Analysis of Side-Chain Conformational Distributions in Neutrophil Peptide-5 NMR Structures

DOROTHEA KOMINOS,[1] DONNA A. BASSOLINO,[1] RONALD M. LEVY,[1]* and ARTHUR PARDI[2]*

[1]Department of Chemistry, Rutgers University, New Brunswick, New Jersey 08903, and [2]Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, Colorado 80309

SYNOPSIS

The side-chain conformations have been analyzed in the antimicrobial peptide, Neutrophil Peptide-5 (NP-5), whose structure was independently generated from nmr-derived distance constraints using a distance geometry algorithm. The side-chain and peptide dihedral angle distributions in the nmr structures were compared with those constructed from a data base of high-resolution protein crystal structures. The side-chain conformational preferences for NP-5 in solution are significantly different from those observed in the crystal structure data base. These results indicate that the side-chain conformations are quite disordered for many of the residues of NP-5. The absence of a correlation between the width of the conformational distribution and surface accessibility suggests that the disorder may be due to limitations in the structural information extracted from the nmr data rather than to molecular motion. However, it is also observed that the degree of conformational disorder is only weakly correlated with the number of nuclear Overhauser enhancements to a given side chain. Possible reasons for this are discussed. Molecular mechanics refinement of these structures did not significantly change the side-chain populations. Anomolously wide distributions are observed for rotations about the peptide bonds and the disulfide bonds in the NP-5 distance geometry structures, which are improved by the refinement. The very high degree of order observed for the central dihedral angle of the disulfide bond in the high-resolution crystal data base suggests that the rotation about this bond in proteins is determined by the local potential.

INTRODUCTION

The determination of the three-dimensional structure of a protein at high resolution is required in order to understand the molecular basis for its mechanism of action. Until recently, the only method capable of providing such high-resolution structural information was x-ray crystallography. The rapid advances that have occurred in nmr technology have culminated in the reports of the solution structures of a number of proteins. There are several methods capable of converting experimental internuclear distance information obtained from nuclear Overhauser enhancement (NOE) data into three-dimensional structures, including the metric matrix distance geometry method,[1-4] restrained molecular dynamics,[5-8] energy minimization in torsion angle space,[9-12] and an internal coordinate Monte Carlo algorithm.[13,14] All of these algorithms have a statistical component to them where many structures that approximately satisfy the distance and stereochemical restraints are generated and compared. For the most part the analysis has focused on identifying and comparing the overall folding patterns of the proteins, and analysis of side-chain conformations has been given less attention.

One of the important questions that can potentially be addressed with the new nmr technology is, How does the structure of a protein in solution differ from that in the crystalline state? The most dramatic differences would involve rearrangement of the protein backbone. For the few examples studied for which both crystal and solution structures are

© 1990 John Wiley & Sons, Inc.
CCC 0006-3525/90/141807-16 $04.00
Biopolymers, Vol. 29, 1807–1822 (1990)
* To whom correspondence should be addressed.

available,[10,15-17] large changes in the structure have not generally been observed, although Billeter et al.[17] have observed rather subtle differences in side-chain conformation ascribed to intermolecular interactions in the crystal. Perhaps (in retrospect) this is not surprising given that protein crystals contain 40–70% water and thus the protein can be heavily solvated in the crystal. However, at the least, very different patterns are expected for side-chain conformations in solution as compared with the crystal environment. The physical basis for this would include the expected increased mobility of surface-exposed side chains in solution, the absence of crystal packing forces in solution, and the loss of specific intermolecular interactions present in the crystal but not in solution.

Analysis of side-chain conformations in x-ray crystal structures has received a great deal of attention. It has been shown that side-chain conformations of amino acid residues are found in discrete regions of conformational space, as expected from an analysis of the structural chemistry of model compounds.[18,19] Additionally, side-chain conformations more closely resemble predicted results when a group of highly resolved peptide[20] or protein[21,22] crystal structures are considered in the analysis. Detailed studies have correlated the side-chain conformations with the secondary structure of the backbone.[23,24]

We have been studying the solution structure of Neutrophil Peptide-5 (NP-5),[13,14,25,26] a member of a family of peptides known as defensins.[27,28] The defensins constitute a class of 10 homologous proteins whose sizes range from 29 to 34 residues, with 8 residues common to the entire family. Defensins have been shown to exhibit a wide range of antimicrobial action.[29-31] In a previous article we analyzed structural features concerning the overall peptide fold.[14] In this paper we focus on the side-chain conformations observed in the nmr structures. The distribution of side-chain conformations for each of the amino acids in NP-5 has been constructed from eight independently generated NP-5 structures using a distance geometry algorithm. NP-5 has not yet been crystallized, although the crystal structure of a homologous defensin (HNP-3) has very recently been solved (D. Eisenberg, personal communication). The side-chain conformational preferences observed in the NP-5 distance geometry structures before and after energy minimization are compared with the corresponding distributions calculated from a data base constructed from 24 high-resolution protein crystal structures. As has been observed previously,[21] the side-chain distributions in crystal structures are sharply peaked about a small number of rotomeric states. We expected to find that buried side chains in NP-5 would have conformations close to that observed in the crystal for the corresponding amino acid, whereas surface side chains would either have a tendency to populate multiple rotomeric states or would exhibit a more continuous distribution of states. As shown in the following sections, the side-chain conformations as calculated from the eight NP-5 nmr structures are very different from the conformational preferences of these side chains found in high-resolution crystal structures of proteins, and possible reasons for these differences will be discussed.

METHODOLOGY

Generation of the Solution Structures of NP-5

The eight solution structures of NP-5 were generated from a set of distance constraints obtained from two-dimensional (2D) NOE data.[25,26] The set of constraints included 107 distances and 3 disulfide bonds between residues 3–31, 5–20, and 10–30, respectively. The 107 distance constraints consisted of 50 backbone–backbone constraints, 22 backbone–side chain constraints and 35 side chain–side chain constraints. Distances (d) estimated from the 2D NOE spectroscopy were classified into three groups: short ($d < 2.8$ Å), medium ($2.8 \leq d \leq 3.8$ Å) and long ($d > 3.8$ Å).[26] Structures were generated from the set of distance constraints using the distance geometry method[2,32] as implemented in the computer program DSPACE (Hare Research Inc.).

High-Resolution Protein Data Base

A group of 22 protein structures, obtained from the July 1987 release of the Brookhaven Protein Data Bank,[33] were selected for this study on the basis of their high resolution (1.8 Å or higher resolution). The 22 proteins in this study possessed R factors ranging from 11.4 to 24.0% and are listed in Table I. This group of structures is similar to the group used by Ponder and Richards[22] in their analysis of side-chain packing in protein crystal structures. These structures represent the four classes of globular proteins,[34] and include multiple-domain and multiple-subunit proteins. The proteins range in size from 36 to 323 residues. The fractional occurrence of each amino acid type in the high-resolution crystal structure data base is shown in Table II.

Gly was by far the most abundant residue in the

Table I Crystallographic Structures Used in the High-Resolution Data Base

Protein Name	Brookhaven Code	No. of Residues	Resolution	*R* factor
Basic trypsin inhibitor	5PTI	52	1.00	20.0
Rubredoxin	5RXN	54	1.20	11.5
Avian pancreatic hormone	1PPT	36	1.37	21.0
Neurotoxin B	1NXB	62	1.38	24.0
Carboxypeptidase A	5CPA	307	1.50	19.0
Crambin	1CRN	46	1.50	11.4
Proteinase A	2SGA	181	1.50	12.6
Tuna cytochrome C	4CYT	103	1.50	17.3
Lysozyme	1LZ1	130	1.50	17.7
Insulin	1INS	102	1.50	17.9
Ovomucoid (3rd domain)	2OVO	56	1.50	19.9
Bence–Jones immunoglobulin	2RHE	114	1.60	14.9
Myoglobin	1MBO	153	1.60	15.3
Plastocyanin	1PCY	99	1.60	17.0
Thermolysin	3TLN	316	1.60	21.3
γ-Crystallin	1GCR	184	1.60	23.0
Papain	9PAP	212	1.65	16.1
α-Lytic protease	2ALP	198	1.70	13.1
Phospholipase	1BP2	123	1.70	17.1
Trypsin	3PTB	223	1.70	18.2
Penicillopepsin	2APP	323	1.80	12.6
Scorpion neurotoxin	1SN3	65	1.80	16.0

data base. Of the residues considered in the analysis of the NP-5 side chains, however, Ser had the largest frequency, occurring 9.2% of the time. Because of the special importance of disulfide bonds in limiting the allowed peptide conformations, we have included the torsional angle about the sulfur–sulfur bond in our conformational analysis. Cysteine, which has a large fractional occurrence in NP-5 (18%), is reasonably well represented in the crystal structure data base (116 residues, 3.7%). The lowest frequency belonged to Met (0.9%) followed by Trp (1.4%) and His (1.9%). Due to the low frequency of these residues in our high-resolution data base, statistics for these residues may not be as accurate as for those

Table II Frequency of Occurrence of Residues in the High-Resolution Data Base[a]

Residue	No. in Data Base	% Total	Residue	No. in Data Base	% Total
Ala	228	7.3	Leu	203	6.5
Arg	114	3.6	Lys	154	4.9
Asn	179	5.7	Met	29	.9
Asp	155	4.9	Phe	116	3.7
Cys	116	3.7	Pro	122	3.9
Gln	136	4.3	Ser	288	9.2
Glu	123	3.9	Thr	202	6.4
Gly	336	10.7	Trp	44	1.4
His	60	1.9	Tyr	176	5.6
Ile	154	4.9	Val	201	6.4

[a] Total: 3136.

of other residues, but as seen by the list of amino acid distributions in NP-5 given in Table III, of these three residues only His is found in NP-5.

Analysis of Geometric Properties

The geometric analysis was carried out using the program IMPACT (Integrated Modeling Program Using Applied Chemical Theory)[13] coupled to a commercial Relational Data Base Management program, INGRES (see the appendix). The dihedral angle conformations were classified according to rotomer states. For a torsion about a bond between two tetrahedral centers, the rotomeric states—denoted g^+, t, and g^-—correspond to rotations of +60°, 180°, and −60°, respectively. For a torsion about a bond involving an sp^2 carbon the rotomer states are +90°, 0°, and −90°. A dihedral angle was classified according to one of the rotomer states if the calculated torsion angle was within ± 30° of the idealized value; side-chain dihedral angles that did not fall within any of the defined rotomer populations were classified as "other." In the data base of side-chain conformations constructed from the high-resolution crystal structures, very few dihedral angles (12%) lie within the "other" category, with the exception of the χ_2 dihedral angle in Asp and Asn, and the χ_3 dihedral angle in Glu and Gln. Two measures of disorder were employed in the analysis of the rotational distribution of a side-chain dihedral angle presented in the following section. The first measure is the percentage of conformers that lie in the "other" category, the second is the weighted average variance within the rotamer states. The weighted average variance is obtained by calculating the variance within each rotamer state separately and averaging the variances with weights according to the fraction of the total number of conformers in each rotamer. For the calculation of the variance, a dihedral angle was classified according to one of the rotomer states if the calculated torsion angle was within ± 60° of the idealized value.

Side-chain conformations in the solution structures of NP-5 were also correlated with surface exposure. The accessible surface area calculations were carried out using an algorithm based on the one developed by Lee and Richards,[35] using an all-atom representation of the protein that included all hydrogens explicitly. The accessibilities for each residue were normalized by a "standard" accessible surface area for each amino acid residue defined as the surface area in the tripeptide Gly-X-Gly, with X in an extended conformation.[36,37]

Molecular Mechanics Refinement of NP-5 Structures

Molecular mechanics refinement of the NP-5 structures was carried out with IMPACT using the all-atom force field parameters of Weiner et al.[38] The conjugate gradient method was used for energy minimization.[39] The potential energy function had the following form:

$$E = \sum_{\text{all bonds}} k_b(b - b_0)^2 + \sum_{\text{all angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{all dihedrals}} k_\phi[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}}\right.$$

Table III Frequency of Occurrence of Residues in NP-5[a]

Residue	No. of Occurrences	% Total	Sequence Number
Cys	6	18.2	(3, 5, 10, 20, 30, 31)
Arg	5	15.2	(6, 15, 26, 32, 33)
Gly	5	15.2	(7, 11, 13, 18, 24)
Ser	3	9.1	(12, 17, 19)
Thr	3	9.1	(4, 21, 28)
Leu	2	6.1	(9, 29)
Phe	2	6.1	(2, 8)
Val	2	6.1	(1, 25)
Ala	1	3.0	(16)
Asn	1	3.0	(23)
Glu	1	3.0	(14)
His	1	3.0	(27)
Ile	1	3.0	(22)

[a] Total number of residues: 33.

$$-\frac{B_{ij}}{R_{ij}^6}+\frac{q_iq_j}{\epsilon R_{ij}}\Bigg]+\sum_{\text{H bonds}}\left[\frac{C_{ij}}{R_{ij}^{12}}-\frac{D_{ij}}{R_{ij}^{10}}\right]+\sum_{\text{no. constraints}}k_d(r_{\text{target}}-r_{\text{calc}})^2$$

The first five terms account for bond stretching, bond angle bending, hindered rotation about a single bond, van der Waals and electrostatic interactions, and the repulsive portion of the hydrogen-bonding potential. The attractive portion of the hydrogen-bond potential is implicit in the electrostatic terms. A distance-dependent dielectric constant was used. The nonbonded list was updated every 10 steps and the nonbonded cutoff was set to 7.5 Å. The final term is a constraint term used to ensure that NOE distance constraints between atom pairs are maintained during the energy refinement.

RESULTS

Side-Chain Conformational Preferences in the High-Resolution Data Base

Table IV lists the distributions of each side-chain dihedral angle from the high-resolution data base for the 17 different amino acids (excluding Ala, Gly, and Pro). The two extra side-chain dihedral angles that are created upon forming a disulfide bridge in Cys residues are presented in Table V. In general, the preferred conformation for the χ_1 dihedral angle in residues with unbranched C_β atoms is the g^- conformation, followed by the t conformation, and finally the g^+ conformation (see Figure 1a). The g^+ conformation for χ_1 is expected to be least favorable due to steric crowding with the backbone. For branched C_β atoms, Val prefers χ_1 in the t conformation while Ile prefers the standard g^- conformation. For the two side chains containing a hydroxyl group (Thr and Ser), the distribution of χ_1 is binodal with 46% g^+ conformation and 44% g^- conformation for Thr, but shows a marked preference for the g^+ conformation in Ser. The Ser χ_1 distribution was shown to have little conformational preference for the g^+ conformation in early work,[18,19] but more recent studies using high-resolution structures indicate a preference for g^+.[21] This unusual conformational preference for Ser χ_1 may be a result of hydrogen bonding of the hydroxyl group to the backbone. The standard deviations in the χ_1 angles for all residues range from 6.0° to 12.7°, and tend to be small except for very long and hydrophilic side chains (Lys and Arg). Another measure of how well χ_1 falls into the expected regions of conformational space is indicated by the low percentage of residues in the "other" category, which ranges from 0 to 14%. The two largest values of "other" conformations were for the long side-chain residues Arg and Lys.

The χ_2 dihedral angle distribution is less regular than χ_1. The percentage of "other" residues has increased compared to χ_1, and ranges between 8.8 to greater than 60.0% of the conformers. Also, the standard deviations of the distributions are larger, with values between 6.5 and 16.9%. In general, χ_2 is less well defined in the longer side chains. For the aromatic residues, χ_2 is distributed in a narrow range about ± 90° (see Figure 1b), with the indole group of the Trp and the imidazole group of His considered symmetric to help simplify this analysis. In our analysis, Asp and Asn do not form simple χ_2 distributions; this result is in agreement with that of Bhat.[19] The χ_2 angle for these side chains were placed in two categories, *cis* and *trans* according to the two largest rotamer populations, but it is evident this is not an accurate description of their conformational distributions (see Figure 1c). A similar result was observed for χ_3 of Glu and Gln, and so they were treated in the same manner. The side-chain angles further from the backbone are generally more disordered.

Disulfide links between cysteine residues provide a strong conformational constraint for small cystine-rich proteins like NP-5. All the cysteine residues in NP-5 are disulfide linked and therefore an analysis of the torsion angles created upon the formation of a disulfide bridge is important. Surprisingly, to our knowledge there has not been a statistical analysis of the disulfide geometries in proteins, although Richardson[40] observed that the Cys residues have a preference for dihedral angle values of ± 90° for both the torsion about C_β-S_γ and the torsion about S_γ-S_γ. The results of our statistical analysis of the conformational preferences of the torsion angles involving the disulfide bond are presented in Table V. While the torsion angle about C_β-S_γ is not narrowly distributed (38% "other" conformations), the S_γ-S_γ torsion angle is very sharply peaked about ± 90°, and none of the residues lie outside these two rotamer states.

Side-Chain Conformational Preferences for the NP-5 Solution Structures

The side-chain dihedral angle distributions for the NP-5 structures are quite different from the data-base structures (see Table VI). The same categories that were used for a residue in the data base were applied in the analysis of the NP-5 structures. For many of the residues, no clear definition of conformational preference is evident. The percentage of "other" residues in the χ_1 torsion angle ranges from 0 to 100%. In general the percentage of "other" con-

Table IV Side-Chain Distributions in the High-Resolution Data Base

Residue Name	No. in Sample		χ_1		χ_2		χ_3		χ_4	
Arg	114	g^+	58.2 ± 8.4	8.8[a]	67.6 ± 9.3	6.1[a]	61.9 ± 10.8	16.7[a]		
		t	−178.6 ± 12.7	21.9	179.8 ± 13.1	65.8	179.2 ± 15.0	37.7	−178.5 ± 15.5	33.3[b]
		g^-	−65.3 ± 9.8	56.1	−70.7 ± 6.5	7.9	−63.5 ± 14.4	20.2		
		90							90.2 ± 14.1	37.7
		Other[c]		13.2		20.2		25.4		29.0
Asn	179	g^+	64.1 ± 8.1	15.1						
		t	174.7 ± 10.1	28.5	177.5 ± 16.8	5.0				
		g^-	−67.8 ± 10.3	44.7						
		0			−2.0 ± 16.9	34.6				
		Other		7.9		60.4				
Asp	155	g^+	63.4 ± 8.4	20.0						
		t	−171.4 ± 8.3	25.8	170.3 ± 12.7	25.8				
		g^-	−68.1 ± 10.4	49.0						
		0			−3.9 ± 15.6	38.7				
		Other		5.2		35.5				
Cys	116	g^+	64.0 ± 11.5	14.7						
		t	−178.7 ± 10.4	26.7						
		g^-	64.1 ± 11.0	57.8						
		Other		0.9						
Gln	136	g^+	60.7 ± 11.6	5.8	71.4 ± 10.4	11.0				
		t	−178.3 ± 10.8	29.4	−177.8 ± 13.0	63.2	179.5 ± 18.4	12.5		
		g^-	−63.3 ± 12.1	53.7	−63.6 ± 10.8	14.7				
		0					3.7 ± 18.4	24.3		
		Other		11.0		11.0		63.2		
Glu	123	g^+	64.4 ± 12.6	12.2	75.7 ± 8.6	14.6				
		t	−176.2 ± 11.0	26.0	−179.1 ± 12.5	58.5	−177.5 ± 14.8	20.3		
		g^-	−63.6 ± 12.1	48.8	−62.2 ± 13.5	14.6				
		0					−2.5 ± 15.9	32.5		
		Other		13.0		12.2		47.2		
His	60	g^+	63.9 ± 8.4	10.0						
		t	−176.0 ± 8.9	33.3						
		g^-	−65.6 ± 10.8	53.3						
		90			84.6 ± 15.7	75.0				
		Other		3.3		25.0				
Ile	154	g^+	61.5 ± 4.1	17.5	66.1 ± 16.5	5.2				
		t	−169.4 ± 11.7	13.6	169.4 ± 10.4	70.2				
		g^-	−60.3 ± 9.3	68.2	−61.6 ± 12.4	15.6				
		Other		0.6		9.1				
Leu	203	g^+	55.8 ± 6.0	1.5	62.3 ± 10.9	29.6				
		t	−175.5 ± 9.5	27.6	175.8 ± 10.6	59.6				
		g^-	−64.9 ± 9.4	60.1	−73.1 ± 11.0	2.0				
		Other		10.8		8.8				
Lys	147	g^+	66.1 ± 13.3	4.8	68.4 ± 14.0	10.2	67.8 ± 14.9	7.5	58.8 ± 13.2	9.5
		t	−177.5 ± 11.8	34.7	−179.1 ± 11.6	53.7	179.7 ± 11.4	49.0	179.2 ± 12.9	52.4
		g^-	−64.6 ± 12.3	46.9	−60.2 ± 14.8	10.9	−66.0 ± 18.9	9.5	−62.3 ± 15.2	12.9
		Other		13.6		25.2		34.0		25.2
Met	29	g^+	62.3 ± 8.1	6.9	79.9 ± 9.8	10.3	68.8 ± 5.9	17.2		
		t	−177.4 ± 11.8	27.6	176.9 ± 10.3	58.6	−179.0 ± 6.9	31.0		
		g^-	−60.0 ± 12.4	62.1	−68.4 ± 16.9	20.7	−61.0 ± 11.9	20.7		
		Other		3.4		10.3		31.0		
Phe	116	g^+	62.1 ± 11.3	20.7						
		t	−179.0 ± 7.8	25.0						
		g^-	−67.4 ± 10.9	52.6						
		90			89.6 ± 11.8	84.5				
		Other		1.7		15.5				
Ser	288	g^+	65.0 ± 11.6	44.1						
		t	−179.2 ± 10.4	19.1						
		g^-	−64.6 ± 10.9	27.8						
		Other		9.0						

Table IV (*Continued from the previous page.*)

Residue Name	No. in Sample	χ_1	χ_2		χ_3		χ_3
Thr	202	g^+	63.4 ± 8.8	46.0			
		t	−169.9 ± 8.5	5.9			
		g^-	−58.6 ± 8.8	43.6			
		Other		4.5			
Trp	44	g^+	65.2 ± 7.9	22.7			
		t	−178.5 ± 9.2	34.1			
		g^-	−70.0 ± 7.2	43.2			
		90			91.7 ± 13.5	97.7	
		Other		0.0		2.3	
Tyr	176	g^+	62.0 ± 10.3	14.2			
		t	−179.1 ± 9.8	33.5			
		g^-	−64.1 ± 10.4	51.1			
		90			89.5 ± 13.8	83.5	
		Other		1.2		16.5	
Val	201	g^+	64.7 ± 7.7	5.5			
		t	174.1 ± 7.6	65.2			
		g^-	−64.5 ± 7.3	21.4			
		Other		8.0			

[a] Percent of conformers in range ($g^+ \pm 30$, $t \pm 30$, $g^- \pm 30$, 90 ± 30, 0 ± 30, or other).
[b] Placed in this distribution as expected by structural chemistry.[19]
[c] Conformers in the "other" category lie outside the defined rotomer populations, see text.

formations within a residue is significantly larger than the corresponding residue in the data base.

There are some residues that show unusual behavior. The Arg residues in NP-5, for instance, appear to be more ordered in χ_4 than in χ_1 (see Figure 2). One generally expects Arg residues to be exposed in solution and therefore χ_4 is expected to be much more disordered than χ_1. Another residue of interest is Cys where the χ_1 torsion angle for Cys is much more disordered than in the data base. Moreover, the Cys side-chain dihedral angle conformations, which are formed upon cross-linking, are very different from the corresponding distributions calculated from the crystallographic data base (see Figure 3). In the NP-5 structures, the dihedral angle about C_β-S_γ has 62.5% of its residues in conformations other than the 90° conformations, as compared with only 38% of the residues in the crystal data base. Even more striking is the comparison for the dihedral angle about the S_γ-S_γ bond. Whereas in the crystal structures all the torsions about the S_γ-S_γ bond correspond to ± 90 rotomeric states with no structures falling in the "other" category, more than half of the disulfide torsions in the NP-5 structures lie in the "other" category (see Table V). Anomolous conformations are also observed for the histidine and leucine side chains. The His χ_1 distribution calculated from the NP-5 distance geometry structures has no residues occurring in the expected g^-, t, and g^+ states, and the percentage of "other" for χ_1 is quite large (75%). The χ_1 distribution for Leu is dominated by the g^+ and t regions in the solution structures of NP-5, whereas the χ_1 distribution for Leu in the crystal data base is predominantly g^-. The Leu χ_2 side-chain preferences also deviate from the crystal data base and there remains a large per-

Table V Side-Chain Angles About Disulfide Bonds

Dihedral Angle	Conformation	High-Resolution Data Base	Distance Geometry Structures
C_α-C_β-S_g-S'_γ	90^+	80.7 ± 13.6 (21.0%)	89.8 ± 14.1 (22.9%)
	90^-	−83.6 ± 14.8 (42.0%)	−96.0 ± 15.1 (14.6%)
Other		38.0%	62.5%
C_β-S_γ-S'_γ-C'_β	90^+	94.3 ± 10.1 (58.0%)	80.4 ± 14.6 (16.6%)
	90^-	−83.5 ± 7.1 (42.0%)	−113.0 ± 5.6 (29.2%)
Other		0.0%	54.2%

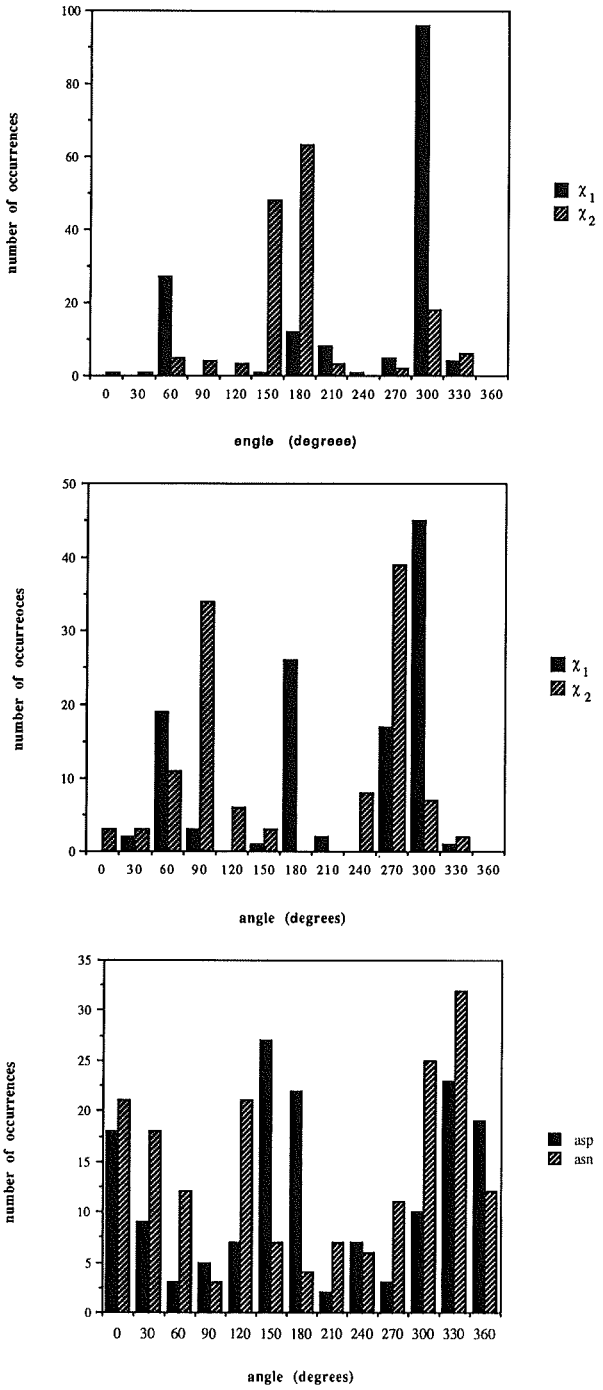

Figure 1 The χ_1 and χ_2 side-chain distributions for typical amino acids in the high-resolution data base. (a) Ile, (b) Phe, and (c) χ_2 distribution for Asp and Asn in the high-resolution data base.

centage of "other" residues in the NP-5 solution structures. The remaining residues occurring in NP-5 tend to behave more like the high-resolution data base. In the NP-5 solution structure the standard deviations for the side-chain dihedral angle distributions are also high relative to the data base.

Correlations of Side-Chain Conformations with Surface Accessibility and Number of Distance Constraints

Many of the side chains in NP-5 are observed to have different geometries in the eight distance geometry structures generated from the NOE data. This disorder may arise from motion of the side chains in solution (dynamic disorder) or from a distribution of protein conformational states (static disorder) probed in the nmr experiment. Alternatively, the disorder may reflect the fact that the number of constraints is insufficient to uniquely position some of the side chains in NP-5. In an effort to distinguish true side-chain disordering (both static and dynamic) from disorder due to insufficient structural information, we analyzed the correlations between side chain disorder, surface accessibility and number of NOE constraints. We have used the standard deviation of the conformational distribution within the defined rotameric states for the side-chain torsion as one measure of the disorder for that internal coordinate. Intuitively, we expected that the conformations of the surface-exposed side chains would be most variable in the different NP-5 solution structures.

In Table VII, the average accessibility for each residue is reported for NP-5. We classify qualitatively, side chains with less than 20% accessibility as buried, those with accessibilities between 20 and 65% as partially exposed and those with accessibilities greater than 65% as exposed. According to this classification scheme, seven side chains (21%) are buried, sixteen side chains (48%) are partially exposed and 10 side chains (30%) are exposed. The pattern of accessibility of side chains by type is generally consistent with previous observations.[41] The least exposed residues are the cysteines. Cys-20 is completely buried (0.1% accessibility). The two cysteine residues closest to the C-terminal of NP-5, Cys-30 and Cys-31, also have a low accessibility. Other buried residues include His-27, Ala-16, and Ile-22. Ile-22 was shown to have a very rigid side-chain structure from the side-chain analysis. Four of the five Arg residues are completely exposed and Arg-6 has 55% of its area accessible to solvent. One puzzling result of this study is that Phe, which is expected to be buried in a hydrophobic core, is exposed in all the structures. This result may reflect the fact that the distance constraints to the ring protons on Phe had very large upper bounds in the

Table VI Side-Chain Distributions in the Eight Distance Geometry Structures of NP-5

Residue Name	No. in Sample		χ_1		χ_2		χ_3		χ_4	
Arg	40	g^+	60.6 ± 19.8	15.0	82.5	2.5	72.8 ± 6.1	5.0		
		t	175.4 ± 14.8	10.0	175.0 ± 15.8	25.0	174.7 ± 14.7	15.0	−176.3 ± 16.9	25.0
		g^-	−56.9 ± 21.6	35.0	−72.2 ± 11.2	27.5	−66.2 ± 18.4	32.5		
		90							85.9 ± 10.4	47.5
		Other		40.0		45.0		47.5		27.5
Asn	8	g^+	58.6 ± 14.1	37.5						
		t	180.0 ± 10.9	25.0	−158.3 ± 6.9	25.0				
		g^-	−51.6 ± 18.2	25.0						
		0			0.0	0.0				
		Other		12.5		75.0				
Cys	48	g^+	53.1 ± 13.9	35.4						
		t	−171.6 ± 2.3	10.4						
		g^-	−57.4 ± 18.0	18.8						
		Other		35.4						
Glu	8	g^+	68.7	12.5	0.0	0.0				
		t	0.0	0.0	0.0	0.0	−169.5 ± 10.1	25.0		
		g^-	0.0	0.0	−60.1	12.5				
		0					0	0.0		
		Other		87.5		87.5		75.0		
His	8	g^+	0.0	0.0						
		t	0.0	0.0						
		g^-	0.0	0.0						
		90			99.5 ± 3.3	25.0				
		Other		100.0		75.0				
Ile	8	g^+	0.0		87.3	12.5				
		t	0.0		168.3 ± 11.7	25.0				
		g^-	−67.1 ± 11.7	100.0	0.0					
		Other		0.0		37.5				
Leu	16	g^+	63.6 ± 13.2	31.2	47.4 ± 7.8	18.8				
		t	−156.6 ± 5.1	25.0	−178.8 ± 10.5	25.0				
		g^-	0.0	0.0	−77.3	6.2				
		Other		43.8		50.0				
Phe	16	g^+	67.8 ± 15.8	37.5						
		t	176.9 ± 2.7	12.5						
		g^-	−77.3	6.2						
		90			87.5 ± 14.9	87.4				
		Other		43.8		12.6				
Ser	24	g^+	70.4 ± 10.2	29.2						
		t	−179.7 ± 9.0	20.8						
		g^-	−59.1 ± 19.3	29.2						
		Other		20.8						
Thr	24	g^+	46.5 ± 12.5	33.3						
		t	−178.1 ± 0.4	8.3						
		g^-	−59.9 ± 15.5	29.2						
		Other		29.2						
Val	16	g^+	63.1	6.2						
		t	176.3 ± 9.1	25.0						
		g^-	−70.1 ± 18.4	43.8						
		Other		25.0						

distance geometry structures (ranging from 9.0 to 12.8 Å) due to the inability to stereospecifically assign the aromatic protons.[26]

In Figure 4 (a and b) we have plotted the deviations in the dihedral angle distributions against the surface accessibility and the number of side-chain NOE constraints for each of the χ_1 and χ_3 side-chain dihedral angles in NP-5. For χ_1 (and χ_2 not shown) the disorder is not well correlated with either the side-chain surface accessibility or the number of NOE constraints, whereas there is a better correlation for χ_3.

In addition to the side-chain conformational analyses, two aspects of the dihedral angles in the

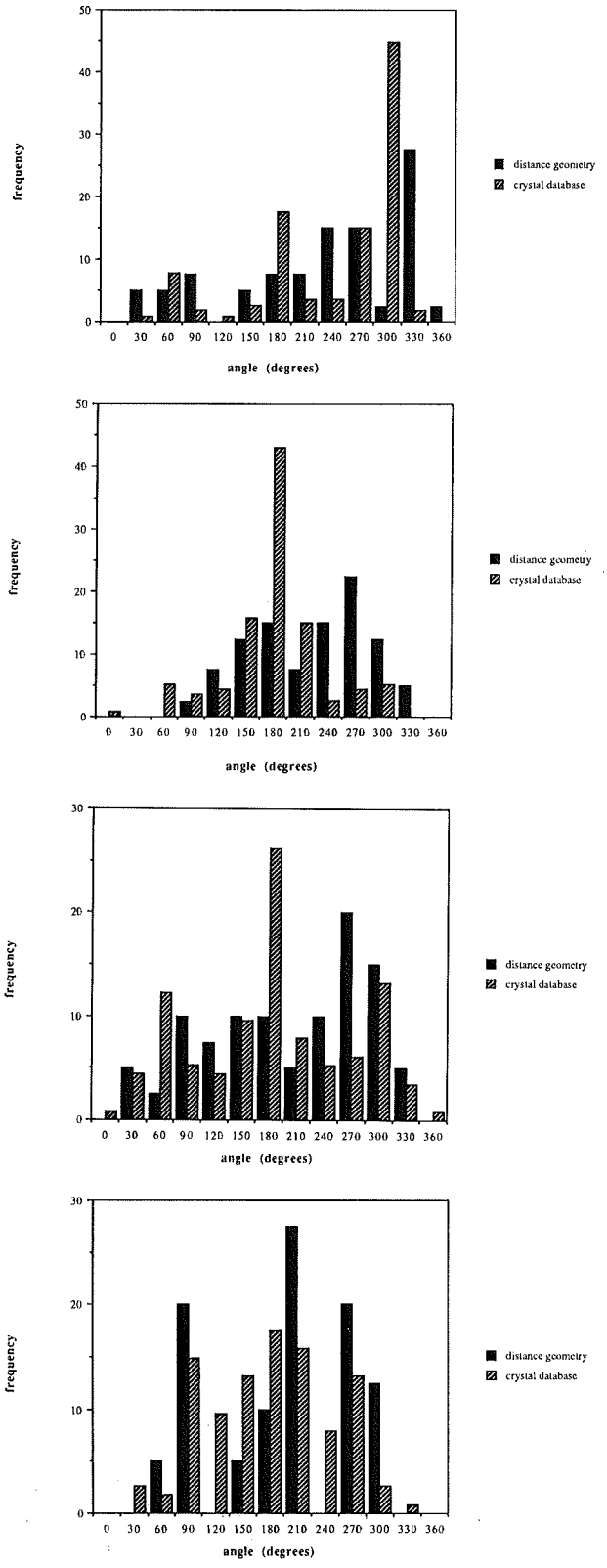

Figure 2 The comparison of the Arg side-chain dihedral angles in the high-resolution data base and the distance geometry structures of NP-5: (a) χ_1, (b) χ_2, (c) χ_3, and (d) χ_4.

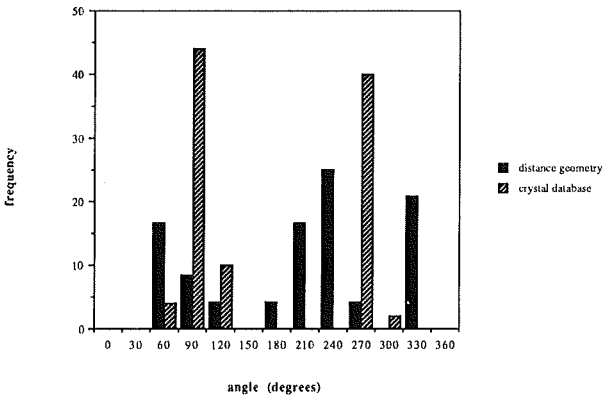

Figure 3 Comparison of dihedral angle about the disulfide bond (C_β-S_γ-S_γ-C_β) in the high-resolution data base and NP-5 distance geometry structures.

main chain of the protein have also been analyzed. The conformation of the β-hairpin extending between residues 19–29 has been analyzed as has the torsion about the peptide bond. These results are briefly reviewed in the following sections.

Conformation of the β-Hairpin

Figure 5 shows a comparison of the ϕ/ψ diagram for the residues in the β-hairpin constructed from the eight NP-5 distance geometry structures with residues in β-hairpins constructed from the crystal structure data base. It is evident that the backbone dihedral angles in the β-hairpin in NP-5 are well within the range of values existing in the high-resolution crystal data base. However, the conformation of the turn in the β-hairpin is not that of a typical β-hairpin. The criteria for classifying turns developed by Lewis et al.[42] is quite lax in its definition and any turn with a single dihedral angle differing by more than 50° from the ideal angles of a turn (types I, I′, II, II′, III, and III′) are classified according to the type it most closely resembles. Any turns with two or more dihedral angles differing by more than 40° from the standard types are assigned to a miscellaneous turn category, type IV. Of the eight turns analyzed in the original distance geometry structures, one was calculated to be in the type I conformation, one was in the type III′ conformation, and six were in the type IV conformation.

Analysis of the Peptide Torsion Angle ω

We have also analyzed the deviation of the peptide torsion angle from planarity in the NP-5 distance geometry structures and compared the results with

Table VII Average Accessibility of the Residues in NP-5 Distance Geometry Structures

Residue No.	Residue Name	Accessibility
1	Val	84.5
2	Phe	76.4
3	Cys	14.8
4	Thr	30.3
5	Cys	14.3
6	Arg	55.5
7	Gly	21.1
8	Phe	74.7
9	Leu	51.2
10	Cys	19.1
11	Gly	48.8
12	Ser	81.9
13	Gly	79.8
14	Glu	38.5
15	Arg	78.1
16	Ala	12.4
17	Ser	23.7
18	Gly	45.9
19	Ser	47.8
20	Cys	0.1
21	Thr	46.9
22	Ile	5.6
23	Asn	86.8
24	Gly	55.2
25	Val	39.3
26	Arg	67.9
27	His	5.5
28	Thr	23.8
29	Leu	4.7
30	Cys	22.7
31	Cys	36.5
32	Arg	41.6
33	Arg	96.1

those from the crystal data base. The average and standard deviation of the peptide dihedral angle calculated from 3104 values in the crystal data base is 179.4° and 6.1°, respectively (excluding the peptide bond in cis peptide proline residues). In comparison, the average and standard deviation of the torsion about the peptide bond in the NP-5 structures is calculated to be 180.8° and 21.0°, respectively. Peptide dihedral angles as small as 136° and as large as 225° were found in the NP-5 structures. The distribution of peptide torsion angles calculated from the crystal structures is compared with the distribution calculated from the NP-5 structures in Figure 6. It is clear that the distribution calculated from the NP-5 structures is too broad and that stronger constraints need to be applied to keep the peptide bond close to planar.

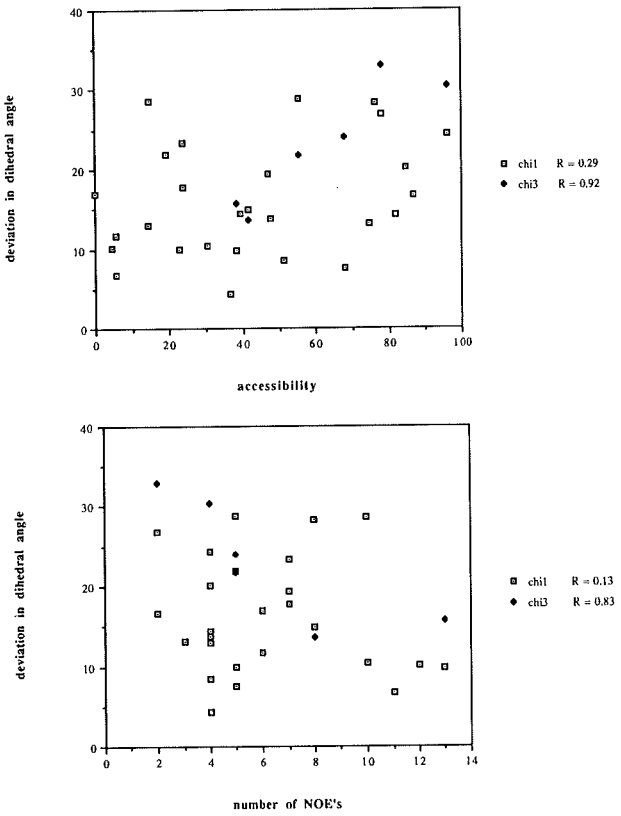

Figure 4 (a) Plot of weighted average standard deviation in the side-chain dihedral angles vs the solvent accessibility for each amino acid in NP-5. (b) Plot of the number of interresidue NOE constraints to a residue vs accessibility for amino acids in NP-5.

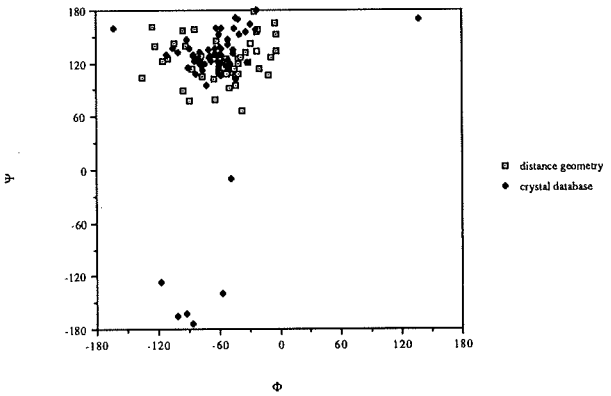

Figure 5 The ϕ/ψ map of the backbone dihedral angles in the sheet region of the β-hairpin in the solution structures and the high-resolution data base. (*) Regions in the high-resolution data base were selected to correspond with those determined by Milner-White and Poet.[46] The following hairpins were selected: 2SGA(34-39), 2SGA(48C-48D), 2SGA(84-85), 2SGA(202-207), 3TLN(36-37), 3TLN(251-252), 2APP(24-25), 2APP(93-94), 2APP(201-202), 2APP(261-262), ISN3(43-44).

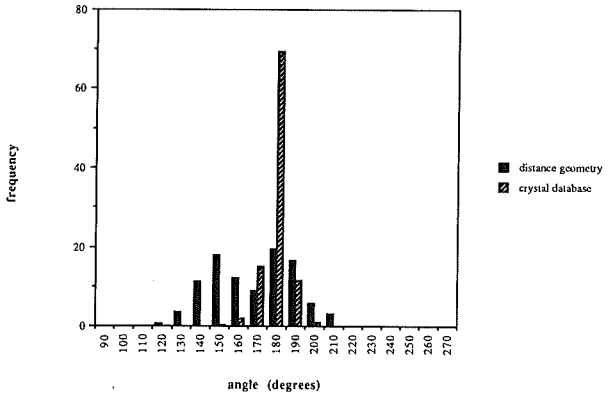

Figure 6 Dihedral angle distributions for the peptide dihedral angle in the high-resolution data base and the distance geometry structures of NP-5.

Constrained Energy Minimization of the NP-5 Structures

After completing the conformational analysis described in the preceding sections, the NP-5 distance geometry structures were refined using constrained energy minimization. The resultant molecular mechanics energies are listed in Table VIII. There is a large decrease in the total energy and in each of the component energies in the NP-5 structures after minimization. Close van der Waals contacts have been relieved, and the bond lengths and angles have improved stereochemistry. We were particularly interested in analyzing the effect of refinement on the dihedral angle distributions analyzed in this paper; specifically the peptide, the disulfide, and the side-chain dihedral angles. Since the molecular mechanics function contains a relatively steep potential restraining the torsions about the peptide bond, it was hoped that refinement of the peptide bond geometry would lead to a narrower distribution about the *trans* peptide geometry. The peptide bond dihedral angle distributions before and after refinement are compared in Figure 7a. While the peptide bond geometry is somewhat improved, the distribution is still too large. Similar results were obtained for the torsion about the disulfide bond. As discussed above, there is a strong preference for this dihedral angle to be ± 90° in the crystal structures. The torsional potential energy function for this dihedral angle has minima at ± 90° and in the energy-minimized structures the disulfide bond geometries are in better agreement with the conformational preferences observed in the crystal structures (Figure 7b). Energy minimization did not lead to more regular β-hairpin turns. After constrained minimization, the β-hairpins in the eight structures were classified as type I′ in two of the structures and type IV in the remaining six structures. We note that the interconversion between different hairpins would require the disruption of the hydrogen bonding in the hairpin. With respect to the side-chain conformational distributions, refinement did not significantly alter the initial distributions calculated from the distance geometry structures, although there is some shift in the values toward local minima in the torsional potentials. However, the side-chain torsional distributions remain much broader than the distributions observed in the crystal data base.

DISCUSSION

The distributions of side-chain conformations have been analyzed in solution structures of NP-5 that

Table VIII Molecular Mechanics Energies[a] of NP-5 Structures Before and After Minimization

Structure	1	2	3	4	5	6	7	8	⟨DG⟩[b]
E_{total}	−580.9	−499.5	−595.2	−483.2	−478.8	−604.5	−518.7	−624.7	7,544.5
E_{bond}	11.9	12.9	13.7	14.8	17.6	12.4	11.0	13.0	713.5
E_{angle}	77.4	124.9	96.3	114.1	131.4	91.8	94.0	90.2	5,726.1
E_{ϕ}	114.1	110.9	109.4	147.2	102.2	110.1	136.2	118.3	182.7
E_{LJ}	−89.0	−79.6	−89.5	−78.3	−70.4	−94.2	−78.3	−93.3	360.8
E_{EL}	−695.3	−668.6	−725.1	−681.0	−659.6	−752.9	−681.6	−752.9	561.9
Root mean square gradient	0.11	0.09	0.40	0.08	0.04	0.07	0.06	0.06	—
d_{viol}[c]	0.89	1.15	0.95	0.92	1.23	1.05	0.87	0.90	9.8[d]

[a] Energies in kcal/mol.
[b] Energies averaged over eight original NP-5 solution structures before minimization.
[c] Square root of the sum of the squares of the distance violations.
[d] The reason for the large distance violation involves use of different upper bounds for the distance geometry and minimization procedures as previously discussed.[14]

Figure 7a

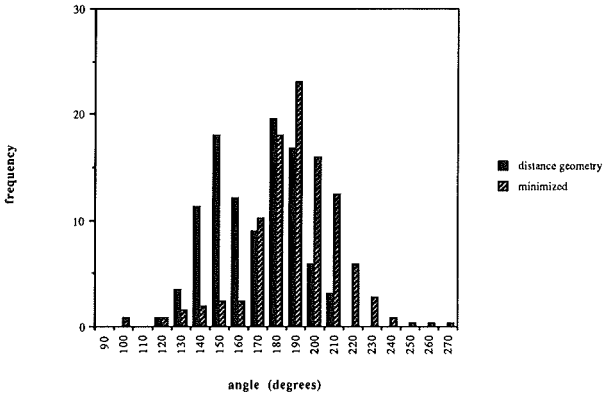

Figure 7b

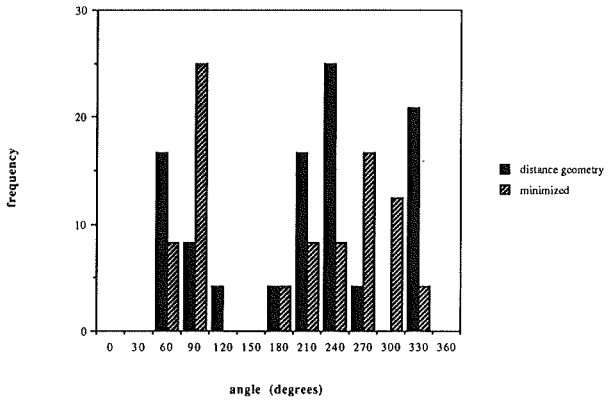

Figure 7 (a) Dihedral angle distributions for rotation about the peptide bond before and after energy refinement. (b) Dihedral angle distribution for rotation about the disulfide bond before and after energy refinement.

were generated from nmr-derived distance data, and these distributions were compared with the distributions of side-chain conformations obtained from a data base consisting of high-resolution x-ray crystal structures of proteins. There is generally a much larger degree of conformational variability in the solution structures than in the data-base structures. It is tempting to ascribe these differences to the greater conformational mobility in solution than in a crystal. However, there are variety of factors, discussed in this section, which could increase the apparent disorder in the solution structures.

One of the most important factors that can contribute to the disorder observed in the NP-5 structures is the number and kind of constraints used to generate the structures. We have attempted to analyze this factor by evaluating the extent to which the side-chain disorder is correlated with the number of NOE constraints to the side chain. Surprisingly, there does not appear to be a strong correlation between the variance in the side-chain conformational distribution and the number of constraints to the side chain. There are a number of possible explanations for this counterintuitive result. First, there is the problem of small sample size; there are very few side chains in NP-5 with large numbers of interresidue NOE constraints to the side chains. There are only three residues that have more than 10 interresidue constraints involving the side chains; the largest number of constraints observed to a residue is 13 (Glu 14). For these residues in fact, the standard deviations in the distributions for χ_1 are among the smallest observed in NP-5. Furthermore, residues with large numbers of constraints tend to have larger side chains and for these residues, correlated changes in the dihedral angles can reduce the displacements of the atoms in the side chain.

Another difficulty encountered in attempting to correlate side-chain disorder with the number of constraints is the measure chosen to quantify disorder. We have chosen to use the weighted average variance in the torsional distributions as a measure of disorder. The side-chain distributions observed in the NP-5 distance geometry structures do not have a simple functional form and are difficult to characterize by a single number such as the variance. For torsional distributions with multiple peaks that are both sharp and widely separated, although the variance will be very large, the disorder is discrete, involving a small number of rotomeric states. This type of disorder is very different from that resulting from a very broad unimodal distribution.

It is important to realize that there is a statistical component to the analysis performed here and that it is not necessarily true that the distribution for a single (or small number of) residue(s) in the NP-5 solution structures reflect the average distribution of a very large number residues in the data base. For example, there is only a single histidine residue in NP-5, and its distributions for the χ_1 and χ_2 dihedral angles are quite different from the average distributions in the data base. This could simply mean that this particular residue is in an unusual conformation and the differences in the two distributions is not meaningful, whereas if there were 10 histidines in NP-5 then it would be much more likely that their average distributions should correspond to the data base. Thus care should be taken to avoid overinterpreting the differences between the NP-5 and data-base side-chain distributions.

That side-chain populations in high-resolution crystal structures have a very strong preference for a small number of rotameric states suggests that *local* (both electronic and nonbonded) interactions

play a major role in determining the conformations of side chains in proteins. In the distance geometry calculations from which the NP-5 structures were generated, however, only interresidue NOEs were included in the constraints because it was felt that for the intraresidue NOEs, the distance ranges were too large to limit the conformational freedom of the side chains. We expect that the quality of the structures, and particularly the placement of the side chains, will improve when both intraresidue NOEs and spin–spin coupling constants are used as additional constraints in the structure generation process. Such improvement was seen in the recent structure determination of the 75-residue protein Tendamistat.[43]

We performed constrained energy minimization on the NP-5 structures, to generate energetically more favorable structures while also trying to eliminate bias in the structures introduced by the distance geometry procedure.[4,44,45] Energy minimization of the distance geometry structures did not appear to significantly alter the distributions of the side-chain conformations in NP-5 and a more robust algorithm such as molecular dynamics or Monte Carlo simulation is needed to surmount local barriers. In any case, side-chain rotational distributions calculated using the molecular mechanics torsional potentials alone do not agree well with those calculated from the crystal data base (Kominos and Levy, unpublished) and further parameterization of the side-chain potential functions is needed.

We comment on the refinement of the peptide dihedral angle and the torsion about the disulfide bond. Although energy refinement lead to more correct distributions for these degrees of freedom, they were still much too broad as compared with the distributions observed in the crystal data base. The torsional force constants used in the molecular mechanics function were parameterized from vibrational data (which is sensitive to the local shape of the potential about the minimum) and much larger restoring forces are required to correct deviations from peptide planarity as large as 100°, as was observed in the distance geometry structures. We have experimented with refinement of NP-5 distance geometry structures using artificially high force constants on the peptide torsion. When the force constant is increased from 10 to 100 kcal/mol, the rms deviation from peptide planarity decreases to 10°, which is in closer agreement with the crystal data base (rms deviation 6°). The very narrow rotational distribution observed for the disulfide bond in the crystal data base is very likely an electronic effect and should not be very sensitive to the crystal vs solution environment. Therefore further increase in the force constant for rotation about the disulfide bond should also improve the refinement of the NP-5 structures.

NP-5 is a member of a family of peptides called "defensins." The high-sequence homology in this family suggests that essentially all the defensins have the same basic backbone topology, yet there is a wide variation in their activity. An understanding of side-chain structure and dynamics in these peptides is required in order to understand the structural basis for differential activity. With the recent crystal structure determination of the HNP-3 structure (Eisenberg et al., to be published), the defensin family is an excellent candidate for high-resolution studies of environmental effects on side-chain properties. We plan to study the NP-5 solution structure at "higher resolution" by incorporating additional nmr-derived information including all intraresidue NOEs and all ϕ and χ_1 coupling constants into the structure determination procedure. It will be of interest to see whether the side-chain conformational patterns in these higher resolution structures more closely resemble those observed in the high-resolution data base.

APPENDIX: USE OF A RELATIONAL DATA BASE

The analysis reported in this paper was carried out with the aid of a commercially available relational data-base management system INGRES.[46] (1986). In a relational data base, information is stored in tables. The rows correspond to objects and the columns correspond to attributes of the objects. The power of the relational data base is the ability to form new tables from existing tables according to the rules of relational algebra. A relational data-base management system has been used previously to analyze protein structural properties.[23] In our construction of a data base for this project two types of tables were built, generic tables and protein specific tables. The generic tables contain information applicable to all the proteins in the data base, for example the dihedral angle definition table contains for each dihedral angle type the following information: a list of the atom types that make up the dihedral angle, and the dihedral angle name and number (e.g., the backbone dihedral angle ϕ contains the atom types N-C-C$^{\alpha}$-N). As another example of

a generic table, we constructed a table of the accessible surface areas of the side chains for each of the 20 amino acids in standard extended conformations.[35] For each protein in the data base several protein-specific tables were constructed. For example, an atomic coordinate table was constructed for each protein, which consisted of the following information for each atom of the protein: atom number, atom name, residue number, residue name, cartesian coordinates, and accessible surface area. The dihedral angle table had a row entry for each dihedral angle in the protein containing the residue number, the residue name, the dihedral name, and the numerical value. The protein structural data that formed the basic information contained in the protein specific tables were calculated using the IMPACT modeling program.[13] An important attribute of a relational data base is that new data can be incorporated in a new table, which can then be easily correlated with existing data. For example, once the dihedral angle tables were constructed for each of the residues in the data base, the side-chain properties could be correlated with additional geometric features such as secondary structure at a later date without having to rebuild the tables.

Although the use of INGRES facilitated the work described in this paper, there are several drawbacks to using a commercially available data-base management package for scientific work. The commercially available data-base managers have been designed for business applications. Thus, such simple scientific functions as sine and cosine are not available. Furthermore, the commercial software is not readily integrated into a large modeling program like IMPACT. There are conflicts involving file structure, portability, and command language. Subsequent to the calculations described in this paper, we have written relational data-base software that has been incorporated directly into IMPACT and we plan to use these software tools in further studies of the kind described in this paper.

This work has been supported by grants from the National Institutes of Health, NIH GM-30580 (to RML) and NIH AI-27026, and the Searle Scholars Program of the Chicago Community Trust 85-C110 (to AP), by a grant from the Johnson and Johnson Research Foundation, and by the award of Supercomputer Graduate Fellowships (to DK and DAB) from the New Jersey Commission on Science and Technology. We thank the National Allocation Committee of the John Von Neumann Supercomputer Center for the generous award of computer time and Dr. Douglas Kitchen for help with the refinement of the NP-5 structures.

REFERENCES

1. Kuntz, I. D., Crippen, G., Kollman, P. A. & Kimmelman, D. (1976) *J. Mol. Biol.* **106,** 983–994.
2. Crippen, G. M. (1977) *J. Comp. Phys.* **24,** 96–107.
3. Braun, W., Bösch, C., Brown, L. R., Go, N. & Wüthrich, K. (1981) *Biochim. Biophys. Acta* **667,** 377–396.
4. Havel, T. F. & Wüthrich, K. (1985) *J. Mol. Biol.* **182,** 281–294.
5. Clore, G. M., Gronenborn, A. M., Brünger, A. T. & Karplus, M. (1985) *J. Mol. Biol.* **185,** 435–455.
6. Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Brelens, R. & van Gunsteren, W. F. (1985) *J. Mol. Biol.* **182,** 179–182.
7. Brünger, A. T., Clore, G. M., Gronenborn, A. & Karplus, M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3801–3805.
8. Clore, G. M., Brünger, A. T., Karplus, M. & Gronenborn, A. M. (1986) *J. Mol. Biol.* **191,** 523–551.
9. Braun, W. & Go, N. (1985) *J. Mol. Biol.* **186,** 611–626.
10. Wagner, G., Braun, W., Havel, T, Schaumann, T., Go, N. & Wüthrich, K. (1987) *J. Mol. Biol.* **196,** 1–29.
11. Holak, T. A., Prestegard, J. H. & Forman, J. D. (1987) *Biochemistry* **26,** 4652–4660.
12. Holak, T. A., Kearsley, S. K., Kim, Y. & Prestegard, J. H. (1988) *Biochemistry* **27,** 6135–6142.
13. Bassolino, D. A., Hirata, F., Kitchen, D., Kominos, D., Pardi, A. & Levy, R. M. (1988) *Intl. J. Supercomp. Appl.* **12,** 41–61.
14. Levy, R. M., Bassolino, D. A., Kitchen, D. B. & Pardi, A. (1989) *Biochemistry* **28,** 9361–9372.
15. Clore, G. M., Gronenborn, A. M., Nilges, M. & Ryan, C. A. (1987) *Biochemistry* **26,** 8012–8023.
16. Kline, A. D., Braun, W. & Wüthrich, K. (1988) *J. Mol. Biol.* **204,** 675–724.
17. Billeter, M., Kline, A., Braun, W., Huber, R., & Wüthrich, K. (1989) *J. Mol. Biol.* **206,** 677–687.
18. Janin, J., Wodak, S., Levitt, M. & Maigret, B. (1978) *J. Mol. Biol.* **125,** 357–386.
19. Bhat, T. N., Sasisekharan, V. & Vijayan, M. (1979) *Int. J. Peptide Protein Res.* **13,** 170–184.
20. Benedetti, E., Morelli, G., Nemethy, G. & Scheraga, H. A. (1983) *Int. J. Peptide Protein Res.* **22,** 1–15.
21. James, M. N. G. & Sielecki, A. R. (1983) *J. Mol. Biol.* **163,** 299–361.
22. Ponder, J. W. & Richards, F. M. (1987) *J. Mol. Biol.* **193,** 775–791.
23. McGregor, M. J., Islam, S. A. & Sternberg, M. J. E. (1987) *J. Mol. Biol.* **198,** 295–310.
24. Piela, L., Nemethy, G. & Scheraga, H. A. (1987) *Biopolymers* **26,** 1273–1286.
25. Bach, A. C., II, Selsted, M. E. & Pardi, A. (1987) *Biochemistry* **26,** 4389–4397.
26. Pardi, A., Hare, D. R., Selsted, M. E., Morrison, R. D., Bassolino, D. A. & Bach, A. C. (1988) *J. Mol. Biol.* **201,** 625–636.

27. Selsted, M. E., Brown, D. M., DeLange, R. J., Harwig, S. S. L. & Lehrer, R. I. (1985) *J. Biol. Chem.* **260,** 4579–4584.
28. Selsted, M. E., Harwig, S. S. L., Ganz, T., Schilling, J. W. & Lehrer, R. I. (1985) *J. Clin. Invest.* **76,** 1436–1439.
29. Selsted, M. E., Szklarek, D. & Lehrer, R. I. (1984) *Infect. Immunol.* **45,** 150–154.
30. Lehrer, R. I., Szlarek, D., Ganz, T. & Selsted, M. E. (1985) *Infect. Immunol.* **49,** 207–211.
31. Lehrer, R. I., Szklarek, D., Ganz, T. & Selsted, M. E. (1985) *J. Virol.* **54,** 467–472.
32. Havel, T. F., Kuntz, I. D. & Crippen, G. M. (1983) *Bull. Math. Biol.* **45,** 665–720.
33. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
34. Levitt, M. & Chothia, C. (1976) *Nature* **261,** 552–558.
35. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400.
36. Shrake, A. & Rupley, J. A. (1973) *J. Mol. Biol.* **79,** 351–371.
37. Chothia, C. (1976) *J. Mol. Biol.* **105,** 1–14.
38. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comp. Chem.* **7,** 230–252.
39. Fletcher, R. & Reeves, C. M. (1983) *Computer J.* **7,** 149–154.
40. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339.
41. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. (1985) *Science* **229,** 834–838.
42. Lewis, P. N., Momany, F. A., & Scheraga, H. A. (1973) *Biochim. Biophys. Acta* **303,** 211–229.
43. Kline, A. D., Braun, W. & Wüthrich, K. (1986) *J. Mol. Biol.* **189,** 377–387.
44. Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988) *FEBS Lett.* **299,** 317–324.
45. Metzler, W. J., Hare, D. R. & Pardi, A. (1989) *Biochemistry* **28,** 7045–7052.
46. INGRES (1986) Relational Database Management System, Relational Technology, Inc., Alameda, CA.
47. Milner-White, E. J. & Poet, R. (1986) *Biochem. J.* **240,** 289–292.
48. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comp. Chem.* **4,** 187–217.
49. Pflugrath, J. W., Wiegand, G., Huber, R. & Vertesy, L. (1986) *J. Mol. Biol.* **189,** 377–387.
50. Sibanda, B. L. & Thornton, J. M. (1985) *Nature* **316,** 170–174.
51. Summers, N. L., Carlson, W. D. & Karplus, M. (1987) *J. Mol. Biol.* **196,** 175–198.

Received August 29, 1989
Accepted December 11, 1989