.15 ± 0.18
all heavy atoms	1.50 ± 0.16
residues 2−40 and residues 49−67	
main chain	0.53 ± 0.08
all heavy atoms	0.95 ± 0.08
loop region (residues 41−48)	
main chain	1.50 ± 0.38
all heavy atoms	2.49 ± 0.51

[a] SA represents the 18 individual structures refined by the simulated annealing method, and $\langle SA \rangle_r$ represents the refined mean structure, as described by Cai et al. (1995a). [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 constrains and torsional angle constrains, 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) for distance constraints and 500 kcal/(more rad^2) for repulsion terms. F_{noe}, F_{tor}, and F_{repel} are, respectively, the constraint violation energies associated with distance constrains and torsional angle constraints and 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). $E_{L\text{-}J}$ is the van der Waals energy recalculated with the same coordinates, but with the electrostatic terms included.

parameters calculated with those reported by others. Heteronuclear NOEs were calculated by

$$\eta = I_{sat}/I_{unsat} \quad (3)$$

in which I_{sat} and I_{unsat} are the experimental peak heights measured from spectra recorded with and without proton irradiation during the recycling delay. Uncertainties were calculated by the sum of the root-mean-square (RMS) of the uncertainties of peak heights about the RMS baseline noise level for I_{sat} and I_{unsat}. Repeated NOE measurements were then averaged.

Computation of Model-Free Parameters. The spin relaxation of amide ^{15}N nuclei in a protein is dominated by the dipolar interaction between the ^{15}N and the attached ^{1}H nuclei as well as the chemical shift anisotropy (CSA) mechanism. The relaxation parameters, as described by Abragam (1961), are given as

$$R_1 = (d^2/4)[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \quad (4)$$

$$R_2 = (d^2/8)[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + c^2/6[3J(\omega_N)4J(0)] + R_{ex} \quad (5)$$

$$\eta = 1 + (d^2/4R_1)(\gamma_N/\gamma_H)[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad (6)$$

where $d = (\mu_0 h \gamma_N \gamma_H/8\pi^2)(1/r^3_{NH})$ and $c = (\omega_N/\sqrt{3})\Delta\sigma$. In eqs 4−6, μ_0 is the permeability of free space, h is Planck's constant, γ_i is the gyromagnetic ratio of spin I, r_{NH} is the nitrogen−hydrogen bond length; ω_H and ω_N are the Larmor frequencies of ^{1}H and ^{15}N, respectively, and $\Delta\sigma$ is the chemical shift anisotropy, which has been determined to be −160 ppm for ^{15}N nuclei in helical polypeptide chains (Hiyama et al., 1988).

The amplitudes and time scales of intramolecular motions of individual N−H bond vectors in a protein are obtained from relaxation data by the model-free approach of Lipari and Szabo (1982a,b; Clore et al., 1990). In this approach, a dimensionless generalized order parameter (S^2) is used to describe the spatial extent of motion of the N−H bond vector. Value of S^2 ranges from 0 for completely free internal motion to 1 for completely restricted motion. The effective correlation time (τ_e) depends on the rate of motion of the N−H bond vector. A physical interpretation of τ_e is generally difficult and requires a specific motional model. R_{ex} is involved to compensate contributions to the ^{15}N spin−spin relaxation rate due to chemical exchange and/or conformational averaging on a time scale much slower than the overall rotational correlation time (τ_m).

The spectral density function for a protein governed by isotropical tumbling is

$$J(\omega) = (2/5)[S^2\tau_m/(1 + (\omega\tau_m)^2) + (1 - S^2)\tau/(1 + (\omega\tau)^2)] \quad (7)$$

where $1/\tau = 1/\tau_m + 1/\tau_e$. If $\tau_e \ll \tau_m$ and τ_e is small, then eq 7 can be reduced to

$$J(\omega) = (2/5)[S^2\tau_m/(1 + (\omega\tau_m)^2) + (1 - S^2)\tau_e] \quad (8)$$

In the limit that $\tau_e \rightarrow 0$, eq 8 further reduces to

$$J(\omega) = (2/5)[S^2\tau_m/(1 + (\omega\tau_m)^2)] \quad (9)$$

The model-free parameters were extracted by substitution of either eqs 7, 8, or 9 into eqs 4, 5, and 6 for R_1, R_2, and NOE, respectively, followed by nonlinear least-squares fitting of the experimental data. Whereas τ_m is the same for all N−H vectors, τ_e is a variable. The optimization minimized the function

$$\chi^2 = \sum[(R_{1i} - R^c_{1i})^2/\sigma^2_{1i} + (R_{2i} - R^c_{2i})^2/\sigma^2_{2i} + (\eta_i - \eta^c_i)^2/\sigma^2_{\eta i}] \quad (10)$$

where the index, i, indicates the residue number, σ_{1i}, σ_{2i}, and $\sigma_{\eta i}$ are the uncertainties for experimental relaxation parameters R_{1i}, R_{2i}, and η_i of residue number i, respectively,

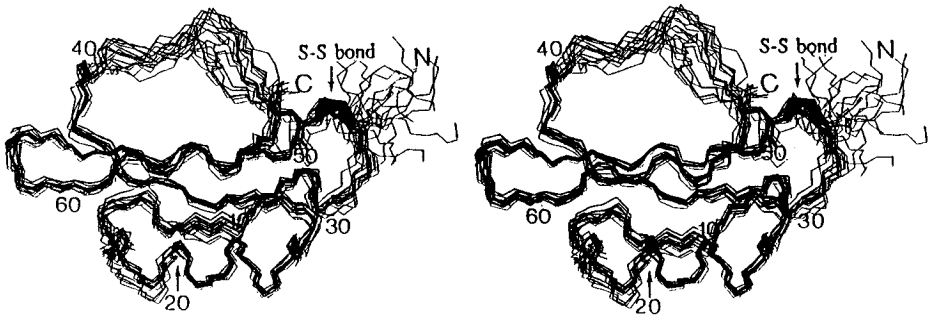

FIGURE 2: Stereoview of the best fit superposition of the 18 SA structures. The main chain atoms (N, C_α, and $C_{C=O}$), including the Cys^3−Cys^{48} disulfide bridge, are shown.

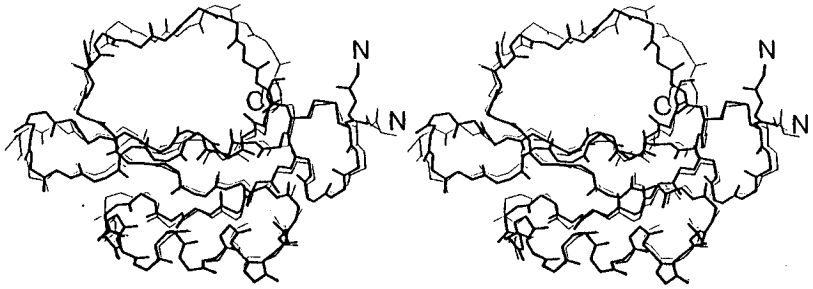

FIGURE 3: Superposition of stereoviews of refined mean ⟨SA⟩ structures of CMTI-V (broken thin line; Cai et al., 1995a) and rCMTI-V (solid thick line). The main-chain atoms (N, C_α, and $C_{C=O}$), including the Cys^3−Cys^{48} disulfide bridge, are shown.

and R^c_{1i}, R^c_{2i}, and η^c_i are the calculated relaxation parameters of residue *i*, respectively. Calculations were carried out using the Model-free version 3.0 software provided by Dr. Arthur G. Palmer, III, Columbia University, New York.

RESULTS

Three-Dimensional Solution Structure of rCMTI-V. Statistical details for the best 18 SA structures and the refined mean ⟨SA⟩ structure are listed in Table 1. Superposition of the main-chain atoms (C, C^α, and N atoms) for the 18 SA structures is shown in Figure 2. The conformation of all main chain atoms for the 18 SA structures is well-conserved, except for the loop region (40−48), several N-terminal residues, and the C-terminal residue, and is the same as that of native CMTI-V (Cai et al., 1995a), as expected; thus, the same secondary structure elements are found: one α-helix (residues 18−28) and three pairs of β-sheets (residues 8−9 and 62−60 antiparallel to residues 67−66 and 54−56, respectively; and residues 32−37 parallel to 51−55), connected by the binding loop and four turns (type II, 12−15; type I, 28−31 and 47−50; and type III, 56−59). Figure 3 depicts superposition of the backbone ⟨SA⟩ structures of rCMTI-V and CMTI-V. The average RMSD between the two refined mean structures is 0.95 Å, which is smaller than the average RMSD (1.15 Å) of 18 SA structures of rCMTI-V from their mean ⟨SA⟩ structure for all main-chain atoms. The average RMSD as a function of residue number is given for rCMTI-V in Figure 4. Main-chain atoms in the binding loop (residues 40−48) and the N-terminal region display higher RMSD values.

Relaxation Data and Order Parameters. The 2D ^1H−^{15}N chemical shift correlation map of rCMTI-V is shown in Figure 5. Data sets for 62 of the 69 residues, excluding those for the N-terminal Gly and six Pro residues, are expected. R_1, R_2, and NOE could be reliably measured for 60 backbone ^{15}NHs. Data obtained for Ser^1 and Ser^2 were not used for analysis because of poor quality. Typical R_1 and R_2 decay curves and complete relaxation data are shown in Figures 6 and 7, respectively. Based on the trimmed weighted average value of the R_2/R_1 ratio, which was 2.54 ± 0.14, the overall correlation time, τ_m, was estimated to be 4.61 ± 0.24 ns. This value of τ_m was used to extract the Lipari and Szabo parameters (Table S-3, Supporting Information). These are depicted in Figure 8. The final globally optimized τ_m value for rCMTI-V was 4.43 ± 0.02 ns.

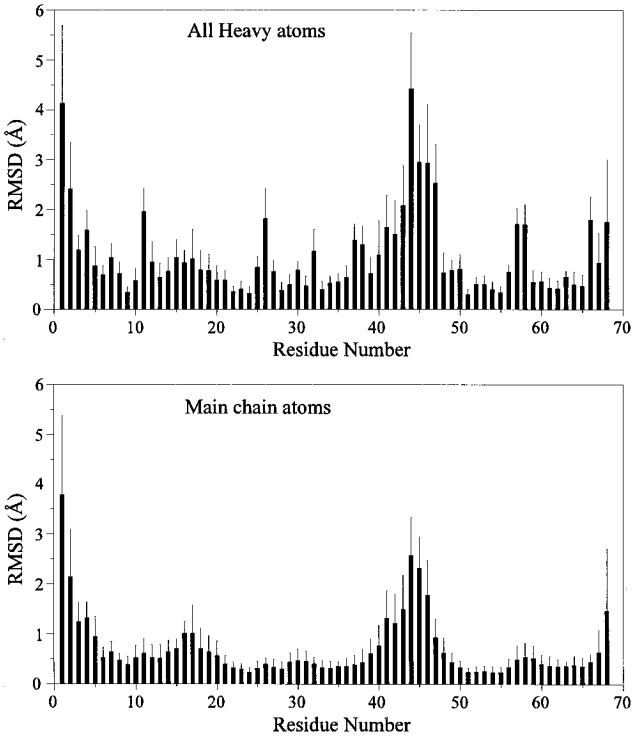

FIGURE 4: Atomic RMSD distribution of the individual SA structures about the mean ⟨SA⟩ structure compared for rCMTI-V. The filled bars indicate the average RMSD values of 18 structures, and the standard deviation is indicated by the vertical error bar.

The order parameters calculated for rCMTI-V (Figure 8; Table S-3 in the Supporting Information) indicate that the

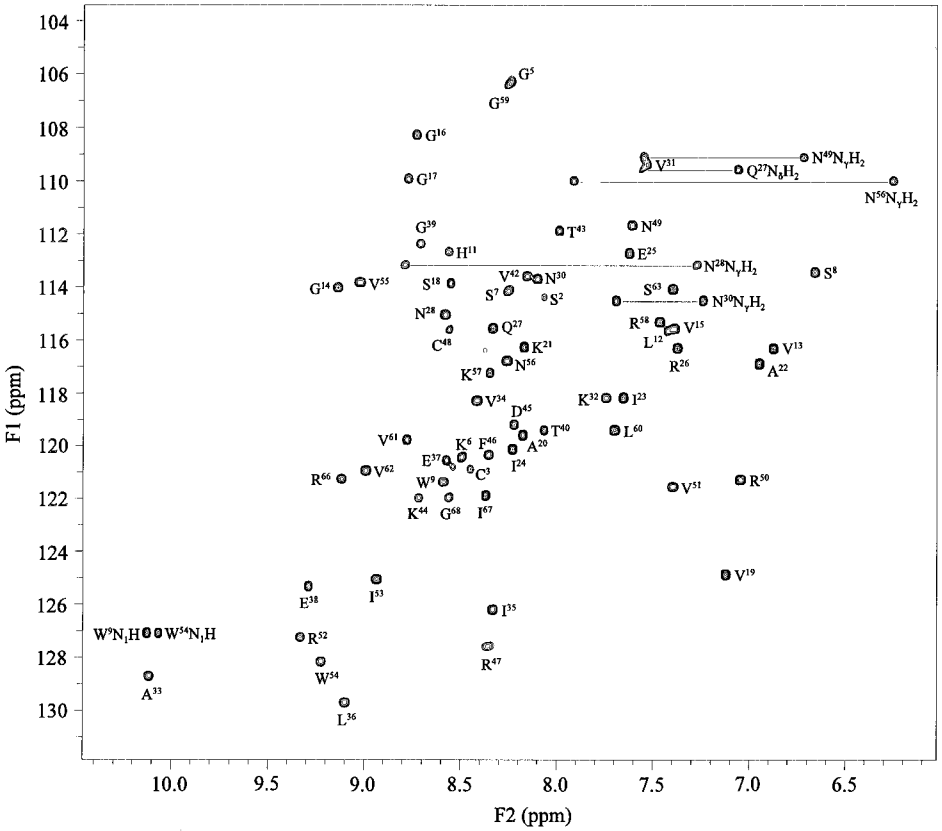

FIGURE 5: $^{1}H-^{15}N$ HSQC map of rCMTI-V, with most NH assignments marked by the one-letter name and residue number.

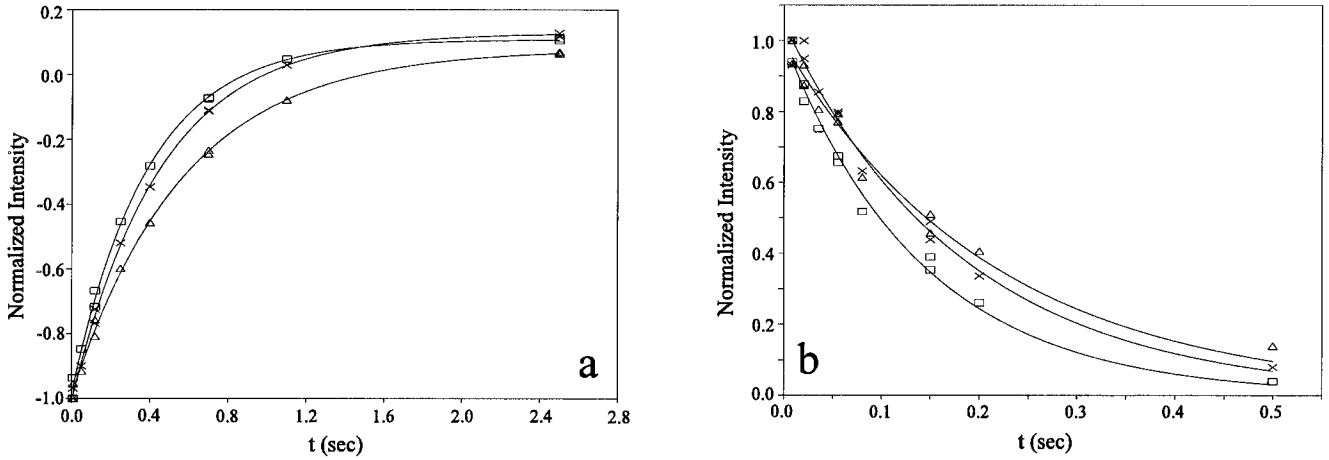

FIGURE 6: ^{15}N relaxation data for Ala[20] (□), Val[30] (×), and Asp[45] (△). (a) R_1 relaxation decay curves; (b) R_2 relaxation decay curves. Uncertainties are smaller than the size symbol and thus not shown.

inhibitor's backbone is highly constrained, except for the region consisting of the N-terminal residues (<7) and the loop region (41−48). The average S^2 calculated is 0.83 ± 0.04 for the constrained region, 0.83 ± 0.03 for β-sheets, 0.85 ± 0.02 for the α-helix, 0.81 ± 0.04 for turns, and 0.73 ± 0.08 for the loop region. The effective correlation times, τ_e, for most residues are in the range of 25−84 ps. Conformational exchange terms were included for 14 residues. The most significant exchange terms are found to be associated with different turns in the inhibitor molecule (Figure 8; Table S-3 in the Supporting Information).

DISCUSSION

The experimental ^{15}N relaxation parameters (R_1, R_2, and $\{^{1}H\}-^{15}N$ NOE) measured for rCMTI-V are sensitive to molecular motions in the picosecond and nanosecond time scale (Clore et al., 1990; Cheng et al., 1993). Data analyses were made assuming an overall isotropic tumbling for the protein molecule. The model-free parameters (S^2, τ_e) derived from experimental data reflect protein mobility on a time scale that is faster than the overall correlation time (τ_m). Motions which occur on a time scale slower than overall correlation time may be reflected by R_2 values. An R_{ex} term included in the model-free formalism generally accounts for such effects as chemical exchange and conformational averaging. However, the R_{ex} values reported here are the results of correction terms to R_2 experimental data to achieve a better overall fit between theory and experiment. Therefore, no attempt is made to quantitatively interpret the R_{ex} values. A relatively large R_{ex} value may indicate the

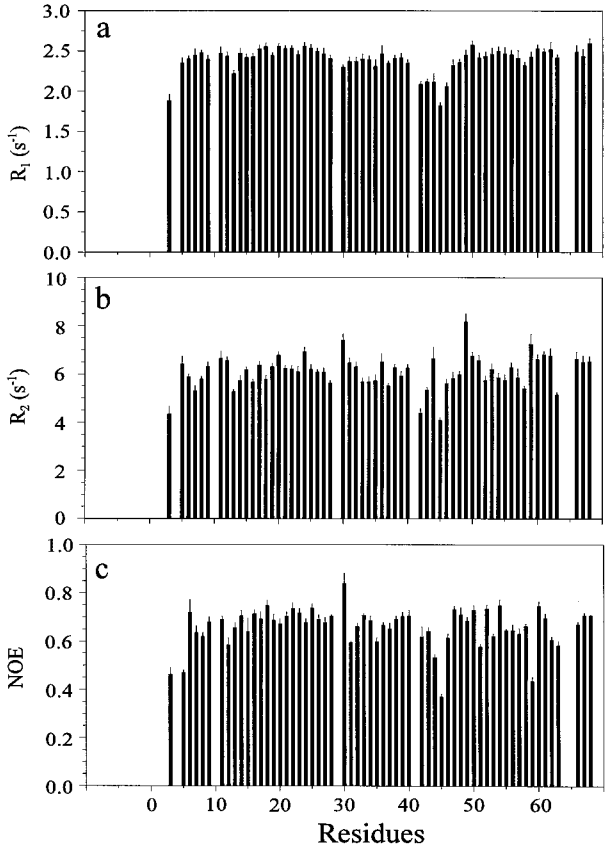

FIGURE 7: Plot of the measured ^{15}N relaxation parameters as a function of residue number: (a) spin−lattice relaxation rate constant (R_1); (b) spin−spin relaxation rate constant (R_2); (c) $\{^1H\}-^{15}N$ steady-state heteronuclear NOE. Error bars represent standard deviations calculated from the optimization of eqs 1, 2, and 3. Data are missing for the N-terminal glycine and six proline residues and are not shown for Ser^1 and Ser^2.

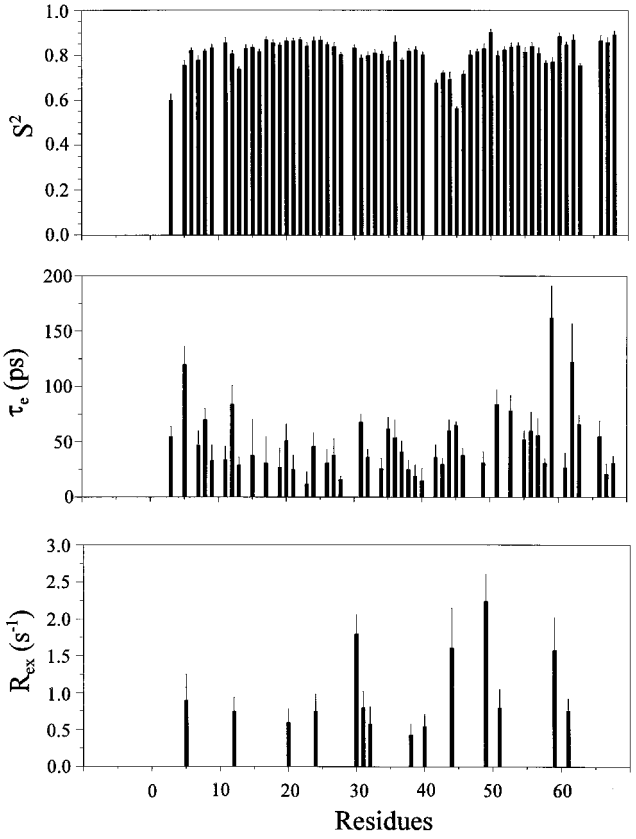

FIGURE 8: Plot of the optimized model free parameters as a function of residue number: (a) generalized order parameter (S^2); (b) effective correlation time for internal motions (τ_e) in picosecond; (c) chemical exchange term (R_{ex}) in Hertz.

existence of an intramolecular conformational exchange process.

Secondary Structure Elements. The α-helix of rCMTI-V shows more restricted motions ($\langle S^2\rangle = 0.852 \pm 0.019$) than the different turns ($\langle S^2\rangle$ values ranging from 0.792 ± 0.010 to 0.817 ± 0.013). The β-sheets show different motions over a relatively large range of average order parameters (0.809 ± 0.028 to 0.867 ± 0.018). There seems to be no correlation between the average order parameter and the nature of secondary structure elements in rCMTI-V (Table 2).

In the case of the α-helix region (residues 18−28), the R_{ex} values obtained for Ala^{20} and Ile^{24} are not significantly large (0.60 and 0.75 Hz, respectively). Only Asn^{28} has an S^2 value (0.804 ± 0.010) that is much lower than those of other residues forming the helix. This may be due to the fact that Asn^{28} is also involved in the type I turn made up of residues 28−31 (average $S^2 = 0.792 \pm 0.010$). The average order parameter for residues 18−27 is 0.856 ± 0.011.

The six strands of β-sheets in rCMTI-V exhibit different mobilities. The average order parameters for β5 (residue 60−62) and β6 (65−67) are surprisingly high (0.867 and 0.863, respectively; Table 2). This may be attributed to the fact that the backbone NHs of residues 60, 62, and 67 are involved in hydrogen bonds with residues 56, 54, and 7, respectively (Cai et al., 1995a). Residues 51−54 are involved in two pairs of β-sheets, but there is no noticeable increase in their order parameters. The average order parameter of the parallel β-sheet involving β2 (residues 32−38) and β3 (residues 51−56) is lower than that of the two antiparallel β-sheets of β1−β6 (7→9 with 62→60) and β4−β5 (67→65 with 54→56). This lends support to the notion that antiparallel β-sheets may be more stable than parallel ones (Creighton, 1984). No significant R_{ex} value (>1 Hz) is observed for any residue in the β-sheets.

The average order parameters are lower for residues in different turns than for those in the α-helix or in the β-sheets (Table 2). No significant differences in average order parameters are observed among the different types of turns. However, at least one residue associated with each turn

Table 2: Mean Order Parameters of Secondary Structure Elements in rCMTI-V

secondary structure	sequence	number of residues[a]	$\langle S^2\rangle^b$
a	18−28	11	0.852 ± 0.019
β1	7−9	3	0.811 ± 0.028
β2	32−38	7	0.809 ± 0.028
β3	51−56	6	0.827 ± 0.018
β4	54−56	3	0.834 ± 0.015
β5	60−62	3	0.867 ± 0.018
β6	65−67	2	0.863 ± 0.006
T(I)	12−15	4	0.794 ± 0.033
$T(II_a)$	28−31	3	0.792 ± 0.010
$T(II_b)$	47−50	3[c]	0.817 ± 0.013
T(III)	57−60	4	0.808 ± 0.053

[a] Number of residues in each secondary structural element for which relaxation data were obtained. [b] Average order parameter, calculated from Table S-3 (Supporting Information). [c] Data for residue 50 are not included in the average because of ill-fitting behavior during the extraction of order parameters.

requires a nonzero value for the R_{ex} term. This suggests the presence of a chemical exchange or conformational averaging process on a time scale range of millisecond to second.

As expected, the loop region demonstrates less restricted motions in comparison with the rest of the protein molecule, with Asp[45] showing the minimum order parameter of 0.563 ± 0.010. The order parameters for residues of the binding loop on the N-terminal side from the reactive site, which is the peptide bond between Lys[44] and Asp[45], are smaller than those of the residues on the C-terminal side (Figure 6). The R_{ex} term is required for Lys[44] for the extraction of the order parameter, thus suggesting the presence of motion on millisecond to second time scale.

Earlier studies of backbone dynamics of several proteins (Kay et al., 1989b; Clore et al., 1990; Kördel et al., 1992) indicate that N- and C-termini have considerably lower order parameters than secondary structure elements. In the case of rCMTI-V, while the N-terminal residues have notably lower than average order parameters, the C-terminal residues present average to above average order parameters. This may be attributed to the fact that the C-terminal residues 65−67 form an antiparallel β-sheet with residues 7−9; furthermore, the backbone NH of Ile[67] is hydrogen-bonded to residue Ser[7]. The above average order parameter observed for residue 68 is not explicable at this time.

There are only three residues (Gly[5], Gly[59], and Val[62]) which are associated with τ_e values larger than 100 ps. Nineteen residues require τ_e values that are either zero or smaller than 25 ps. Most (38) of the residues have τ_e values in the range of 25−84 ps. A physical interpretation of τ_e was not attempted.

Comparison with Chymotrypsin Inhibitor 2 (CI-2) and Eglin c. The sequence from Asn[19] to Gly[83] in CI-2 can be aligned with the sequence from Pro[4] to Gly[68] in CMTI-V (Krishnamoorthi et al., 1990). The alignment indicates about 62% differences in the sequences. Nonetheless, the NMR solution structure of CMTI-V (Cai et al., 1995a) is found to be highly similar to the crystal structure of CI-2 (McPhalen & James, 1987): the folding patterns are similar and the secondary structure elements are highly conserved. Eglin c is another member of the pototo I inhibitor family that differs in sequence (63%; Krishnamoorthi et al., 1990) but possesses a very similar three-dimensional structure (Hyberts & Wagner, 1990).

The overall correlation time, τ_m, reported for the 64-residue truncated form of CI-2 (residues 1−19 removed) is 4.77 ± 0.07 ns (Shaw et al., 1995); a value of 4.15 ± 0.05 ns is reported for the 70-residue eglin c (Peng & Wagner, 1992). These values compare well with the value of 4.43 ± 0.02 ns obtained for the 69-residue rCMTI-V. The discrepancies may arise from the way τ_m is obtained: For CI-2 it is obtained from a grid search and no further optimization is performed. In the present study as well as in the case of eglin c, further optimization after the initial grid search on τ_m has been employed.

Both the trend and magnitudes of order parameters compare well among CI-2, eglin c, and rCMTI-V. CI-2 lacks a flexible N-terminus due to the removal of residues 1−19. The average S^2 for the structured regions are 0.83, 0.87, and 0.83 for eglin c, CI-2, and rCMTI-V, respectively. However, τ_e values in the range of 100−200 ps are determined for the N-terminal residues in eglin c, but not in rCMTI-V; the average order parameter for the N-terminal residues (ranging from 0.08 to 0.60) is lower than that observed in the case of rCMTI-V (0.599, 0.755, and 0.824 for Cys[3], Gly[5], and Lys[6], respectively), thus indicating a greater N-terminal flexibility for the former molecule. This may be attributed to the presence of the Cys[3]−Cys[48] disulfide bond that exists in rCMTI-V, but not in eglin c. The N terminal and binding loop residues in eglin c require different τ_e values. That is not the case with rCMTI-V. The τ_e values are, however, similar for both the structured regions and the binding loop among rCMTI-V, CI-2 and eglin c.

CONCLUSIONS

The three-dimensional solution structure of rCMTI-V was determined to be the same as that of the native protein, although it has an additional N-terminal Gly residue and is not N-acetylated. This validates the interpretation of the present protein backbone dynamics study in structural terms. The dynamical features of rCMTI-V include a generally restricted backbone and C-terminus and a less restricted binding loop and N-terminus. No significant motions on microsecond to millisecond time scale are observed. The less-ordered segments of the N-terminal residues and the binding loop region in CMTI-V or rCMTI-V display greater flexibility or internal motions; the structural uncertainties associated with these regions are not simply due to a dearth of NMR constraints. This also implies the importance of protein flexibility in function. Recent results obtained in our laboratory indeed imply an interaction between the N-terminus of the inhibitor and trypsin or factor XIIa (Huang et al., to be published). The Cys[3]−Cys[48] bridge in CMTI-V somewhat dampens internal motion in the N-terminal region. Similarly, involvement of the C-terminus in a β-sheet structure reduces its mobility.

ACKNOWLEDGMENT

We thank Dr. Arthur G. Palmer III, Columbia University, New York, for providing us the model-free software, Dr. Charlotta Johansson, Swedish NMR Center, Stockholm, Sweden, for some of the pulse sequences used, and Mr. Dave Manning for his technical assistance.

SUPPORTING INFORMATION AVAILABLE

Three tables containing ^{1}H and ^{15}N assignments, R_1, R_2, and NOE values, and backbone dynamical parameters (S^2, τ_e, R_{ex}) for rCMTI-V (10 pages). Ordering information is given on any current masthead page.

REFERENCES

Abragam, A. (1961) *Principles of Nuclear Magnetism*, pp 264−353, Clarendon Press, Oxford.
Anil Kumar, Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1−6.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−360.
Bode, W., Greyling, H. J., Huber, R., Otlewski, J., & Wilusz, T. (1989) *FEBS Lett. 242*, 285−292.
Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185−189.
Brooks, R. R., Bruccoleri, R. E., Olafson, R. E., State, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
Brünger, A. T. (1992) X-PLOR version 3.1, Yale University Press, New Haven and London.
Cai, M., Gong, Y., Kao, J. L.-F., & Krishnamoorthi, R. (1995a) *Biochemistry 34*, 5201−5211.

Cai, M., Gong, Y., Prakash, O., & Krishnamoorthi, R. (1995b) *Biochemistry 34*, 12087−12094.
Cai, M., Huang, Y., Liu, J., & Krishnamoorthi, R. (1995c) *J. Biomol. NMR 6*, 123-128.
Cai, M., Huang, Y., Prakash, O., Wen, L., Han, S. K., & Krishnamoorthi, R. (1995d) *J. Magn. Reson. Ser. B 108*, 189−191.
Cavanagh, J., Palmer, A. G., Wright, P. E., & Rance, M. (1991) *J. Magn. Reson. 91*, 429−436.
Cheng, J.-W., Lepre, C. A., Chamber, S. P., Fulghum, J. R., Thomson, J. A., & Moore, J. M. (1993) *Biochemistry 32*, 9000−9010.
Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., & Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 4989−4991.
Creighton T. E. (1984) *Proteins*, pp 1−515, Freeman, New York.
Driscoll, P. C., Gronenborn, A. M., Beress, L., & Clore, G. M. (1989) *Biochemistry 28*, 2188−2198.
Driscoll, P. C., Gronenborn, A. M., Wingfield, P. T., & Clore, G. M. (1990) *Biochemistry 29*, 4668−4682.
Fesik, S. W., & Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 78*, 588−593.
Hayashi, K., Takashita, T., Hamato, N., Takano, H., Hara, S., Miyata, T., & Kato, H. (1994) *J. Biochem. (Tokyo) 116*, 1013−1018.
Hiyama, Y., Niu, C.-H., Silverton, J. V., Bavaso, A., & Torchia, D. A. (1988) *J. Am. Chem. Soc. 110*, 2378−2383.
Hojima, Y., Pierce, J. V., & Pisano, J. J. (1982) *Biochemistry 21*, 3741−3746.
Holak, T. A., Gondol, D., Otlewski, J., & Wilusz, T. (1989a) *J. Mol. Biol. 210*, 635−648.
Holak, T. A., Bode, W., Huber, R., Otlewski, J., & Wilusz, T. (1989b) *J. Mol. Biol. 210*, 649−654.
Hyberts, S. G., & Wagner, G. (1990) *Biochemistry 29*, 1465−1474.
Karplus, M., & McCammon, J. A. (1983) *Annu. Rev. Biochem. 52*, 263−300.
Kay, L. E., Marion, D., & Bax, A. (1989a) *J. Magn. Reson. 84*, 72−84.
Kay L. E., Torchia, D. A., & Bax, A. (1989b) *Biochemistry 28*, 8972−8979.
Kördel, J., Skelton, H. J., Akke, M., Palmer A. G., & Chazin, W. J. (1992) *Biochemistry 31*, 4856−4866.
Krishnamoorthi, R., Gong, Y., & Richardson, M. (1990) *FEBS Lett. 273*, 163−167.
Krishnamoorthi, R., Gong, Y., Lin, C.-L. S., & VanderVelde, D. (1992a) *Biochemistry 31*, 898−904.
Krishnamoorthi, R., Lin, C.-L. S., & VanderVelde, D. (1992b) *Biochemistry 31*, 4965−4969.
Laskowski, M. Jr., & Kato, I. (1980) *Annu. Rev. Biochem. 49*, 593−626.
Lerner, L., & Bax, A. (1986) *J. Magn. Reson. 69*, 375−380.
Lipari, G., & Szabo, A. (1982a) *J. Am. Chem. Soc. 104*, 4546−4559.
Lipari, G., & Szabo, A. (1982b) *J. Am. Chem. Soc. 104*, 4559−4570.
Live, D. H., Davis, D. G., Agosta, W., C., & Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1939−1941.
Ludvigsen, S., Shen, H., Kjaer, M., Madsen, J. C., & Poulsen, F. M. (1991) *J. Mol. Biol. 222*, 621−635.
Marion, D., Ikura, M., & Bax, A. (1989) *J. Magn. Reson. 84*, 425−430.
McPhalen, C. A., & James, M. N. G. (1987) *Biochemistry 26*, 261−269.
McWherter, C. A., Walkenhorst, W. F., Campbell, E. J., & Glover, G. I. (1989) *Biochemistry 28*, 5708−5714.
Meiboom S., & Gill, D. (1958) *Rev. Sci. Instrum. 29*, 688−691.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.
Noggle, J. H., & Shirmer, R. E. (1971) *The Nuclear Overhauser Effect: Chemical Applications*, Academic Press, New York.
Otlewski, J., & Zbyryt, T. (1994) *Biochemistry 33*, 200−207.
Otlewski, J., Zbyryt, T., Krokoszynska, I., & Wilusz, T. (1990) *Biol. Chem. Hoppe Seyler 371*, 589−594.
Palmer, A. G., Cavanagh, J., Wriet, P. E., & Rance, M. (1991) *J. Magn. Reson. 93*, 151−170.
Peng, J. W., & Wagner, G. (1992) *Biochemistry 31*, 8571−8586.
Prat Gay, G. de, & Fersht, A. R. (1994) *Biochemistry 33*, 7957−7963.
Shaw, G. L., Davis, B., Keeler, J., & Fersht, A. (1995) *Biochemistry 34*, 2225−2233.
Skelton, N. J., Palmer, A. G., Akke, M., Kördel J., Rance, M., & Chazin, W. (1993) *J. Magn. Reson., Ser. B 102*, 253−264.
Vold, R. L., Waugh, J. S., Klein, M. P., & Phelps, D. E. (1968) *J. Chem. Phys. 48*, 3831−3832.
Wen, L., Kim, S., Tinn, T. T., Huang, J., Krishnamoorthi, R., Gong, Y., Lwin, Y. N., & Kyin, S. (1993) *Protein Express. Purif. 4*, 215−222.
Wieczorek, M., Otlewski, J., Cook, J., Parks, M., Leluk, J., Wilimowska-pelc, A., Polanowski, A., Wilusz, T., & Laskowski, M., Jr. (1985) *Biochem. Biophys. Res. Commun. 126*, 646−652.
Wijmenga, S. S., Hallenga, K., & Hilbers, C. W. (1989) *J. Magn. Reson. 84*, 634−642.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
Wynn, R., & Laskowski, M., Jr. (1990) *Biochem. Biophys. Res. Commun. 166*, 1406−1410.
Zuiderweg, E. R. P., & Fesik, S. W. (1989) *Biochemistry 28*, 2387−2391.

BI952466D