Three-Dimensional Structure of Potato Carboxypeptidase Inhibitor in Solution. A Study Using Nuclear Magnetic Resonance, Distance Geometry, and Restrained Molecular Dynamics†

G. Marius Clore,*,‡ Angela M. Gronenborn,*,‡ Michael Nilges,‡ and Clarence A. Ryan§

Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, FRG, and Department of Agricultural Chemistry, Washington State University, Pullman, Washington 99163

Received March 27, 1987; Revised Manuscript Received May 26, 1987

ABSTRACT: The solution conformation of potato carboxypeptidase inhibitor (CPI) has been investigated by ^{1}H NMR spectroscopy. The spectrum is assigned in a sequential manner by using two-dimensional NMR techniques to identify through-bond and through-space (<5 Å) connectivities. A set of 309 approximate interproton distance restraints is derived from the two-dimensional nuclear Overhauser enhancement spectra and used as the basis of a three-dimensional structure determination by a combination of metric matrix distance geometry and restrained molecular dynamics calculations. A total of 11 converged distance geometry structures were computed and refined by using restrained molecular dynamics. The average atomic root mean square (rms) difference between the final 11 structures and the mean structure obtained by averaging their coordinates is 1.4 ± 0.3 Å for residues 2–39 and 0.9 ± 0.2 Å for residues 5–37. The corresponding values for all atoms are 1.9 ± 0.3 and 1.4 ± 0.2 Å, respectively. The larger values for residues 2–38 relative to those for residues 5–37 arise from the fact that the positions of the N- (residues 1–4) and C- (residues 38–39) terminal tails are rather poorly determined, whereas those of the core of the protein (residues 5–37) are well determined by the experimental interproton distance data. The computed structures are very close to the X-ray structure of CPI in its complex with carboxypeptidase, and the backbone atomic rms difference between the mean of the computed structures and the X-ray structure is only 1.2 Å. Nevertheless, there are some real differences present which are evidenced by significant deviations between the experimental upper interproton distance limits and the corresponding interproton distances derived from the X-ray structure. These principally occur in two regions, residues 18–20 and residues 28–30, the latter comprising part of the region of secondary contacts between CPI and carboxypeptidase in the X-ray structure.

Carboxypeptidase inhibitor (CPI)[1] is a small 39-residue protein that binds tightly to both carboxypeptidase A ($K_I \sim 5 \times 10^{-9}$ M) and carboxypeptidase B ($K_I \sim 5 \times 10^{-8}$ M) (Ryan et al., 1974; Haas & Ryan, 1981). Its sequence (Haas et al., 1975) and the pairing of its three disulfide bridges (Leary et al., 1979) have been established, and a 2.5-Å resolution X-ray structure of the CPI–carboxypeptidase complex has been determined (Rees & Lipscomb, 1980, 1982). At the present time, there is no information available on the three-dimensional structure of uncomplexed CPI.

Over the past few years advances in both NMR methodology and computational techniques have made it possible to determine the three-dimensional structures of protein in solution [for detailed reviews, see Wüthrich (1986) and Clore and Gronenborn (1987)]. Although a number of structures have been solved in this manner (Braun et al., 1983, 1986; Williamson et al., 1985; Kline et al., 1986; Kaptein et al., 1985; Clore et al., 1985, 1986a, 1987a–c) and model calculations on the basis of interproton distances derived from X-ray structures have been carried out successfully by using three different approaches (Havel & Wüthrich, 1984, 1985; Braun & Go, 1986; Brünger et al., 1986; Clore et al., 1986b, 1987d), only two detailed comparisons between an X-ray structure and a set of structures derived from NMR measurements in solution have been published to date (Clore et al., 1987e,f; Wagner et al., 1987). Given that the determination of three-dimensional protein structures by NMR is still in its infancy, such are clearly of value to validate the method. CPI provides a particularly interesting test case for two reasons. First, it possesses no regular secondary structure (Rees & Lipscomb, 1980, 1982). Consequently, any approach that depends on the prior delineation of secondary structure elements derived from a qualitative interpretation of the NOE data (Wüthrich, 1984) in order to treat these as rigid bodies whose relative orientations can be modeled on the basis of long-range ($|i - j| > 5$) interproton distance restraints (Zuiderweg et al., 1984; Altman & Jardetzky, 1986) is not applicable in this instance. Further, no benefit can be gained from the fact that regular secondary structure elements are generally well-defined by short-range ($|i - j| \leq 5$) interproton distances alone (Wüthrich et al., 1984). Second, CPI provides an ideal case with which to probe possible differences in the structure of a protein inhibitor bound to its target enzyme and free in solution.

† This work was supported by the Max-Planck Gesellschaft, by Grant 321/4003/0318909A from the Bundesministerium für Forschung und Technologie, and by Grant Cl 86/1-1 from the Deutsche Forschungsgemeinschaft (G.M.C. and A.M.G.).

‡ Max-Planck-Institut.

§ Washington State University.

[1] Abbreviations: CPI, potato carboxypeptidase inhibitor; NMR, nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser enhancement or effect; NOESY, two-dimensional nuclear Overhauser enhancement spectroscopy; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; 1 cal = 4.183 J; rms, root mean square; <Glu, pyroglutamic acid; DEAE, diethylaminoethyl; DG, distance geometry; RD, restrained dynamics.

0006-2960/87/0426-8012$01.50/0

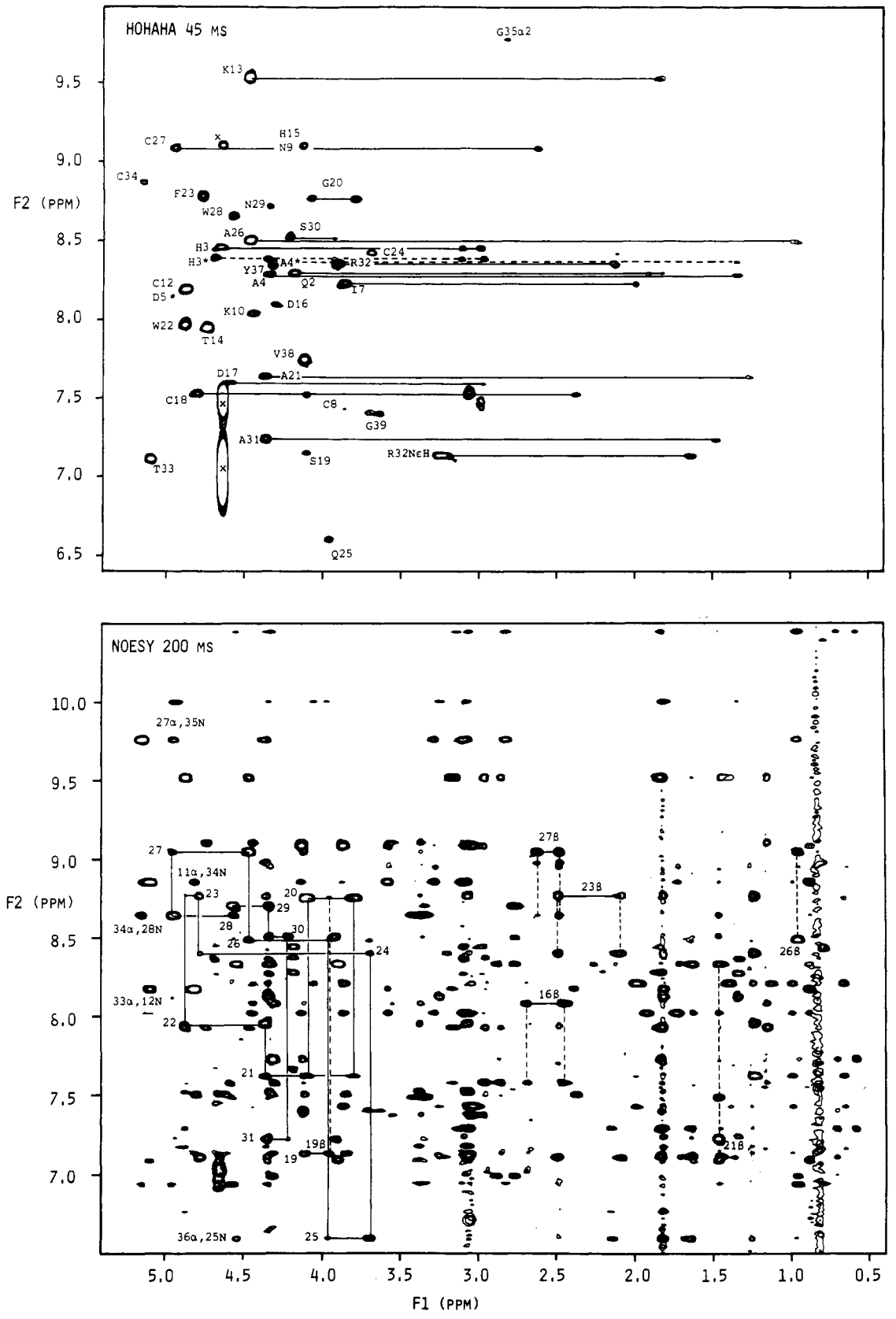

FIGURE 1: NH/aromatic (F2 axis)–aliphatic (F1 axis) region of the 45-ms HOHAHA and 200-ms NOESY spectra of CPI in H_2O. In the case of the HOHAHA spectrum some relayed connectivities are indicated by continuous lines and the labels are at the positions of the direct NH–C^{α}H cross-peaks. H3* and A4* arise from the 38- and 37-residue minor (<10% each) species present in the preparation, and the three cross-peaks labeled with an x arise from hydrogen exchange between amide protons and water occurring during the MLEV17 pulse sequence and the subsequent homospoil and recovery delays (see Experimental Procedures). In the NOESY spectrum a sequence of $C^{\alpha}H(i)$–$NH(i + 1)$ connectivities from residues 20–31 is indicated by continuous lines (—) with peaks labeled by residue number at the position of the intraresidue $C^{\alpha}H(i)$–$NH(i)$ cross-peak. Some $C^{\beta}H(i)$–$NH(i + 1)$ NOEs are also indicated by dashed lines (- - -) with the peaks labeled by residue number followed by the letter β at the position of the intraresidue $C^{\beta}H(i)$–$NH(i)$ cross-peak. In addition, a few long-range ($|i - j| > 5$) $C^{\alpha}H(i)$–$NH(j)$ cross-peaks are labeled. The spectra are unsymmetrized.

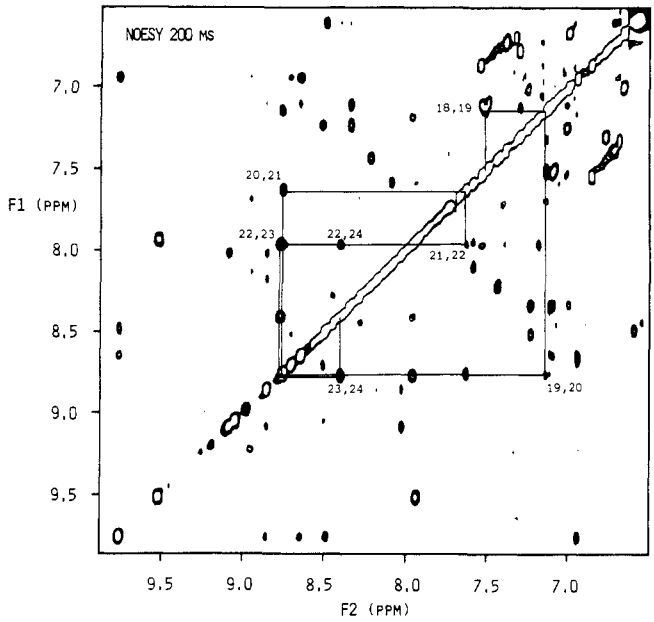

FIGURE 2: NH (F1 axis)–NH (F2 axis) of the 200-ms NOESY spectrum of CPI in H_2O. The sequence of NH(i)–NH(i + 1) connectivities extending from residues 18–24 is indicated by the continuous line (—). The spectrum is unsymmetrized.

To this end we have carried out a ^1H NMR study on CPI with the aim of determining its three-dimensional structure in solution. First, we assign the proton resonances in a sequential manner using two-dimensional NMR techniques to identify through-bond and through-space (<5 Å) connectivities. We then derive a set of 309 approximate interproton distance restraints from the NOESY spectra that form the basis of a three-dimensional structure determination using a combination of metric matrix distance geometry (Crippen & Havel, 1978; Havel et al., 1984; Havel & Wüthrich, 1984, 1985; Sippl & Scheraga, 1986) and restrained molecular dynamics (Kaptein et al., 1985; Clore et al., 1985,1986a,b; Brünger et al., 1986) calculations.

Experimental Procedures

Sample. CPI was prepared from Russet Burbank potato tubers by an adaptation of the method of Pearce and Ryan (1983). The inhibitor is a member of the heat-stable proteins in potato tubers and is soluble in 80% ethanol. After ethanol extraction and removal of ethanol by vacuum evaporation as previously described, the "crude inhibitor" fraction was dialyzed (M_r 2000 cutoff) against several changes of 50 mM ammonium bicarbonate and lyophilized. The lyophilized in-

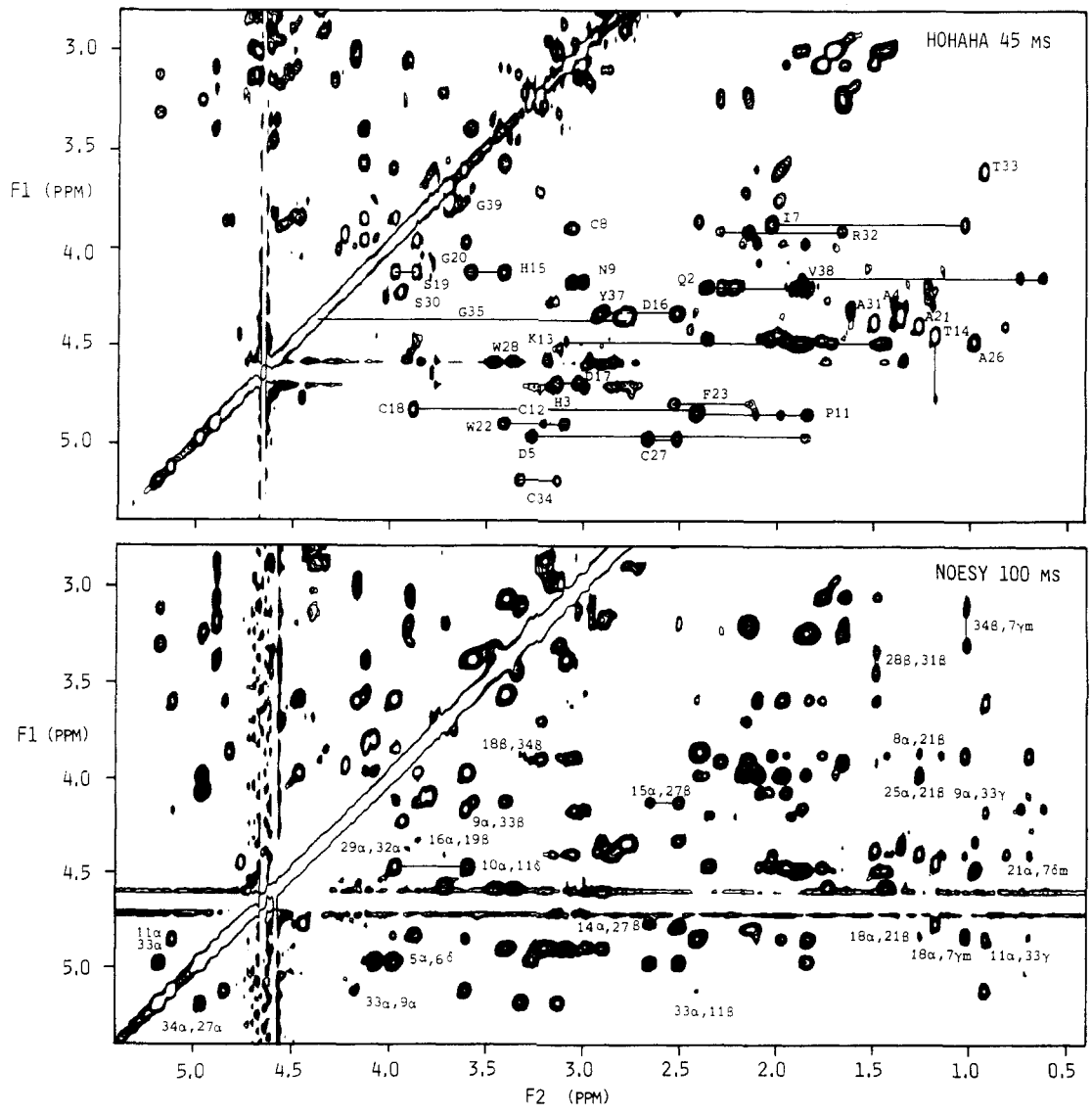

FIGURE 3: Portion of the C$^\alpha$H (F1 axis)–aliphatic (F2 axis) of the 45-ms HOHAHA and 100-ms NOESY spectra of CPI in D_2O. Direct and relayed connectivities are present in the HOHAHA spectrum, and some spin networks originating from the C$^\alpha$H protons are indicated by continuous lines. Cross-peaks arising from long-range ($|i - j| > 5$) NOEs are labeled in the NOESY spectrum. The spectra are unsymmetrized.

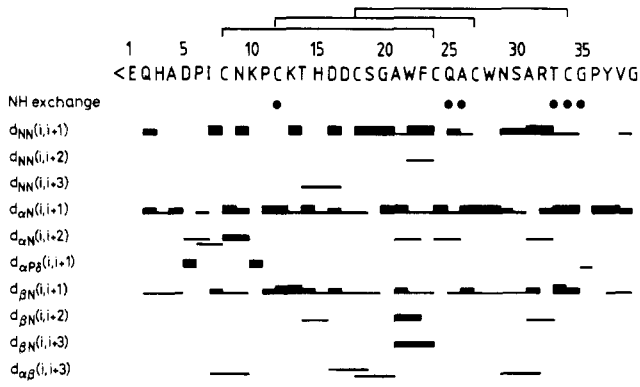

FIGURE 4: Sequence of CPI together with a summary of the observed short range ($|i - j| \leq 5$) NOEs involving the NH, $C^{\alpha}H$, and $C^{\beta}H$ protons as well as the $C^{\delta}H$ protons of proline residues. The NOEs are classified accordingly as strong, medium, or weak by the thickness of the line. NH protons that are still present after dissolving the protein in D_2O are indicated by closed circles (●). The location of the disulfide bridges is indicated above the sequence.

hibitor was then chromatographed on sulfopropyl-Sephadex (Pearce & Ryan, 1983), and the fractions containing the inhibitor were pooled, dialyzed against several changes of distilled water, and lyophilized. The inhibitor preparation had a slight brown coloration that was removed by dissolving the inhibitor in water and passing it through a 1.5 × 15 cm column of DEAE, previously washed thoroughly with water. The breakthrough fractions containing the inhibitor were pooled and dialyzed (M_r 2000 cutoff) against several changes of distilled water and lyophilized. No attempts were made to separate the major (>80%) CPI species (39 residues <Glu-Gln-His-Ala-Asp...) from the two minor (<10% each) isoinhibitors (38 residues <Glu-His-Ala-Asp... and 37 residues His-Ala-Asp...) present in the preparation (Haas & Ryan, 1981). The inhibitor migrated as a single, homogeneous band on SDS–polyacrylamide gels.

The samples for NMR contained 7 mM CPI in either 90% H_2O/10% D_2O or 99.996% D_2O at pH 3.82. All experiments were carried out at 35 °C.

NMR Spectroscopy. NMR spectra were recorded on a Bruker AM500 spectrometer equipped with digital phase shifters and an ASPECT 3000 computer. All two-dimensional spectra were recorded in the pure phase absorption mode by using the time proportional incrementation method (Redfield & Kuntz, 1975; Bodenhausen et al., 1980) as described by Marion and Wüthrich (1983). The following spectra were recorded in D_2O and H_2O: NOESY (Jeener et al., 1979; Macura et al., 1981) and MLEV17 HOHAHA (Davis & Bax, 1985; Bax & Davis, 1985). NOESY spectra were recorded with mixing times of 100 and 200 ms and the HOHAHA spectra at several mixing times ranging from 18 to 65 ms. For measurements in H_2O, the water resonance was suppressed by a semiselective excitation pulse sequence. In the case of the NOESY spectra this involved replacing the last 90° pulse in the sequence by the jump–return sequence (90_x–τ–90_{-x}) with the carrier placed at the position of the solvent (Plateau & Gueron, 1982) and a value of 80 μs for τ. In the case of the HOHAHA spectra this involved the addition of the pulse sequence 90_x–*H*–*d*–90_x–τ–90_{-x}–Δ–90_{ϕ}–2τ–$90_{-\phi}$–Δ at the end of the $MLEV17_y$ sequence (Sklenar & Bax, 1987; Bax et al., 1987), where *H* is a strong homospoil pulse (8 ms), *d* a recovery delay (8 ms), 90_x–τ–90_{-x} the jump–return sequence with a value of 80 μs for τ, and Δ a short delay (100 μs) at the beginning and end of the refocusing echo sequence 90_{ϕ}–2τ–$90_{-\phi}$ in which the phase ϕ is cycled along the *x*, *y*, *−x*, and *−y* axes in the exorcycle manner.

Table I: Proton Resonance Assignment of CPI at pH 3.8 and 35 °C

	chemical shift (ppm)			
residue	NH	α	β	others
<E1[a]				
Q2	8.29	4.18	2.21, 1.85	$C^{\gamma}H$ 2.32, 2.25
H3	8.46	4.67	3.09, 2.98	$C^{\delta}H$ 7.10; $C^{\epsilon 1}H$ 8.40
A4	8.19	4.23	1.32	
D5	8.14	4.94	3.25, 1.82	
P6		4.43	2.31, 1.98	$C^{\gamma}H$ 2.05, 1.90; $C^{\delta}H$ 4.03, 3.94
I7	8.21	3.85	1.99	$C^{\gamma}H$ 1.40, 1.13; $C^{\gamma}H_3$ 0.99; $C^{\delta}H_3$ 0.66
C8	7.43	3.86	2.25, 3.04	
N9	9.09	4.13	3.04, 2.97	NH_2 7.34, 6.59
K10	8.03	4.44	1.94, 1.75	$C^{\gamma}H$ 1.64, 1.47; $C^{\delta}H$ 1.72; $C^{\epsilon}H$ 3.04; NH_3^+ 7.52
P11		4.81	2.37, 1.79	$C^{\gamma}H$ 2.06, 1.94; $C^{\delta}H$ 3.94, 3.56
C12	8.18	4.87	3.19, 2.96	
K13	9.53	4.46	1.84, 1.84	$C^{\gamma}H$ 1.60, 1.45; $C^{\delta}H$ 1.66; $C^{\epsilon}H$ 2.95; NH_3^+ 7.44
T14	7.95	4.73	4.41	$C^{\gamma}H_3$ 1.14
H15	9.10	4.11	3.53, 3.36	$C^{\delta 2}H$ 7.42; $C^{\epsilon 1}H$ 8.58
D16	8.09	4.30	2.69, 2.43	
D17	7.60	4.59	2.96, 2.86	
C18	7.52	4.80	3.84, 2.37	
S19	7.13	4.10	3.94, 3.84	
G20	8.75	4.09, 3.79		
A21	7.63	4.26	1.23	
W22	7.96	4.87	3.37, 3.07	$C^{\epsilon 3}H$ 7.54; $C^{\zeta 2}H$ 7.27; $C^{\zeta 3}H$ 7.11; $C^{\eta}H$ 7.02; $N^{\epsilon 1}H$ 10.00
F23	8.78	4.77	2.49, 2.09	$C^{\delta}H$ 7.12; $C^{\epsilon}H$ 7.32; $C^{\zeta}H$ 7.24
C24	8.41	3.69	3.18, 2.11	
Q25	6.60	3.96	1.68, 1.63	$C^{\gamma}H$ 2.14, 1.82; NH_2 7.30, 6.77
A26	8.49	4.46	0.96	
C27	9.05	4.94	2.61, 2.47	
W28	8.65	4.56	3.42, 3.34	$C^{\epsilon 3}H$ 7.52; $C^{\zeta 2}H$ 7.33; $C^{\zeta 3}H$ 7.16; $C^{\eta}H$ 7.11; $N^{\epsilon 1}H$ 10.45
N29	8.72	4.23	2.92, 2.76	NH_2 7.55, 6.86
S30	8.52	4.21	3.92, 3.92	
A31	7.23	4.35	1.47	
R32	8.35	3.89	2.26, 2.12	$C^{\gamma}H$ 1.62; $C^{\delta}H$ 3.25, 3.18; $N^{\epsilon}H$ 7.13
T33	7.10	5.09	3.57	$C^{\gamma}H_3$ 0.89
C34	8.86	5.14	3.28, 3.09	
G35	9.77	4.27, 2.82		
P36		4.54	1.70, 1.30	$C^{\gamma}H$ 1.60, 1.39; $C^{\delta}H$ 3.11, 3.03
Y37	8.34	4.22	2.87, 2.76	$C^{\delta}H$ 6.99; $C^{\epsilon}H$ 6.66
V38	7.74	4.12	1.83	$C^{\gamma}H_3$ 0.70, 0.58
G39	7.40	3.71, 3.63		
H3*[b]	8.39	4.68	3.09, 2.96	
A4*[b]	8.37	4.23	1.31	

[a] <E1 is pyroglutamic acid, and its resonances were not assigned. [b] H3* is the histidine of the 38-residue minor species <Glu-His-Ala-Asp...; A4* is the alanine of the 37-residue minor species His-Ala-Asp-.... No difference in the chemical shifts of the other residues could be detected between these two minor species (<10% each) and the major (>80%) 39-residue species.

Calculations. Metric matrix distance geometry calculations were carried out with the program DISGEO (Havel & Wüthrich, 1984, 1985; Havel, 1986). All energy minimization and restrained molecular dynamics calculations were carried out as described previously (Clore et al., 1986a,b) on a CRAY-XMP using the program XPLOR (A. T. Brünger, unpublished data), which is a vectorized version of the program CHARMM (Brooks et al., 1983) especially adapted for restrained molecular dynamics. Displaying of the structures was carried out using a modified version of the function network of FRODO (Jones, 1978) interfaced with CHARMM on an Evans & Sutherland PS330 color graphics system. The smooth backbone (N, C^{α},

Table II: Protocol of the Refinement Stage[a]

	⟨DG⟩
	↓
phase 1	600-cycle restrained energy minimization c = 40 kcal mol^{-1} Å^{-2}
	↓
	⟨DGm⟩
	↓
phase 2	4-ps restrained dynamics[b] T = 300–800 K c = 1.25–40 kcal mol^{-1} Å^{-2}
	↓
	2-ps restrained dynamics T cooled to 300 K c = 40 kcal mol^{-1} Å^{-2}
	↓
	400-cycle restrained energy minimization c = 40 kcal mol^{-1} Å^{-2}
	↓
	⟨RDDG⟩

[a] The notation of the structures is as follows: ⟨DG⟩ comprise the 11 converged distance geometry structures, ⟨DGm⟩ the structures derived from the DG structures by restrained energy minimization, and ⟨RDDG⟩ the structure derived from the DGm structures by restrained molecular dynamics. [b] The temperature of the system was adjusted to lie between 300 and 800 K by scaling the velocities of the atoms upward by a factor of 1.25 if the temperature fell below 300 K and downward by a factor of 0.75 if the temperature rose above 800 K. The velocity scaling was carried out every 0.25 ps. The NOE restraints force constants, c, were increased from 1.25 kcal mol^{-1} Å^{-2} to a maximum value of 40 kcal mol^{-1} Å^{-2} by doubling their value every 0.25 ps.

C) atom representations were obtained as described previously (Feldmann et al., 1986).

Results and Discussion

Assignment of the ^1H NMR Spectrum. Sequence specific resonance assignments were made in a sequential manner (Wüthrich et al., 1982; Billeter et al., 1982; Wüthrich, 1986). The HOHAHA spectra, recorded at several mixing times in order to demonstrate successively direct, single-relayed, and multiple-relayed through-bond connectivities (Davis & Bax, 1985; Braunschweiler & Ernst, 1985) were used to identify spin systems. The NOESY spectra were used to demonstrate through-space (<5 Å) connectivities in order to sequentially assign resonances via short-range ($|i - j| \leq 5$) NOEs involving the NH, C$^\alpha$H, and C$^\beta$H protons as well as the C$^\delta$H protons of proline. For this purpose, the most useful NOEs are the NH(i)–NH(i + 1), C$^\alpha$H(i)–NH(i + 1), C$^\beta$H(i)–NH(i + 1), C$^\alpha$H(i)–NH(i + 2), C$^\alpha$H(i)–NH(i + 3), and C$^\alpha$H(i)–C$^\beta$H(i + 3) NOEs. Examples of HOHAHA and NOESY spectra are shown in Figures 1–3, a summary of the short-range NOEs is shown in Figure 4, and the complete list of assignments is given in Table I.

From the assignment viewpoint, two features are noteworthy. First, all residues were assigned completely with the exception of the N-terminal pyroglutamic acid (<Glu). Second, the chemical shifts of the N-terminal but one residue are slightly different in the two shorter minor inhibitor species than in the major one. Thus the chemical shifts of the NH and C$^\alpha$H protons of His-3 are slightly different in the 38 residue <Glu-His-Ala-Asp... minor species than in the 39 residue <Glu-Gln-His-Ala-Asp... major species, and the chemical shifts of the NH and C$^\alpha$H protons of Ala-4 are

Table III: Atomic rms Differences[a]

	atomic rms differences (Å)			
	residues 2–39		residues 5–37	
	backbone atoms	all atoms	backbone atoms	all atoms
	(A) rms Distributions			
⟨DG⟩ vs ⟨DG⟩	1.7 ± 0.2	2.5 ± 0.3	1.6 ± 0.2	2.2 ± 0.3
⟨DGm⟩ vs ⟨DGm⟩	1.7 ± 0.2	2.4 ± 0.3	1.5 ± 0.2	2.1 ± 0.2
⟨RDDG⟩ vs ⟨RDDG⟩	2.1 ± 0.4	2.8 ± 0.4	1.3 ± 0.2	2.0 ± 0.3
⟨DG⟩ vs $\overline{\text{DG}}$	1.2 ± 0.1	1.7 ± 0.2	1.1 ± 0.1	1.5 ± 0.2
⟨DGm⟩ vs $\overline{\text{DGm}}$	1.1 ± 0.1	1.7 ± 0.2	1.0 ± 0.1	1.4 ± 0.2
⟨RDDG⟩ vs $\overline{\text{RDDG}}$	1.4 ± 0.3	1.9 ± 0.3	0.9 ± 0.2	1.4 ± 0.2
⟨DG⟩ vs ($\overline{\text{DG}}$)m	1.3 ± 0.1	1.9 ± 0.2	1.2 ± 0.1	1.6 ± 0.2
⟨DGm⟩ vs ($\overline{\text{DGm}}$)m	1.3 ± 0.1	2.8 ± 0.2	1.1 ± 0.1	1.5 ± 0.2
⟨RDDG⟩v vs ($\overline{\text{RDDG}}$)m	1.5 ± 0.3	2.1 ± 0.3	1.0 ± 0.2	1.5 ± 0.2
	(B) rms Shifts			
⟨DG⟩ vs ⟨DGm⟩	0.4 ± 0.04	0.4 ± 0.03	0.4 ± 0.07	0.5 ± 0.08
⟨DGm⟩ vs ⟨RDDG⟩	2.2 ± 0.3	2.7 ± 0.4	1.5 ± 0.1	2.0 ● 0.2
⟨DG⟩ vs ⟨RDDG⟩	2.3 ± 0.3	2.8 ± 0.4	1.6 ± 0.1	2.1 ± 0.2
$\overline{\text{DG}}$ vs $\overline{\text{DGm}}$	0.2	0.3	0.2	0.3
$\overline{\text{DGm}}$ vs $\overline{\text{RDDG}}$	1.4	1.7	0.9	1.2
$\overline{\text{DG}}$ vs $\overline{\text{RDDG}}$	1.5	1.8	1.0	1.4
$\overline{\text{DG}}$ vs ($\overline{\text{DG}}$)m	0.5	0.8	0.5	0.6
$\overline{\text{DGm}}$ vs ($\overline{\text{DGm}}$)m	0.5	0.7	0.5	0.7
$\overline{\text{RDDG}}$ vs ($\overline{\text{RDDG}}$)m	0.6	0.9	0.4	0.6
($\overline{\text{DG}}$)m vs ($\overline{\text{DGm}}$)m	0.4	0.5	0.3	0.5
($\overline{\text{DGm}}$)m vs ($\overline{\text{RDDG}}$)m	1.6	2.0	0.9	1.4
($\overline{\text{DG}}$)m vs ($\overline{\text{RDDG}}$)m	1.6	2.1	1.1	1.5

[a] The notation of the structures is as follows: ⟨DG⟩ comprise the 11 converged distance geometry structures, ⟨DGm⟩ the structures derived from the DG structures by restrained energy minimization, and ⟨RDDG⟩ the structure derived from the DGm structures by restrained molecular dynamics (see text). $\overline{\text{DG}}$, $\overline{\text{DGm}}$, and $\overline{\text{RDDG}}$ are the mean structures obtained by averaging the coordinates of the DG, DGm, and RDDG structures, respectively, best fitted to residues 2–39. The estimated standard atomic rms errors s_{mean} of these mean structures is given by $[\sum(\text{rmsd}_i)^2/n(n-1)]^{1/2}$, where rmsd_i is the atomic rms difference between the ith structure and the mean structure and n is the number of structures. ($\overline{\text{DG}}$)m, ($\overline{\text{DGm}}$)m, and ($\overline{\text{RDDG}}$)m, are the structures obtained by restrained energy minimization of the mean $\overline{\text{DG}}$, $\overline{\text{DGm}}$, and $\overline{\text{RDDG}}$ structure, respectively.

Table IV: Interproton Distance Deviations and Radii of Gyration[a]

	rms difference between calculated and target interproton distance restraints (Å)				
		interresidue			
structure	all (309)	short range ($\|i-j\| \leq 5$) (127)	long range ($\|i-j\| > 5$) (82)	intraresidue (100)	radii of gyration (Å)
⟨DG⟩	0.46 ± 0.05	0.35 ± 0.06	0.74 ± 0.08	0.26 ± 0.03	10.10 ± 0.10
⟨DGm⟩	0.14 ± 0.02	0.11 ± 0.03	0.19 ± 0.04	0.12 ± 0.01	10.00 ± 0.11
⟨RDDG⟩	0.07 ± 0.01	0.06 ± 0.02	0.09 ± 0.01	0.07 ± 0.01	9.08 ± 0.18
$\overline{DG}$	0.34	0.19	0.57	0.23	9.94
$\overline{DGm}$	0.20	0.10	0.29	0.22	9.85
$\overline{RDDG}$	0.15	0.15	0.16	0.25	8.83
$(\overline{DG})$m	0.14	0.11	0.18	0.13	10.03
$(\overline{DGm})$m	0.11	0.09	0.13	0.11	9.97
$(\overline{RDDG})$m	0.07	0.05	0.09	0.07	9.02
X-ray	0.58	0.51	0.87	0.29	9.29
X-ray–EM	0.54	0.45	0.83	0.29	9.32
X-ray–RM	0.14	0.10	0.23	0.09	9.21

[a] The notation of the structures is the same as that in Table III. In addition, X-ray, X-ray–EM, and X-ray–RM are the X-ray structure of CPI in the CPI–carboxypeptidase complex (Rees & Lipscomb, 1982), the energy minimized X-ray sructure, and the restrained energy minimized X-ray structures, respectively (i.e., the force constants for the NOE restraints energy are 0 kcal mol^{-1} Å^{-2} for X-ray–EM and 40 kcal mol^{-1} Å^{-2} for X-ray–RM). The rms difference (rmsd) between the calculated (r_{ij}) and target restraints is calculated with respect to the upper (r_{ij}^u) andd lower (r_{ij}^l) limits such that rmsd = $[(r_{ij} - r_{ij}^u)^2/n]^{1/2}$ if $r_{ij} > r_{ij}^u$, rmsd = 0 if $r_{ij}^l \leq r_{ij} \leq r_{ij}^u$, and rmsd = $[(r_{ij} - r_{ij}^l)^2/n]^{1/2}$ if $r_{ij} < r_{ij}^l$.

Table V: Energies of the Structures[a]

	energy (kcal/mol)									
structure	total	potential	bond[b] (568)	angle[b] (1014)	dihedral[b] (244)	improper (173)	van der Waals	electro-static	H bond	NOE restraints[c] (309)
⟨DG⟩	3707 ± 1461[d]	898 ± 313[d]	82 ± 8	289 ± 46	280 ± 18	0.8 ± 1.0	368 ± 1257[d]	−26 ± 34	−5 ± 2	2720 ± 641
⟨DGm⟩	674 ± 223	425 ± 143	73 ± 26	375 ± 72	235 ± 17	0.1 ± 0.05	−22 ± 34	−221 ± 40	−13 ± 3	249 ± 83
⟨RDDG⟩	−107 ± 139	−173 ± 116	40 ± 14	270 ± 40	185 ± 9	0.04 ± 0.01	−100 ± 28	−535 ± 45	−34 ± 9	67 ± 25
$\overline{DG}$	>10^6	>10^6	21642	5273	472	0.6	>10^6	5159	−10	1461
$\overline{DGm}$	>10^6	>10^6	21217	4995	491	0.6	>10^6	−152	−14	508
$\overline{RDDG}$	>10^6	>10^6	38368	8309	331	6.2	>10^6	−3000	183	465
$(\overline{DG})$m	1032	801	120	571	292	0.4	35	−199	−18	231
$(\overline{DGm})$m	404	267	46	362	247	0.2	−65	−302	−23	137
$(\overline{RDDG})$m	154	94	39	512	217	0.05	−80	−576	−27	60
X-ray	(5448)[f]	1351	117	405	296	3.1	651[e]	−105	−15	(4097)[f]
X-ray–EM	(3619)[f]	33	19	183	224	0.02	−120	−239	−34	(3586)[f]
X-ray–RM	451	207	72	324	204	0.04	−69	−290	−34	244

[a] The notation of the structures if the same as that in Tables III and IV. The total energy is the sum of the potential and NOE restraints energies, and the potential energy is made up of all the other bonded and nonbonded energy terms. The number of terms for the bond, angle, dihedral, and improper dihedral (planarity) potentials and for the effective NOE interproton distance restraints potential is given in parentheses. [b] The bond, angle, and dihedral potentials for the three disulfide bridges are included in these terms. [c] The restraints force constants [cf. eq 1 of Clore et al. (1986a)] have values of 40 kcal mol^{-1} Å^{-2}. [d] The ranges of the total, potential, and van der Waals energies of the DG structures extend from 2501 to 7565 kcal/mol, from 472 to 4941 kcal/mol, and from −67 to 4248 kcal/mol, respectively. [e] The high van der Waals energy of the X-ray structure arises from five bad nonbonded contacts (see text). [f] As the NOE restraints energy was not included in the energy function used to obtain X-ray–EM from the X-ray structure, the total energy and NOE restraints energy are given in parentheses.

slightly different in the 37 residue His-Ala-Asp... minor species than in the two other species (the numbering of the residues being that of the major species). Other than these few minor differences, no difference in chemical shifts could be detected between the three inhibitor species.

Interproton Distance Restraints. The basis of the structure determination consisted of a set of 309 structurally useful NOEs comprising 127 short- ($|i-j| \leq 5$) and 82 long- ($|i-j| > 5$) range interresidue NOEs and 100 intraresidue NOEs. These NOEs were first identified in the 200-ms mixing time NOESY spectra and then classified into three intensity classes (Williamson et al., 1985), strong, medium, and weak, on the basis of the 100-ms mixing time NOESY spectra. These three classes correspond to distance ranges of 1.8–2.7, 1.8–3.3, and 1.8–5.0 Å, respectively. In the case of NOEs involving methyl groups, an additional 0.5 Å per methyl group was added to the upper distance limit to account for the higher apparent intensity of methyl resonances. The complete list of all NOEs used in the computations is available as supplementary material.

The NOE restraints were supplemented by nine distance restraints for the three disulfide bridges between Cys-8 and Cys-24, Cys-12 and Cys-27, and Cys-18 and Cys-34. For each disulfide bridge there are three distance restraints, S_i–S_j, S_i–C^β_j, and S_j–C^β_i, which are restrained to values of 2.02 ± 0.01, 2.99 ± 0.01, and 2.99 ± 0.01 Å, respectively.

Tertiary Structure Computation. The computation of the tertiary structure employed the same two-stage approach that we previously used for α1-purothionin (Clore et al., 1986a,b), phoratoxin (Clore et al., 1987a), hirudin (Clore et al., 1987b), and the globular domain of histone H5 (Clore et al., 1987c), namely, a structure generation phase using the metric matrix distance geometry program DISGEO (Havel, 1986), followed by a refinement phase using a combination of restrained energy

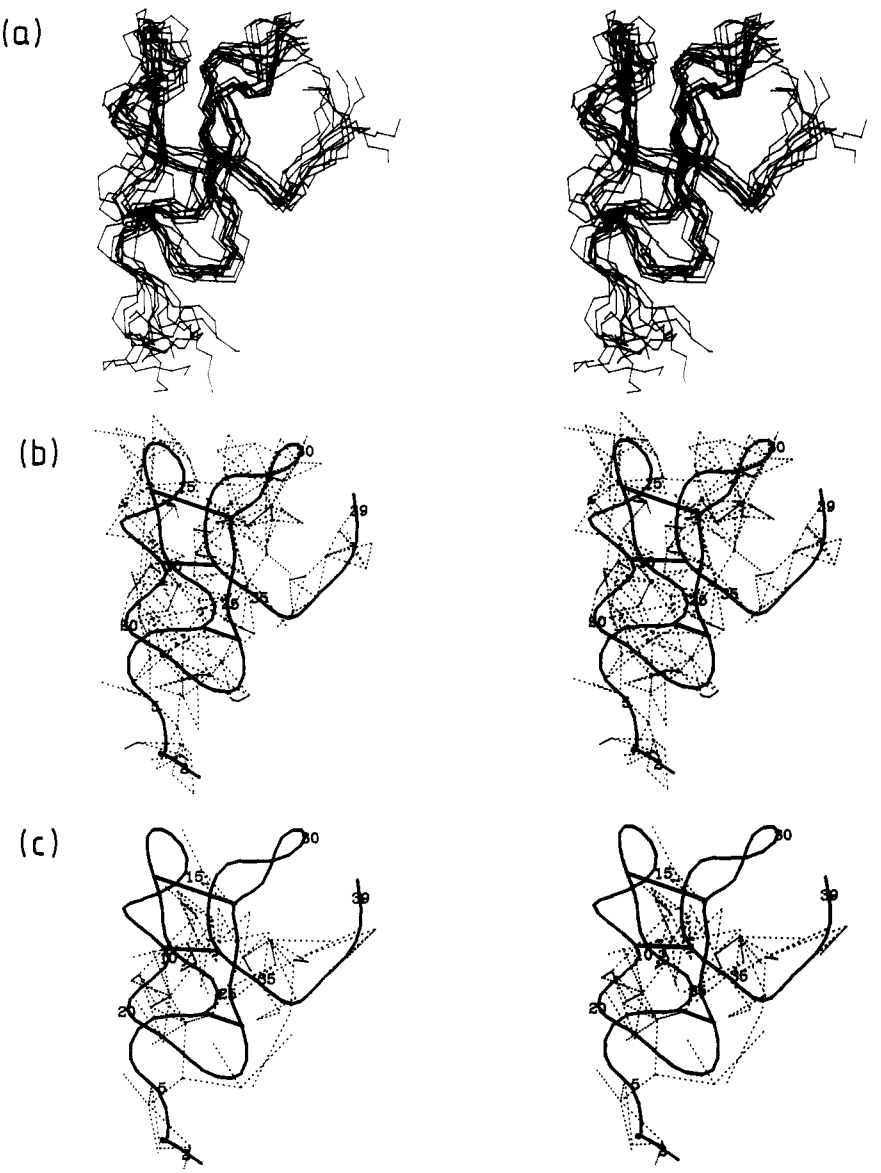

FIGURE 5: (a) Best-fit superposition of the backbone (C, C^{α}, N) atoms of the 11 converged RDDG structures. (b and c) The short- and long-range interproton distance restraints shown as dashed lines, respectively, superimposed on a framework comprising a smoothed backbone (C, C^{α}, N) atom representation of the restrained energy minimized average structure $\overline{(RDDG)}$m.

minimization and restrained molecular dynamics in which the NOE interproton distances are incorporated into the total energy function of the system in the form of effective potentials (Kaptein et al., 1985; Clore et al., 1985, 1986b; Brünger et al., 1986).

A total of 11 converged DG structures were obtained from the DISGEO calculations. In these calculations all the restraints were included explicitly, and the upper limits of distances involving methyl and methylene protons were corrected for the pseudoatom representation employed by DISGEO as described by Wüthrich et al. (1983). In the refinement stage, no corrections to the upper distance limits were used as these distances were referred to single $(\langle r^{-6} \rangle)^{-1/6}$ average distances, the form of the NOE restraints effective potential was a square-well (Clore et al., 1986a), and the disulfide bridge restraints were no longer included explicitly as they were implictly contained in the bond, angle, and dihedral potentials of the total empirical energy function. The refinement stage comprised two phases, the first a restrained energy minimization phase to generate the DGm structures and the second a restrained molecular dynamics phase to generate the RDDG structures. The details of the refinement procedure are summarized in Table II.

In addition, the coordinates of the DG, DGm, and RDDG structures, best fitted to each other (excluding residue 1, as no NOEs to residue 1 were identified), were averaged to give the mean structures $\overline{DG}$, $\overline{DGm}$, and $\overline{RDDG}$, respectively (Clore et al., 1986a,b). As these mean structures are poor in stereochemical and energetic terms, they were subjected to 600 cycles of restrained energy minimization in which the van der Waals radii were increased gradually from one-fourth of their usual values to their full values to generate the structures $\overline{(DG)}$m, $\overline{(DGm)}$m, and $\overline{(RDDG)}$m, respectively (Clore et al., 1986b).

Converged Structures. A total of 11 DG structures were obtained and subjected to refinement. The course of the refinement is summarized in Table III–V, which give the atomic rms distributions and shifts, the interproton distance deviations and radii of gyration, and the energies of the structures, respectively. The best-fit superposition of the backbone (C, C^{α}, N) atoms of the 11 RDDG structures is

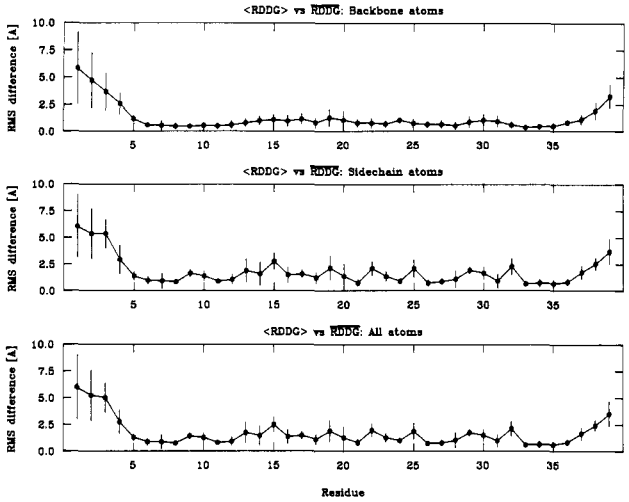

FIGURE 6: Atomic rms distribution of the backbone (C, C^{α}, N, O) atoms, side-chain atoms, and all atoms of the 11 RDDG structures about the mean structure $\overline{\text{RDDG}}$ best fitted to residues 2–39. The closed circles (●) represent the average rms difference at each residue between the RDDG structures and the mean $\overline{\text{RDDG}}$ structure, and the bars represent the standard deviations in these values.

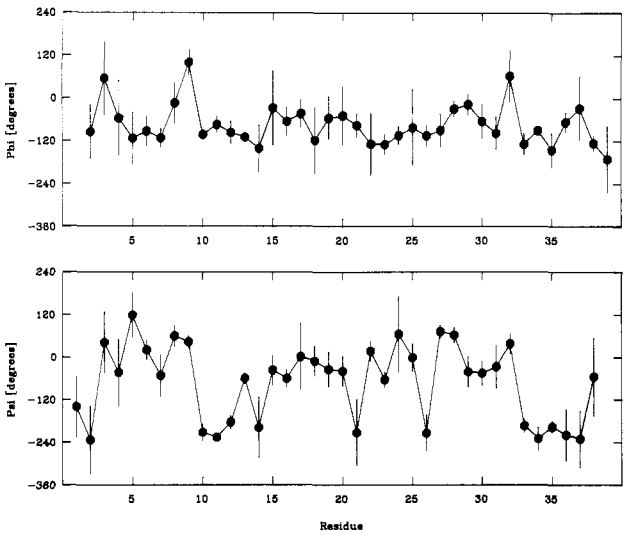

FIGURE 7: Angular rms distribution of the ϕ and ψ backbone torsion angles of the RDDG structures. The closed circles (●) are the values of the ϕ and ψ angles of the restrained energy minimized average structure ($\overline{\text{RDDG}}$)m, and the bars represent the average angular rms deviations between all pairs of RDDG structures.

shown in Figure 5 together with the superposition of the short- and long-range interproton distances on the average restrained energy minimized structures ($\overline{\text{RDDG}}$)m.

The effect of refinement is to produce a set of structures that not only satisfy the experimental interproton distance restraints better (Table IV) but also are better in stereochemical and energetic terms (Table V). The initial restrained energy minimization to generate the DGm structures simply takes the DG structures into the next local energy subminimum and hence is accompanied by only small atomic rms shifts (<0.5 Å; Table III). In contrast, restrained molecular dynamics produces much larger atomic rms shifts (~2.3 Å for the backbone atoms of residues 2–39; Table III) and locates significantly lower energy local subminima within the global minimum region (Table V). Thus the average NOE restraints and total nonbonding energies of the RDDG structures are ~182 and ~413 kcal/mol lower, respectively, than those of the DGm structures. These general features are reflected not

Table VI: Difference between the Experimentally Derived Interproton Distance Restraints in Solution and the Interproton Distances Derived from the X-ray Structure of CPI

obsd NOE	obsd NOE intensity[a]	r_{ij} (X-ray)[b] (Å)	Δ[c] (Å)	r_{ij} (X-ray–RM)[b] (Å)	Δ[c] (Å)
Short-Range ($\|i-j\| \leq 5$) Interresidue					
NH(i)–NH($i+1$)					
C18, S19	s	3.5	−0.8	2.8	−0.1
S19, G20	s	4.0	−1.3	3.0	−0.3
G20, A21	s	3.4	−0.7	2.2	0.0
W22, F23	s	3.6	−0.9	2.8	−0.1
R32, T33	s	3.8	−1.1	2.9	−0.2
NH(i)–NH($i+2$)					
W22, C24	w	5.9	−0.9	4.4	0.0
C^{α}H(i)–NH($i+1$)					
A21, W22	s	3.4	−0.7	3.1	−0.4
C^{α}H(i)–C^{β}H($i+3$)					
D16, S19	w	7.6	−2.6	5.3	−0.3
C18, A21	w	6.3	−1.3	5.1	−0.1
C^{β}H(i)–NH($i+1$)					
C12, K13	s	3.8	−1.1	3.0	−0.3
K13, T14	s	3.4	−0.7	2.7	0.0
A21, W22	m	4.2	−0.7	3.3	0.0
A31, R32	s	4.4	−1.2	3.4	−0.2
T33, C34	s	4.6	−1.9	2.9	−0.2
others					
A21(C^{β}H)–Q25(C^{α}H)	s	4.3	−1.6	3.0	−0.3
I7($C^{\delta}H_3$)–C8(NH)	w	6.3	0.8	5.5	0.0
I7($C^{\delta}H_3$)–K10(C^{δ}H)	s	5.4	−2.2	3.8	0.0
T33($C^{\gamma}H_3$)–C34(NH)	s	3.9	−0.7	2.3	0.0
Long-Range ($\|i-j\| > 5$) Interresidue					
C8(C^{α}H)–A21($C^{\beta}H_3$)	m	4.9	−1.1	4.2	−0.4
N9(C^{α}H)–T33(C^{α}H)	m	4.1	−0.8	3.9	−0.6
N9(C^{α}H)–T33(C^{β}H)	s	4.2	−1.5	3.3	−0.6
P11(C^{α}H)–T33($C^{\gamma}H_3$)	s	5.6	−2.4	2.9	0.0
C12(NH)–T33($C^{\gamma}H_3$)	m	6.3	−2.5	4.1	−0.3
A21(NH)–I7($C^{\gamma}H_3$)	m	5.1	−1.3	3.8	0.0
A21(NH)–I7($C^{\delta}H_3$)	m	4.7	−0.9	4.1	−0.3
A21(C^{β}H)–I7($C^{\gamma}H_3$)	s	4.6	−0.9	2.9	0.0
F23(C^{δ}H)–D5(C^{β}H)	s	3.7	−1.0	3.1	−0.4
F23(C^{δ}H)–P36(C^{γ}H)	m	4.4	−1.1	3.5	−0.2
W28(NH)–C34(C^{α}H)	m	4.1	−0.8	3.4	−0.1
W28($C^{\delta 1}$H)–G35(NH)	m	4.0	−0.7	4.3	0.0
T33(C^{α}H)–I7($C^{\delta}H_3$)	w	11.2	−5.7	7.0	−1.5
C34(NH)–N9(C^{α}H)	m	4.3	−1.0	2.9	0.0
C34(NH)–P11(C^{α}H)	m	4.0	−0.7	3.7	−0.4
Intraresidue					
I7:C^{α}H–$C^{\delta}H_3$	s	4.0	−0.8	2.7	0.0
W22:C^{α}H–$C^{\delta 1}$H	m	4.6	−1.3	3.6	−0.3
W22:C^{α}H–$C^{\epsilon 3}$H	m	4.2	−0.9	3.7	−0.4
W28:C^{α}H–$C^{\epsilon 3}$H	m	4.7	−1.4	2.1	0.0

[a] The NOE intensities are classified into strong (s), medium (m), and weak (w) and are taken to correspond to distance ranges of 1.8–2.7, 1.8–3.3, and 1.8–5.0 Å, respectively. In the case of NOEs involving a methyl group(s), an additional 0.5 Å per methyl group is added to the upper distance limit to account for the higher intensity of methyl resonances. [b] X-ray is the X-ray structure of CPI in the complex with carboxypeptidase (Rees & Lipsomb, 1982). X-ray–RM is the restrained energy minimized X-ray structure obtained with the interproton distance limits derived from the experimental NOE measurements as restraints. [c] Δ represents the differences between the upper distance limits deduced from the NOE measurements on the one hand and the X-ray and restrained energy minimized X-ray structure on the other.

only in the energies of the individual DG, DGm, and RDDG structures but also in those of the restrained energy minimized average structures ($\overline{\text{DG}}$)m, ($\overline{\text{DGm}}$)m, and ($\overline{\text{RDDG}}$)m (Table V). In energetic terms, the ($\overline{\text{DG}}$)m and ($\overline{\text{DGm}}$)m structures are slightly better than the individual DG and DGm structures, whereas the ($\overline{\text{RDDG}}$)m structure is comparable to the individual RDDG structures (Table V). In terms of atomic rms displacements, however, the restrained energy minimized av-

Table VII: Atomic rms Differences between the Solution and X-ray Structures of CPI[a]

	atomic rms difference (Å)							
	X-ray				X-ray–RM			
	residues 2–38		residues 5–37		residues 2–38		residues 5–37	
	backbone atoms	all atoms	backbone atoms	all atoms	backbone atoms	all atoms	backbone atoms	all atoms
⟨DG⟩	1.8 ± 0.2	2.7 ± 0.2	1.5 ± 0.2	2.4 ± 0.2	1.8 ± 0.2	2.7 ± 0.2	1.5 ± 0.2	2.4 ± 0.2
⟨DGm⟩	1.8 ± 0.2	2.7 ± 0.2	1.5 ± 0.2	2.3 ± 0.2	1.8 ± 0.2	2.6 ± 0.2	1.4 ± 0.2	2.3 ± 0.2
⟨RDDG⟩	2.1 ± 0.2	3.0 ± 0.3	1.5 ± 0.2	2.4 ± 0.3	2.0 ± 0.2	2.9 ± 0.3	1.4 ± 0.2	2.3 ± 0.3
$\overline{DG}$	1.4	2.2	1.1	1.9	1.4	2.1	1.1	1.9
$\overline{DGm}$	1.4	2.1	1.1	1.9	1.4	2.1	1.1	1.8
$\overline{RDDG}$	1.6	2.3	1.2	2.0	1.5	2.2	1.1	1.8
$\overline{(DG)}$m	1.5	2.4	1.1	2.0	1.4	2.3	1.0	1.9
$\overline{(DGm)}$m	1.5	2.4	1.1	1.9	1.4	2.3	1.0	1.9
$\overline{(RDDG)}$m	1.7	2.5	1.3	2.1	1.7	2.4	1.2	2.0
X-ray–EM	0.25	0.24	0.25	0.24	0.36	0.41	0.38	0.43
X-ray–RM	0.46	0.50	0.49	0.52				

[a] The notation of the structures is the same as that in Tables III and IV. In the X-ray structure of the CPI–carboxypeptidase complex (Rees & Lipscomb, 1982) <Glu-1 is not visible in the electron density map and Gly-39 is cleaved from the rest of CPI.

erage structures are closer to their respective mean structures than any of the individual structures (Table III).

It is generally observed that the application of restrained molecular dynamics to a set of restrained energy minimized distance geometry structures results in a larger atomic rms distribution of the structures (Clore et al., 1986a, 1987a–c). In the case of CPI, this is indeed true for residues 2–39 considered as a whole. Examination of Figure 5 as well as the plot of atomic rms distribution as a function of residue number shown in Figure 6 clearly shows, however, that the atomic positions of residues 1–4 and 38–39 are poorly determined compared to that of residues 5–37. When this is taken into account, the effect of restrained dynamics on the atomic rms distribution of the structures can be dissected into two effects: namely, the atomic rms distribution of the poorly defined residues (1–4 and 38–39) is increased, whereas that of the well-defined region (residues 5–37) is decreased (Table III). Thus for residues 5–37 the backbone atomic rms distribution of the RDDG structures about the mean structure $\overline{RDDG}$ is 0.9 ± 0.2 Å compared to 1.1 ± 0.1 Å for the DG structures, whereas for residues 2–39 the respective values are 1.4 ± 0.1 and 1.2 ± 0.3 Å. Consequently, one can consider CPI to be made up of a core (residues 5–37), principally stabilized by the three disulfide bridges, and small N- (residues 1–4) and C- (residues 38–39) terminal tails. The core is additionally stabilized by four long-range backbone hydrogen bonds between Cys-12(N) and Arg-32(O), Ala-26(N) and Gly-35(O), Cys-34(N) and Lys-10(O), and Gly-35(N) and Ala-26(O), characterized by slowly exchanging NH protons (see Figure 4).

With the exception of a small helix extending from residues 16–21, there are no regular secondary structure elements and the structure principally consists of a series of turns connected by short stretches of polypeptide chain. Five turns can be clearly identified comprising residues 5–7, 7–10, 13–15, 22–24, and 28–31. Of these, only two, namely those from residues 7–10 and 28–31, are characterized by a CO(i)–NH(i + 4) backbone hydrogen bond, and only one, namely that from residues 7–10, is classical in nature, being of the type II variety. Considering these structural features, one might expect the structure to be rather poorly determined as, on the whole, the interproton distances that can be derived from NMR measurements are rather good at defining regular secondary structure elements but rather poor at defining irregular structures. In the case of CPI, however, the network of interproton distances obtained is sufficient to define the atomic positions of the backbone atoms of the core residues 5–37 extremely well (see Table III and Figures 5 and 6). Further, this is also extended to the ϕ and ψ backbone torsion angles (Figure 7). As expected, the side chains are somewhat less well-defined, particularly as most of these are surface residues. Indeed, only nine residues have an accessible surface area of <10 Å^2, of which five are cysteines (Cys-12, Cys-18, Cys-24, Cys-27, and Cys-34) and the other four have small side chains (Asp-5, Ala-21, Ala-26, and Gly-35). The positions of these buried side chains are well-defined (see Figure 6) as their location is dependent not only on the interproton distance restraints but also on packing requirements. In addition to these buried residues, quite a large number of surface residues (11 of 24) within the 5–37 segment are well-defined by the extensive NOE network with values of <1.3 Å for their atomic rms distributions about the mean $\overline{RDDG}$ structure (Figure 6).

Comparison with the X-ray Structure of CPI. In comparing the NOE-derived "solution" structures of CPI with that of the X-ray structure of CPI in its complex with carboxypeptidase (Rees & Lipscomb, 1980, 1982), one must bear two questions in mind. In particular, to what extent do the differences between the solution and X-ray structures arise from (a) genuine differences as reflected in differences between the experimental interproton distance restraints and a corresponding set of X-ray-derived distances, on the one hand, and from (b) inadequacies in the input data used to determine the solution structures as reflected by the limitations in the number, range, and accuracy of the experimental restraints, on the other.

That there are genuine differences between the solution and X-ray structures can be easily ascertained by the fact that the interproton distance deviations between the calculated and experimental distances are considerably large for the X-ray structure than for either the DGm and RDDG structures (Table IV). There are 37 deviations larger than 0.5 Å (Table VI), of which 35 involve at least one backbone proton.

The X-ray structure itself has a very high van der Waals energy (651 kcal/mol), which arises from five bad nonbonded contacts involving the following atom pairs: Ala-4(O)–Trp-22 ($C^{\epsilon 1}$H), 2.17 Å; Ile-7($C^{\delta 2}$H)–Gly-20(NH), 1.48 Å; Ala-21(O)–Phe-23(NH), 1.14 Å; Ala-31(C^{α}H)–Arg-32(HH11), 1.44 Å; and Arg-32(NH)–Arg-32(HH12), 1.17 Å. These can be completely relieved by 600 cycles of energy minimization (EM) to generate the structure X-ray–EM with minimal atomic rms shifts (0.24 Å for all atoms), which are within the upper limit

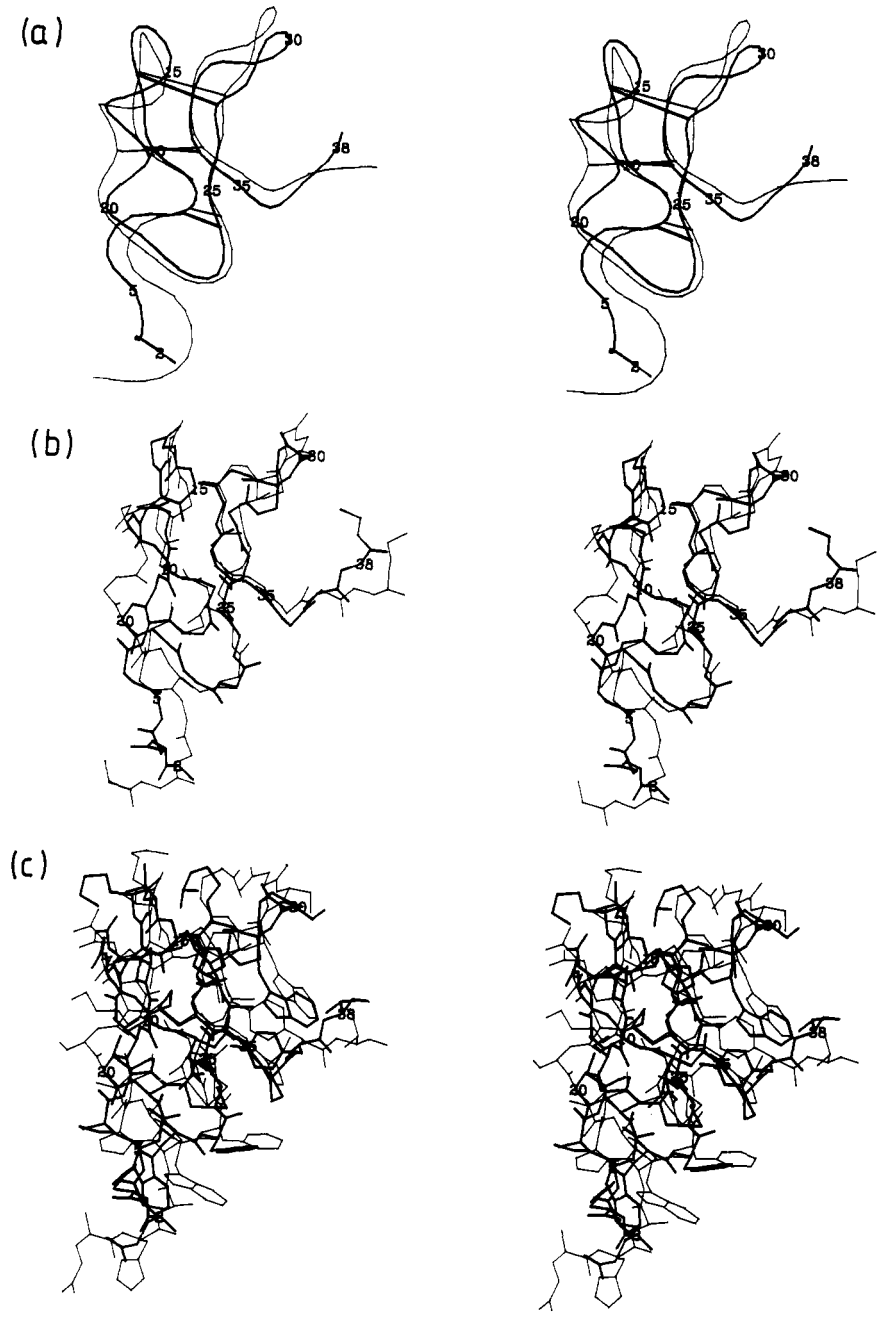

FIGURE 8: Best-fit superposition (residues 2–38) of the restrained energy minimized average structure ($\overline{\text{RDDG}}$)m (thick lines) with the X-ray structure of CPI (*thin lines*). (a) Smoothed backbone (C, C^{α}, N) atom representation with the location of the disulfide bridges indicated by lines joining the appropriate C^{α} atoms. (b) Backbone (C, C^{α}, N, O) atoms. (c) All atoms with the exception of hydrogens. Only residues 2–38 are shown, as <Glu-1 is not visible in the electron density map and Gly-39 is cleaved from the rest of the protein in the CPI–carboxypeptidase complex from which the X-ray structure of CPI is derived (Rees & Lipscomb, 1982).

of the error estimate of 0.3 Å in the X-ray coordinates (Rees & Lipscomb, 1982) as derived from a Luzzati (1952) plot. In addition to relieving the bad nonbonded contacts, energy minimization results in a small improvement in the interproton distance deviations (Table V). To see whether the experimental interproton distance restraints could potentially be accommodated by only small atomic rms shifts in the X-ray structure, we subjected the X-ray structure to 600 cycles of restrained energy minimization (RM) to generate the structure X-ray–RM. This resulted in much larger improvements in the interproton distance deviations, and the atomic rms shifts produced by this procedure, though naturally larger than those produced by energy minimization alone, were still small (~0.5 Å for all atoms; Table VII). Moreover, X-ray–RM is comparable to the DGm structures with respect to both interproton distance deviations (Table IV) and energies (Table V) and is significantly closer to the X-ray structure in terms of atomic positions than the DGm, $\overline{\text{DGm}}$, and ($\overline{\text{DGm}}$)m structures or the DG, $\overline{\text{DG}}$, and ($\overline{\text{DG}}$)m structures (Table VII). Nevertheless, a few interproton distance deviations larger than 0.5 Å still remain (Table VI). Further, the agreement between the experimental interproton distance restraints and the calculated distances is still much better for the RDDG and ($\overline{\text{RDDG}}$)m structures than for X-ray–RM (Table IV) and, in addition, the nonbonding energies of the RDDG and ($\overline{\text{RDDG}}$)m structures are significantly lower than those of either the X-ray–RM or X-ray–EM structures. Thus, some of the differences between the RDDG structures and the X-ray structure must be considered to reflect genuine differences.

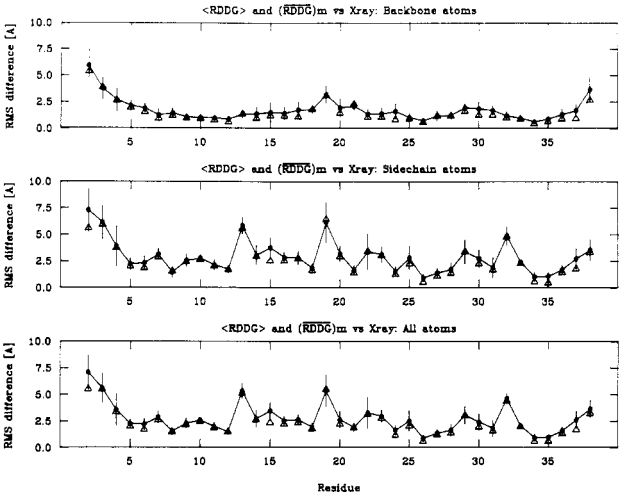

FIGURE 9: Atomic rms difference between the RDDG (●) and (RDDG)m (△) structures on the one hand and the X-ray structure on the other. The closed circles represent the average rms difference between the RDDG structures and the X-ray structure at each residue, and the bars represent the standard deviations in these values.

A best-fit superposition of the X-ray and (RDDG)m structure and a plot of atomic rms differences between the RDDG and (RDDG)m structures on the one hand and the X-ray structure on the other are shown in Figures 8 and 9, respectively. A summary of the atomic rms differences between the calculated structures and the X-ray structure is given in Table VII. As expected, the largest deviations occur at the N- (residues 1–4) and C- (residue 38) terminal ends (Figures 8 and 9). As the positions of these residues are poorly determined by the experimental data (cf. Figures 5 and 6), these differences cannot be regarded as significant. Rather, they simply reflect the paucity of experimental restraints within these two regions, most likely due to a larger degree of flexibility of these terminal arms. In this respect one should always bear in mind when comparing NMR solution and X-ray structures that there will in general be a correlation between the regions most poorly defined by the NMR data and those regions exhibiting the largest deviations from the X-ray structure. Indeed, any close coincidence of a poorly defined region with the corresponding region in the X-ray structure should be regarded as purely fortuitous. The close coincidence of the backbone atoms within the 5–37 residue segment is impressive with a backbone atomic rms difference of only 1.3 Å between (RDDG)m and the X-ray structure. Moreover, it is worth noting that the average restrained energy minimized structure (RDDG)m is closer to the X-ray structure than any of the individual RDDG structures with the exception of 1 of the 11 RDDG structures, which has a backbone atomic rms difference of only 1.1 Å with respect to the X-ray structure, and that the mean structure RDDG is slightly closer still. Further, the four long-range backbone hydrogen bonds stabilizing the 5–37 core of CPI as well as the type II turn (residues 7–10) discussed in the previous section are also present in the X-ray structure. Thus, the differences between the X-ray structure in the complex and the structure in free solution are clearly small. Nevertheless, there are probably two regions where the backbone atom differences are significant and are easily appreciated both from the best-fit superposition (Figure 8) as well as from the backbone atomic rms plot (Figure 9). The first region comprises the segment from residues 18–20, which are part of a helix extending from residues 16–21 in the solution structures, whereas in the X-ray structure this region is slightly distorted. The second region comprises the turn formed by residues 28–31, which has a slightly different orientation with respect to the rest of the protein in the solution and X-ray structures. This latter difference could easily be due to changes on complexation with carboxypeptidase as residues 27–30 constitute the secondary contact region between CPI and carboxypeptidase (Rees & Lipscomb, 1982). Interestingly, this difference is larger than the difference involving the residues of the primary contact region (residues 36–38), despite the fact that the conformation of residue 38 is poorly determined by the experimental restraints.

As to the side chains, a critical assessment of the differences cannot be carried out because their positions are clearly less well determined by the experimental restraints than those of the backbone atoms. Indeed, the largest differences involve those residues with the poorest determined atomic positions (compare the side chain atom atomic rms plot in Figures 6 and 9). The differences in the conformations of Pro-36 and Tyr-37, however, which constitute part of the primary contact region (residues 36–38), are relatively small.

CONCLUSIONS

In this paper we have presented the determination of the three-dimensional structure of CPI in solution on the basis of NOE derived interproton distance restraints. With the exception of small N- and C-terminal arms, the structure is well-defined by the experimental restraints despite the fact that it possesses only one small region (residues 16–21) of regular secondary structure. The comparison of the NMR and X-ray structures of CPI represents one of the first detailed comparisons of a protein structure determined by two independent methods, one in solution and the other in the solid state. The NMR structure is similar to the previously determined X-ray structure, although some small but distinct differences are apparent. Whether these differences arise from crystal-packing forces or from conformational changes occurring upon complexation with carboxypeptidase in the crystal cannot be ascertained. Nevertheless, the results clearly show that the structure of uncomplexed CPI in solution is very similar to that of the complexed form in the crystal, confirming the view that a number of protease inhibitors are small, rigid proteins which do not undergo substantial conformational changes upon binding to their target molecule.

ACKNOWLEDGMENTS

We thank Gregory Pearce for excellent technical assistance in preparing CPI.

SUPPLEMENTARY MATERIAL AVAILABLE

One table giving the complete list of NOE interproton distance restraints used in the computation of the three-dimensional structure of CPI (8 pages). Ordering information is given on any current masthead page.

Registry No. Carboxypeptidase inhibitor, 11075-17-5.

REFERENCES

Altman, R. B., & Jardetzky, O. (1986) *J. Biochem. (Tokyo) 100*, 1403–1423.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–366.

Bax, A., Sklenar, V., Clore, G. M., & Gronenborn, A. M. (1987) *J. Am. Chem. Soc.* (in press).

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 153*, 321–345.
Bodenhausen, G., Vold, R. L., & Vold, R. R. (1980) *J. Magn. Reson. 37*, 93–106.
Braun, W., & Go, N. (1985) *J. Mol. Biol. 186*, 611–626.
Braun, W., Wider, G., Lee, K. H., & Wüthrich, K. (1983) *J. Mol. Biol. 169*, 921–948.
Braun, W., Wagner, G., Wörgötter, E., Vasch, M., Kägi, J. H. R., & Wüthrich, K. (1986) *J. Mol. Biol. 187*, 125–129.
Braunschweiler, L. R., & Ernst, R. R. (1985) *J. Magn. Reson. 53*, 521–528.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swamirathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.
Brünger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3801–3805.
Clore, G. M., & Gronenborn, A. M. (1987) *Protein Eng. 1*, 275–288.
Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1985) *J. Mol. Biol. 186*, 435–455.
Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986a) *EMBO J. 5*, 2729–2735.
Clore, G. M., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986b) *J. Mol. Biol. 191*, 523–551.
Clore, G. M., Sukumaran, D. K., Nilges, M., & Gronenborn, A. M. (1987a) *Biochemistry 26*, 1732–1745.
Clore, G. M., Sukumaran, D. K., Nilges, M., Zarbock, J., & Gronenborn, A. M. (1987b) *EMBO J. 6*, 529–537.
Clore, G. M., Gronenborn, A. M., Nilges, M., Sukumaran, D. K., & Zarbock, J. (1987c) *EMBO J. 6*, 1833–1842.
Clore, G. M., Nilges, M., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1987d) *FEBS Lett. 213*, 269–277.
Clore, G. M., Gronenborn, A. M., Kjaer, M., & Poulsen, F. (1987e) *Protein Eng. 1*, 305–311.
Clore, G. M., Gronenborn, A. M., James, M. N., Kjaer, M., McPhalen, C., & Poulsen, F. (1987f) *Protein Eng. 1*, 313–318.
Crippen, G. M., & Havel, T. F. (1975) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A34*, 282–284.
Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2821–2822.
Feldman, R. J., & Brooks, B. R. (1986) *Tools for Each Age: Understanding Protein Architecture Through Simulated Unfolding*, Division of Computer Research and Technology, National Institutes of Health, Bethesda, MD.
Hass, G. M., & Ryan, C. A. (1981) *Methods Enzymol. 80*, 778–791.
Hass, G. M., Nan, H., Bitmann, K., Grahn, D. T., Ericsson, L. M., & Neurath, H. (1975) *Biochemistry 14*, 1334–1342.
Havel, T. F. (1986) *QCPE*, No. 507.
Havel, T. F., & Wüthrich, K. (1984) *Bull. Math. Biol. 45*, 673–698.
Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281–294.
Havel, T. F., Kuntz, I. D., & Crippen, G. M. (1983) *Bull. Math. Biol. 45*, 665–720.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.
Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179–182.
Kline, A. D., Braun, W., & Wüthrich, K. (1986) *J. Mol. Biol. 183*, 503–507.
Leary, T. R., Grahn, D. T., Neurath, H., & Hess, G. M. (1979) *Biochemistry 18*, 2252–2256.
Luzzati, V. (1952) *Acta Crystallogr. 5*, 802–810.
Macura, C., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259–281.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Pearce, G., & Ryan, C. A. (1983) *Anal. Biochem. 130*, 223–225.
Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310–7311.
Redfield, A. G., & Kuntz, S. D. (1975) *J. Magn. Reson. 19*, 250–254.
Rees, D. C., & Lipscomb, W. N. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4633–4637.
Rees, D. C., & Lipscomb, W. N. (1982) *J. Mol. Biol. 160*, 475–4998.
Ryan, C. A., Hass, G. M., & Kuhn, R. W. (1974) *J. Biol. Chem. 249*, 5495–5499.
Sippl, M. J., & Scheraga, H. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2283–2287.
Sklenar, V., & Bax, A. (1987) *J. Magn. Reson.* (in press).
Wagner, G., Braun, W., Havel, T. F., Schauman, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611–640.
Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295–315.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol. Biol. 155*, 311–319.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949–961.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 415–740.
Zuiderweg, E. R. P., Billeter, M., Boelens, R., Scheek, R. M., Wüthrich, K., & Kaptein, R. (1984) *FEBS Lett. 174*, 243–247.