Solution Structures of Proteins from NMR Data and Modeling: Alternative Folds for Neutrophil Peptide 5†

Ronald M. Levy,* Donna A. Bassolino, and Douglas B. Kitchen
Department of Chemistry, Rutgers University, New Brunswick, New Jersey 08903
Arthur Pardi*
Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, Colorado 80309
Received April 20, 1989; Revised Manuscript Received July 14, 1989

ABSTRACT: The structure of neutrophil peptide 5 in solution has recently been reported (Pardi et al., 1988). The structure determination was accomplished by using a distance geometry algorithm and 107 interproton distance constraints obtained from 2D NMR data. In each of the eight independent solutions to the distance geometry equations, the overall fold of the polypeptide backbone was identical and the root mean square (rms) deviation between backbone atoms of the superimposed structures was small (~2.4 Å). In this paper we report additional NP-5 structures obtained by using a new structure generation algorithm: a Monte Carlo search in torsion angle space. These structures have a large rms backbone deviation from the distance geometry structures (~5.0 Å). The backbone topologies differ in significant respects from the distance geometry structures and from each other. Structures are found that are pseudo mirror images of part or all of the fold corresponding to that first obtained with the distance geometry procedure. For small proteins, the problem of distinguishing the correct structure among pseudo mirror images is likely to be greater than previously recognized. When a set of test distance constraints constructed from a novel Monte Carlo structure is used as input in the distance geometry algorithm, the fold of the resulting structure does not correspond to that of the target. The results also demonstrate that the previously accepted criteria (the magnitude of the rms deviation between multiple solutions of the distance geometry equations) for defining the accuracy and precision of a peptide structure generated from NMR data are inadequate. An energetic analysis of structures corresponding to the different folding topologies has been carried out. The molecular mechanics energies obtained by minimization and molecular dynamics refinement provide sufficient information to eliminate certain alternative structures. On the basis of a careful comparison of the different trial structures with the experimental data, it is concluded that the NP-5 peptide fold which was originally reported is most consistent with the data. An alternative fold corresponding to structures with low energies and small total distance violations is ruled out because for this fold predicted NOEs are not observed experimentally.

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

During the past few years enormous progress has been achieved in the determination of the structure of proteins in solution using information derived from 2D NMR data (Wüthrich, 1986). There are two conceptually distinct steps in the structure determination process: (1) the use of the NMR data to generate a set of structural constraints that consists primarily of internuclear distance ranges and (2) the determination of three-dimensional structures based on these constraints using an appropriate computer algorithm. The distance geometry algorithm was the first algorithm successfully applied to this problem (Crippen, 1977; Havel et al., 1985; Kline et al., 1986); it is an efficient method for generating structures and is widely used. Recently, questions have been raised concerning the sampling properties of the distance geometry algorithm (de Vlieg et al., 1988; Brünger et al., 1987; Metzler et al., 1989). de Vlieg et al. (1988) have shown that further refinement of distance geometry structures using restrained molecular dynamics improves the conformational sampling. Brünger et al. (1987) compared restrained molecular dynamics with the distance geometry method for generating solution conformations for a system without long-range distance data and found that from the distance geometry results alone it was possible to incorrectly conclude that there was a unique solution conformation. Metzler et al. (1989) have shown that, in the limit of minimal distance information, the distance geometry procedure tends to generate overly extended structures and sample only a small region of conformational space. We have obtained further information regarding the conformational sampling limitations of the distance geometry algorithm in our structural studies of rabbit neutrophil defensin peptide NP-5.[1] The problem, which is expected to be of general importance, concerns the need to distinguish among solutions to the distance constraint equations which consist of pseudo mirror image structures for part or all of the peptide backbone.

NP-5, a 33-residue polypeptide, is a member of the defensin family of peptides. These low molecular weight cationic peptides are active antimicrobial agents and rely on an oxygen-independent mechanism for their activity (Selsted et al., 1985). Ten members of the defensin family have been isolated to date. The structure of NP-5 in solution has recently been reported (Pardi et al., 1988). The structure determination was reported by using a distance geometry algorithm and 107 interproton distance constraints obtained from 2D NMR data. In each of the eight independent solutions to the distance geometry equations, the overall fold of the polypeptide back-

† This work has been supported by grants from the National Institutes of Health, NIH GM-30580 (to R.M.L.) and NIH AI-27026, and the Searle Scholars Program of the Chicago Community Trust, 85-C110 (to A.P.), by a grant from the Johnson and Johnson Research Foundation, and by the award of a Supercomputer Graduate Fellowship (to D.A.B.) from the New Jersey Commission on Science and Technology.
* To whom correspondence should be addressed.

[1] Abbreviations: DG, distance geometry; MC, Monte Carlo; MD, molecular dynamics; NP-5, neutrophil peptide 5.

0006-2960/89/0428-9361$01.50/0 © 1989 American Chemical Society

bone was identical and the root mean square (rms) deviation between backbone atoms of the superimposed structures was small (~2.4 Å). In this paper we report additional structures obtained by using a different structure generation algorithm: a Monte Carlo search in torsion angle space (Bassolino et al., 1988). These structures have a large rms backbone deviation from the distance geometry structures. Furthermore, while some of the structures have the same fold as previously reported, for others, the backbone topologies differ in significant respects from the distance geometry structures and from each other. These results demonstrate that the current implementation of the distance geometry procedure does not adequately sample conformational space, even at the level of determining the polypeptide backbone geometry. The results also demonstrate that the previously accepted criteria (the magnitude of the rms deviation between multiple solutions of the distance geometry equations) for defining the accuracy and precision of a peptide structure generated from NMR data are inadequate. In the following section we present the internal coordinate Monte Carlo structure generation algorithm. Under Results we describe the different overall folds for NP-5 obtained as solutions to the distance constraint equations as well as those obtained by using the Monte Carlo search procedure. The molecular mechanics energies and distance violations corresponding to the different structures are reported. Under discussion we discuss the criteria used to choose the correct overall fold for NP-5. Special attention is paid to the way the NOEs are distributed throughout the molecule and the importance of specific NOEs in determining the fold.

Methods

We have incorporated the NMR information into the structure generation and refinement by creating a set of distance constraints from the NOE data and adding the constraints as a harmonic penalty term in the target function (Bassolino et al., 1988). This is the standard procedure that has been employed by most groups. We note however that it may be advantageous to refine against NOE intensities directly, since a unique distance constraint can be derived from an NOE only when the following conditions are satisfied: resonances are not overlapping, the effects of spin diffusion and internal motions are neglected, and overall tumbling is isotropic. In practice these simplifying approximations may be adequate for many situations of interest, although for the generation of high-resolution structures, these effects require further study.

The distance constraints employed in the present calculations were extracted from 2D NOE spectra (Pardi et al., 1988; Bach et al., 1987). The set of constraints included 94 NOE distances and three disulfide bonds between residues 3–31, 5–20, 10–30, respectively. The fewer distance constraints used in the Monte Carlo calculations as compared with the distance geometry calculations simply reflect how NOE constraints for nonequivalent protons on prochiral centers are handled, as discussed below. Following the procedure employed in the distance geometry calculations on NP-5 (Pardi et al., 1988) the distances were classified into three groups: short ($d < 2.8$ Å), medium (2.8 Å $\leq d \leq$ 3.8 Å), and long ($d > 3.8$ Å). The following bounds were added to the distances in each group: short, $d \pm 0.5$ Å; medium, $d - 0.5/d + 1.0$ Å; and long, $d \pm 1.0$ Å. Unlike the distance geometry calculations, no pseudo atoms were employed and there was no relaxation of chirality constraints for nonequivalent protons on prochiral centers (Pardi et al., 1988). Instead, for methylene, methyl, and aromatic groups a r^{-6} weighted average distance over "equivalent" protons was used (Brünger et al., 1986). For example, for an NOE between the methyl protons on a residue and a proton j, the weighted average distance used is given by

$$\langle r^{-6}\rangle^{-1/6} = \left\{\frac{(r_{C\gamma H1-j})^{-6} + (r_{C\gamma H2-j})^{-6} + (r_{C\gamma H3-j})^{-6}}{3}\right\}^{-1/6} \quad (1)$$

This method for calculating an "effective" distance to a methyl group appropriately weights the distance from proton j to the nearest methyl proton much more heavily. Although we have followed the procedure of Brünger et al. (1987) and divided the sum of the individual NOE interactions by the number of overlapping resonances, another reasonable choice would be to define an effective distance by using the sum of the separate interactions instead of the average; this would in fact more closely correspond to refining against intensities. A related issue concerns the procedure used to extract a target distance from experimental NOE intensities when there are overlapping resonances. In the present study the assignment to strong, medium, or weak NOE for methyls is based on the total NOE intensity for the methyl group. Considering that the classification of NOEs is qualitative and that relatively large error bounds are added to the target distances, the errors introduced by this procedure are small. For example, the difference between the use of a weighted average distance and the distance to the closest proton is always less than 12% for methylene protons and even smaller for methyl protons. These differences are well within the error bounds added to the target distances. Also, for resolved resonances where stereospecific assignments are not available, the r^{-6} weighted distance averaging corresponds to the limit of rapid spin equilibration due to spin diffusion (Olejniczak et al., 1984). Thus for methylene protons that are separated by ~1.7 Å this is expected to be a reasonably good approximation, even when stereospecific assignments are available. The use of a weighted average distance instead of the introduction of pseudoatoms leads to a set of much tighter distance constraints. The bounds used in the present study are comparable to those used in a recent protein structure determination by restrained molecular dynamics (Clore et al., 1985) but much tighter than those normally used in distance geometry and restrained molecular dynamics calculations (Pardi et al., 1988; Wüthrich, 1986; Kline et al., 1986; de Vlieg et al., 1988; Williamson et al., 1985; Wagner et al., 1987; Braun & Gō, 1985).

From a qualitative analysis of the 2D NOE patterns for NP-5 it was possible to identify a secondary structural feature consisting of a β-hairpin between residues 19 and 29 with residues 23 and 24, forming a type I′ β-turn (Pardi et al., 1987). In the structure used to initiate the Monte Carlo sampling, the backbone torsion angles between residues 19 and 29 were initially set to ideal β values; all other backbone (ϕ,ψ) pairs were set to trans values. The β-turn was not built into the initial structure. All bond lengths and angles were set to standard values during the torsion angle sampling. The target function for the Monte Carlo search included distance constraint terms and additional terms of the molecular mechanics type. The distance constraints were incorporated by using a harmonic penalty function as described previously (Bassolino et al., 1988). In some of the calculations, the "molecular mechanics" term consisted of a single uniform nonbonded repulsion between all nonbonded pairs which switched on if $d_{ij} \leq 2.5$ Å; a standard molecular mechanics expression was used in most calculations which included a Lennard–Jones 6–12 potential as well as electrostatic and hydrogen bond interactions.

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

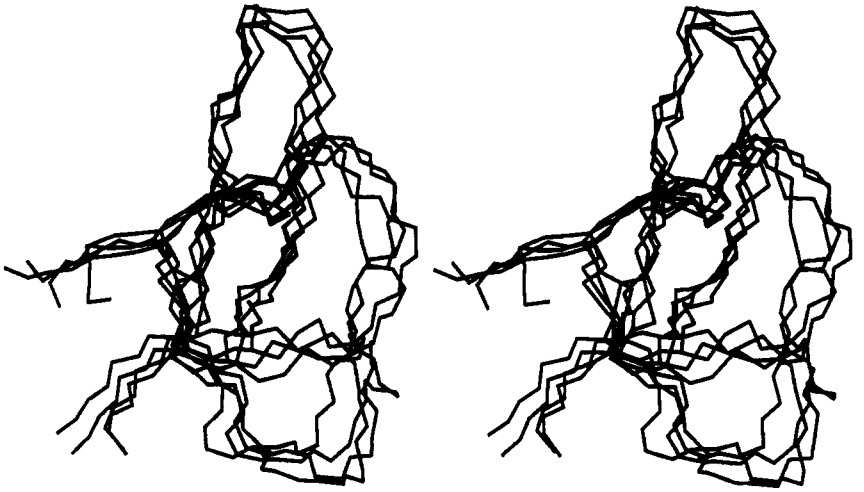

FIGURE 1: Stereoview of the standard fold of NP-5. Four structures (DG-1–4) representative of those generated by the distance geometry algorithm are superimposed, as the best fit of the N, C^{α}, C, O, and C^{β} atoms.

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

The internuclear distance constraints were incorporated into the calculations in stages as first proposed by Braun and Gō (1985). The 94 distance constraints were divided into five sets representing increasingly distant (in sequence) interactions (Table I). In the first stage of the structure generation only the first set containing 44 distances corresponding to sequential $[i-(i+1)]$ backbone protons were included in the target function. The Monte Carlo sampling was performed for 10000 steps at 300 K with an initial step size $-5° \leq \Delta\alpha \leq +5°$. The step size was adjusted continuously to ensure a 40%–60% acceptance rate. The sampling was performed over all backbone and side-chain torsions. The additional distance constraints (sets 2–5, Table I) were added sequentially; between 2000 and 20000 MC steps were performed at each stage. In the latter part of this study the structure generation process was simplified; the distance constraints were divided into two sets; i.e., sets 1–3 were merged as were sets 4 and 5. The resulting structures were minimized (in Cartesian space) for 2000 conjugate gradient cycles. Several minimizations were carried out; the relative weights on the distance constraint terms and the molecular mechanics terms were varied in different trials. The results of refinement by quenched molecular dynamics are also reported. All the calculations (structure generation and optimization by minimization and molecular dynamics) reported in this paper were carried out with the program package IMPACT (Bassolino et al., 1988). The total CPU time required to generate and refine an NP-5 structure by the Monte Carlo procedure is approximately 4–7 h on a Convex C210 minisupercomputer. For comparison, approximately 1 h is required to generate a structure by the distance geometry procedure. Recent modifications to the Monte Carlo search subsequent to this study have resulted in a severalfold increase in the efficiency of the algorithm.

Table I: Constraint Sets[a] Used in Monte Carlo Calculations

set no.	type	no. of constraints	%
1	backbone–backbone sequential only	43	45.7
2	backbone–backbone backbone–side chain $[i-(i+5)$ residues]	7	7.4
3	side chain–side chain $[i-(i+5)$ residues]	12	12.7
4	backbone–backbone backbone–side chain side chain–side chain $[i-(i+10)$ residues]	5	4.2
5	long-range interactions ($\Delta i > 10$)	27	28.7

[a] Set 1 includes 43 sequential backbone–backbone interactions involving H^{α} and/or HN protons. Set 2 includes interactions between backbone–backbone and backbone–side chain protons separated by up to 5 residues. Set 3 includes all side chain–side chain proton interactions for protons separated by up to 5 residues. Set 4 contains all the medium-range interactions, which includes all protons separated by up to 10 residues. Set 5 contains all the long-range interactions between protons separated in the sequence by more than 10 residues.

RESULTS

Overall Fold. Figure 1 shows a stereopair superposition of the peptide backbone for four structures generated by the distance geometry procedure. The polypeptide fold has been described previously (Pardi et al., 1988). The most readily distinguished feature of the structure is a β-hairpin formed by residues 19–29. The hairpin is oriented vertically in Figure 1. From the N terminus, the chain proceeds toward the hairpin, where it makes a sharp bend at residue 6 in front of the hairpin. The chain then proceeds approximately parallel to the hairpin until it forms a loose type I β-turn between residues 11 and 14. From here it enters the β-hairpin at residue 18; after completing the β-hairpin at residue 29, the N and C termini of the chain end in close proximity. It is apparent from the picture and the rms deviation between pairs of structures (Table II) that the polypeptide fold is the same for each of the structures.

Sixteen NP-5 structures have been generated with the Monte Carlo procedure. Of these, 13 structures had acceptably low values for the total violation of the distance constraints after refinement. The pairwise rms deviation between superimposed structures is listed in Table II. It is evident there is a much larger variation in these structures, for which the average rms deviation between superimposed backbone atoms is ~5 Å as compared with ~2 Å among the structures generated by the distance geometry procedure. Visual inspection on an Evans and Sutherland PS 390 showed that these structures could be classified into a number of different folding topologies according to the relative orientation of three molecular domains. Domain 1 corresponds to N- and C-terminal sequences (residues 1–10 and 30–33) which are "bonded" via two disulfide bonds between residues 3–31 and 10–30. Domain 2 corresponds to an extended loop between residues 11 and 18, and domain 3 corresponds to the β-hairpin between residues 19 and 29. Five distinct backbone topologies have been observed among the 16 structures generated by the Monte Carlo procedure. The "standard" fold, described

Table II: RMS Deviation between NP-5 Structures

	Distance Geometry Structures[a]						
	set 1	set 2	set 3	set 4	set 5	set 6	set 7
set 0	1.88	1.81	1.55	2.18	2.80	1.99	1.81
set 1		2.30	1.98	1.83	3.10	2.00	1.89
set 2			1.57	2.20	2.69	2.72	2.41
set 3				2.23	2.69	2.43	2.01
set 4					2.90	2.65	2.23
set 5						3.49	3.23
set 6							1.81

	Monte Carlo Structures[b]							
	MC-3	MC-4	MC-5	MC-6	MC-7	MC-8	MC-13	DG-3
MC-2	5.29	3.62	5.55	1.74	2.06	1.86	5.05	4.82
MC-3		4.98	4.32	5.58	5.71	5.54	4.27	4.70
MC-4			5.53	4.16	3.59	3.83	5.73	5.68
MC-5				5.46	5.73	5.60	5.59	6.07
MC-6					2.29	1.64	5.26	5.11
MC-7						1.98	5.54	5.58
MC-8							5.24	5.11
MC-13								4.03

[a] All distance geometry structures have the standard backbone fold corresponding to that shown in Figures 1 and 2. [b] The different folds corresponding to the Monte Carlo structures are shown in the figures. MC-13, which has the standard fold, is shown in Figure 2. The pseudo mirror image fold MC-4 is shown in Figure 3. MC-5 is shown in Figure 4. MC-2 and -6–8 have the same backbone topology, shown in Figure 5. MC-3 is shown in Figure 6.

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

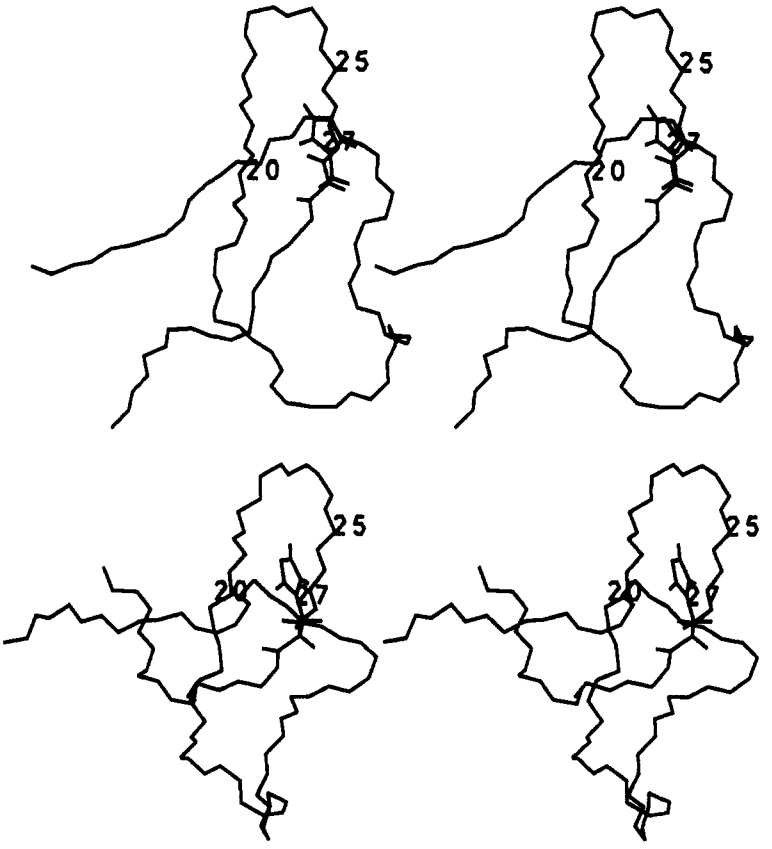

FIGURE 2: Stereoview of the standard fold of NP-5. The structure on top (DG-3) was generated by the distance geometry algorithm. The lower structure, MC-13, was generated by the Monte Carlo procedure discussed in the text. In the orientation shown in the figure, domain 1 (residues 1–10 and 30–33) is above the plane of the paper. The side chains of adjacent residues in the β-hairpin (residues 20–29) alternate above and below the plane of the paper. The side chain H27, which is above the paper plane, is shown for reference.

previously, is shown in Figures 1 and 2. The β-hairpin is oriented vertically in the picture; residues 19–23 form the left fork of the hairpin. In this orientation, when the C^{α} proton of a residue in the hairpin is pointed into the hairpin, the corresponding side chain is pointed into the page. This corresponds to a "right-handed" β-hairpin (Pardi et al., 1988). In the pictures, domain 1 is in front of the β-hairpin and the N-terminal residue is on the left.

The other backbone topologies obtained from the Monte Carlo search contain "pseudo" mirror images of a portion or all of the fold described above. The mirror image is with respect to the qualitative folding pattern and does not imply an exact mirror image, which is not possible because of the chiral centers (L-amino acids). In Figure 3 the Monte Carlo structure MC-4 is shown, which has a backbone topology that is a mirror image of the standard fold. Such mirror image structures were also obtained as solutions to the distance geometry equations. In these structures, the β-hairpin makes a "left-handed turn" (Pardi et al., 1988). Left-handed turns have not been observed experimentally, presumably because

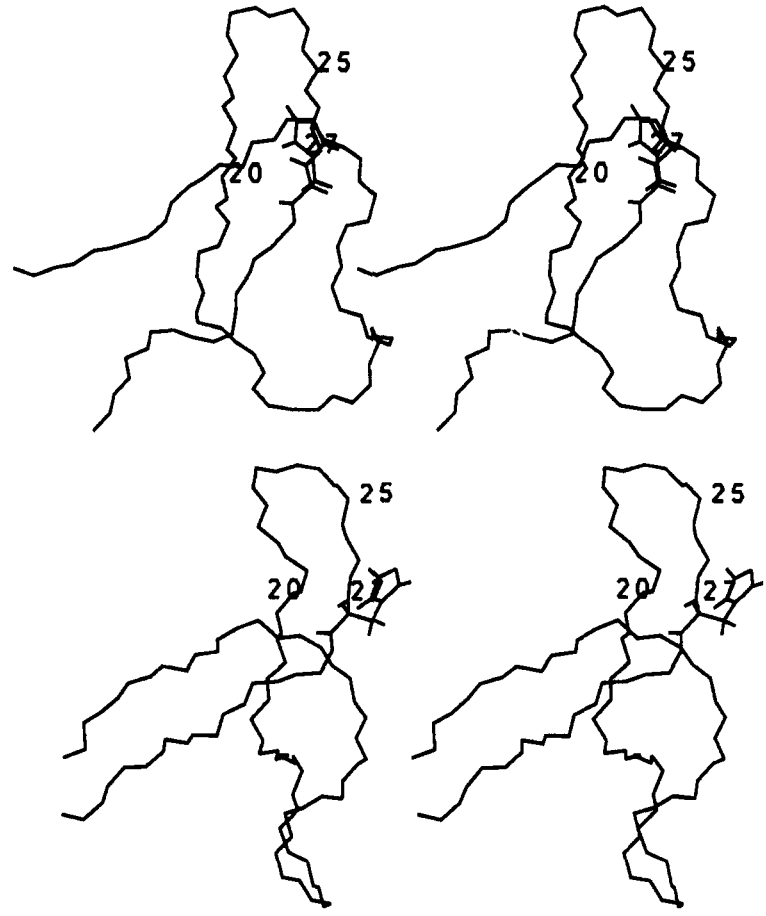

FIGURE 3: Stereoview of a structure with an alternate fold. The structure on the top (DG-3) has the standard fold. The lower structure, MC-4, is a global mirror image of the backbone topology of the standard fold. In the mirror image structure, domain 1 is below the plane of the paper, and the side chains in the β-hairpin have been reflected through the mirror plane; e.g., H27 is below the paper plane.

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

cross-stranded hydrogen bonds cannot be easily formed in this structure with L-amino acids. The structure MC-5 shown in Figure 4 is very similar to MC-4, the complete pseudo mirror image structure (Figure 3), except that the chain from residue 8 to 15 is on the opposite side of the chain segment containing residues 29–33. Thus for MC-4 shown in Figure 3, domain 2 is above the plane containing the β-hairpin, while it is below the plane for MC-5 shown in Figure 4. This illustrates another problem encountered when folding a protein from an extended structure by using distance constraints. A particular piece of the peptide backbone may end up on the wrong side of another part of the chain, and essentially complete unfolding of the molecule is required to correct the relative orientation of the chain segments. The DISMAN algorithm developed by Braun and Gō (1985) attempts to circumvent this problem by allowing atoms (and chains) to pass through one another at specific times in the refinement procedure. A similar procedure has been implemented within IMPACT subsequent to the NP-5 study. One difficulty we have experienced with this method is that structures are generated in which the polypeptide chain passes through aromatic rings; these structures cannot be refined and are discarded.

There are many different ways domains can be assembled in which only a portion of the complete structure forms one or more pseudo mirror images. Four structures (MC-2, 6, 7, and 8) were obtained from the Monte Carlo search which have the same folding pattern for the β-hairpin as the standard structure but for which the rest of the peptide backbone (primarily residues 1–16) forms a pseudo mirror image of the standard fold. Figure 5 shows a stereoview of such a structure. The β-sheet has a right-handed twist as is commonly seen in proteins (Sibanda & Thornton, 1985; Richardson, 1981), and the side chains have an orientation relative to the backbone corresponding to a right-handed turn (Pardi et al., 1988). However, domain 1 is below the plane of the paper in the figure whereas in the standard fold it is above the plane.

In the structure shown in Figure 6 (MC-3), domain 2 is rotated by ~90° with respect to its orientation in the standard fold. The β-hairpin does form a right-handed turn but is somewhat disordered at the base. Also, the fold is similar to the structure shown in Figure 4 in that the chain segment that includes residues 8–15 is on the opposite side of the chain segment containing residues 29–33.

Table II lists the rms deviation among the different structures after superposition. The backbone rms deviations between pairs of structures corresponding to the standard fold generated by the distance geometry procedure are small ($d_{rms} \cong 2.5$ Å). It is of interest to note that rms deviations between pairs of structures that have the standard fold but were generated with the two different procedures is larger ($d_{rms} \cong 4.0$ Å). The rms deviations between pairs of structures with different folding topologies are of course, also larger ($d_{rms} \cong 5$ Å). Thus it is clear that the Monte Carlo search procedure is sampling a much greater region of conformational space than is the distance geometry procedure.

We have also calculated the solvent-accessible surface areas for each of the distance geometry and Monte Carlo structures. The solvent-accessible surface area provides one measure of the compactness of the protein structure. By use of a probe radius of 1.4 Å, the average solvent-accessible surface area for the Monte Carlo structures is 2088 ± 34 Å^2 as compared with an average value of 2336 ± 19 Å^2 for the distance ge-

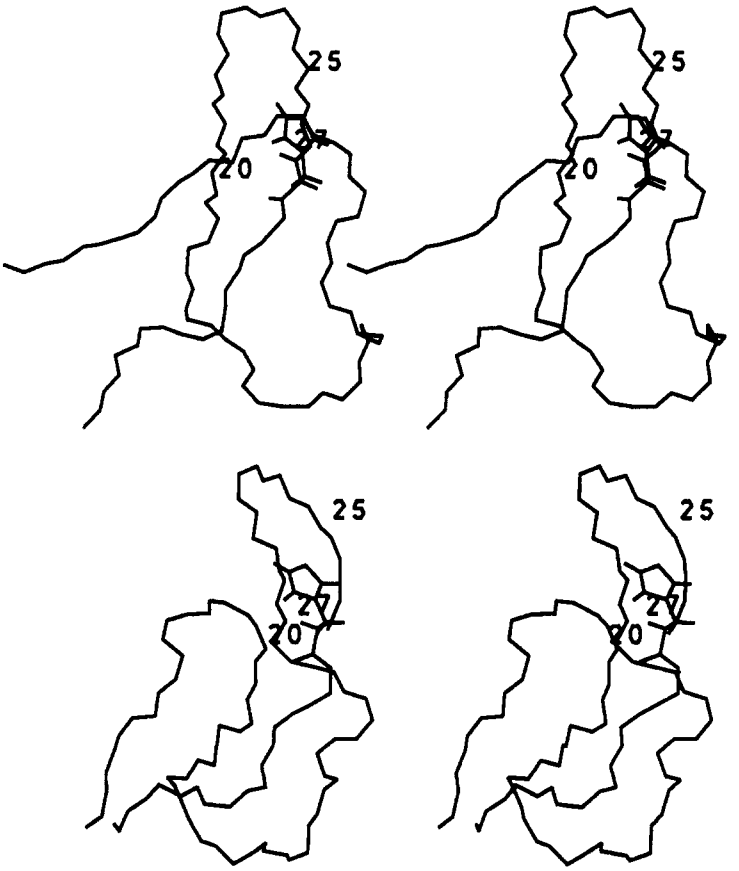

FIGURE 4: Stereoview of a structure with an alternate fold. The structure on the top (DG-3) has the standard fold. The lower structure, MC-5, is similar to the alternate structure shown in Figure 3 except that domain 2 is below the plane of the hairpin.

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

Table III

(A) Molecular Mechanics Energies of NP-5 Structures[a] after Minimization

energy[b]	⟨DG⟩	⟨DG_{min}⟩	DG-5	DG-7	DG-8	MC-13	MC-3	MC-4	MC-5	MC-6
E_{total}	7544.5	−548.2	−478.8	−518.6	−624.7	−469.7	−331.5	−77.3	−210.3	−418.1
E_{bond}	713.5 ± 513.5	13.4 ± 0.9	17.6	11.0	13.0	17.3	29.6	53.9	41.7	21.3
E_{angle}	5726.1 ± 501.4	102.5 ± 6.7	131.4	94.0	90.2	126.9	214.1	325.7	226.6	138.9
E_{ϕ}	182.7 ± 9.1	118.6 ± 5.4	102.2	136.2	118.3	120.8	169.8	171.8	191.3	131.8
E_{LJ}	630.8 ± 72.1	−84.1 ± 3.0	−70.4	−78.8	−93.3	−50.7	−38.4	62.0	11.1	−43.1
E_{rel}	561.9 ± 10.6	−698.6 ± 11.5	−659.6	−681.6	−752.9	−683.8	−706.6	−690.7	−681	−668
rms gr		0.11 ± 0.04	0.04	0.06	0.06	0.06	0.16	0.08	0.04	0.02
d_{viol}[c]	9.8	1.0 ± 0.05	1.2	0.9	0.9	1.4	1.9	1.8	1.6	1.6

(B) Molecular Mechanics Energies[b] of NP-5 Structures after Quenched Molecular Dynamics

energy	DG-7[d]	MC-4[e]	MC-4[d]	MC-6[d]	MC-6[f]	MC-6[g]
E_{total}	−692.4	−296.2	−344.7	−623.4	−620.4	−505.6
E_{bond}	11.5	36.4	38.8	16.1	15.1	41.4
E_{angle}	85.2	222.0	240.7	109	102.5	118.4
E_{ϕ}	94.5	201.0	165.3	166.2	174.9	143
E_{LJ}	−100.3	−5.2	−7.5	−87.1	−96.2	−44.7
E_{el}	−783.3	−750.4	−782.0	−827.6	−816.7	−763.7
rms gr	0.08	0.05	1.30	1.30	0.04	6
d_{viol}[c]	1.1	1.7	1.6	1.1	1.1	0.85

[a] ⟨DG⟩, energies averaged over eight structures before refinement; ⟨DG_{min}⟩, average energies after refinement. DG-5 and DG-8 have the largest and smallest energies of the eight structures generated by the distance geometry procedure. DG-7 has the smallest distance violations of the structures generated by distance geometry. The DG-5, DG-8, and MC-13 structures have the standard fold. [b] Molecular mechanics energies by component in kcal/mol. E_{LJ}, Lennard-Jones energy; E_{el}, electrostatic energy; rms gr, root mean square gradient. [c] Square root of the sum of the squares of the distance violations. [d] 3000 steps of MD at 1000 K, followed by 3000 steps cooling MD to 1 K. [e] 1000 steps of MD at 298 K, followed by 1000 cycles of minimization. [f] 3000 steps MD at 1000 K, followed by 2000 cycles of minimization. [g] 3000 steps of MD, followed by 2000 cycles of minimization with a heavier weight on the distance constraint term.

ometry structures. The tendency of the distance geometry method to generate overly extended structures that are less compact than target structures has been noted in recent tests of the distance geometry method (Metzler et al., 1989).

Distance Violations. Distance violations for representative distance geometry and Monte Carlo structures are included as entries in Table III. The initial distance geometry structures have a large distance violation before refinement because of the replacement of pseudoatoms by protons and the use of tighter constraint ranges in the present calculations. The total distance constraint violations averaged over the eight distance geometry structures is 9.8 Å. The residual distance violations of the distance geometry structures obtained after minimization with a high weight ($W_{constraint}$ = 10) on the distance constraint

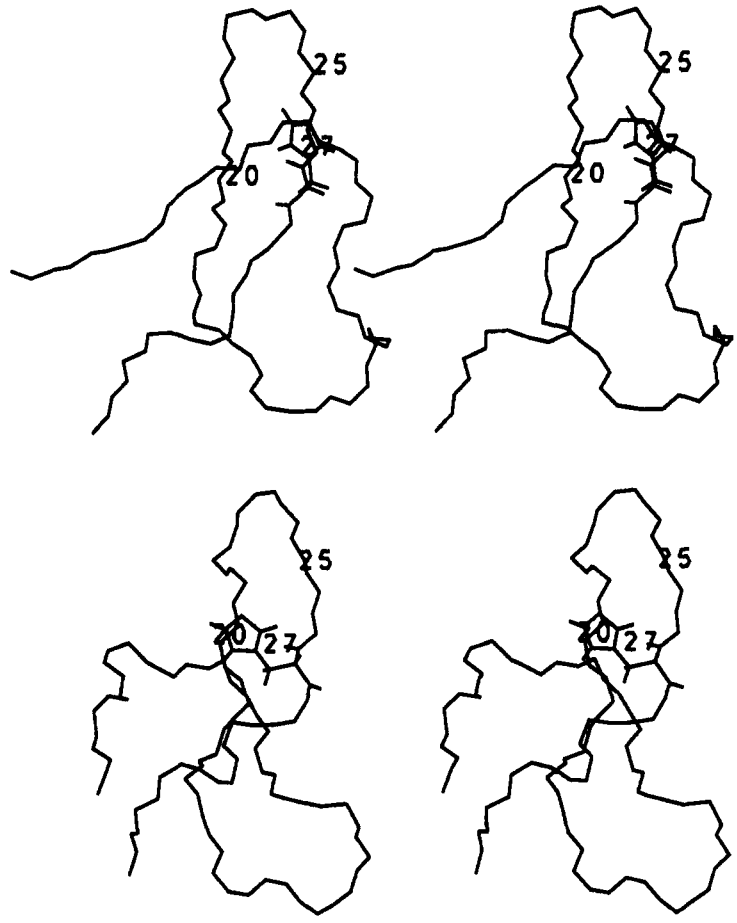

FIGURE 5: Stereoview of a structure with an alternate fold. The structure on the top (DG-3) has the standard fold. The structure on the bottom (MC-6) is a partial pseudo mirror image structure. Four structures (MC-2 and -6–8) were obtained from the Monte Carlo search which have this folding pattern. In this alternate fold, a portion of the peptide backbone (residues 1–16) forms a pseudo mirror image of the standard fold. The orientation of the side chains in the β-hairpin is the same as in the standard fold.

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

term are greatly reduced. The average total violation per structure is reduced to 1.0 Å by this minimization procedure. As noted previously, the use of pseudoatoms in the distance geometry structure generation procedure can lead to structures that are completely consistent with distance constraints used as input in the distance geometry calculations but that are not consistent with the NMR data (Pardi et al., 1988). Thus, with the structure DG-1 as an example, when tighter distance bounds criteria are employed, the structure before minimization has pairs of atoms that violate the distance bounds by more than 3 Å (there are no violations greater than 0.1 Å in this same structure when the larger distance bounds are used). Following minimization using the tighter constraint range however, there are no violations over 1 Å.

The total residual distance violations corresponding to minimized structures with alternative folds are somewhat larger than the values calculated for minimized structures with the standard fold. The total residual distance violations for the structures shown in Figures 3–6 vary between 1.6 and 1.9 Å, whereas for structures with the standard fold, the values range between 0.9 and 1.4 Å. The residual violations are generally reduced by quenched dynamics optimization (see below). The smallest total distance violation calculated for a structure with the standard fold corresponds to 1.05 Å, while for alternative folds the smallest total distance violation, which corresponds to the structure shown in Figure 5, is calculated to be 0.85 Å.

Energetics. The molecular mechanics energies of representative structures corresponding to the different folding topologies are listed in Table III. Structures generated by the distance geometry procedure are highly strained before refinement, especially the bond angles. The large steric strain in the bonds and angles is in part due to the use of a different dictionary for bond lengths and angles used in IMPACT from that used in the distance geometry refinement program. The nonbonded energy of these structures is repulsive, and there is a large repulsive electrostatic component. After minimization of a target function that included both molecular mechanics and distance constraint terms, the bond and angle energies are greatly reduced and the nonbonded energy is attractive.

The minimized molecular mechanics energy of a structure corresponding to the standard fold is considerably lower than that of structures corresponding to any of the alternative folds (−625 kcal/mol for a standard fold structure generated by the distance geometry procedure as compared to −418 kcal/mol corresponding to the partial pseudo mirror image fold of the structure MC-6 shown in Figure 5). However, refinement protocols do have a significant effect on the resulting molecular mechanics energies, and it is difficult to make quantitative comparisons among energies of structures using different protocols. For example, the minimized energy of a structure with the standard fold generated by the Monte Carlo procedure is −469 kcal/mol, which is still lower than the values corresponding to structures with alternative folds, but larger than the corresponding value for a standard fold structure generated by the distance geometry procedure. However, the distance geometry structures had an average total distance violation of 9.8 Å at the start of minimization, whereas the minimization of the Monte Carlo structures was begun when the total

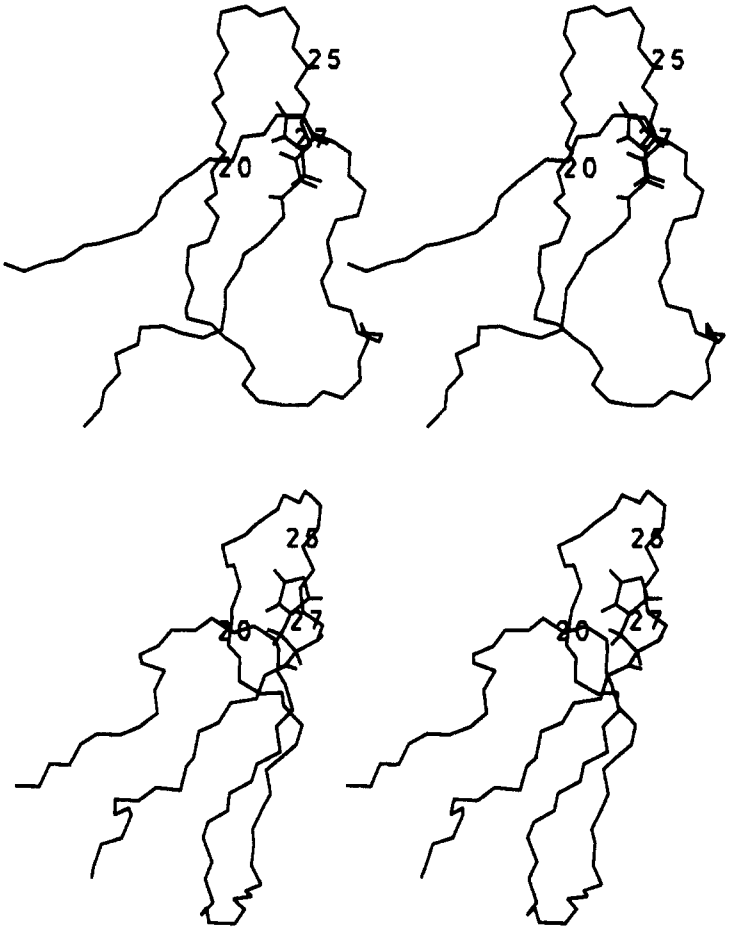

FIGURE 6: Stereoview of a structure with an alternate fold. The structure on the top (DG-3) has the standard fold. In the lower structure (MC-3) domain 2 is rotated by 90° with respect to the standard fold.

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

distance violation was considerably larger (~20 Å). The larger reduction in distance violations during the minimization of structures generated by the Monte Carlo procedure is accompanied by larger steric strain. This reflects the competition between the steric (molecular mechanics) and distance constraint terms during minimization.

The small values of the rms gradients listed in Table III [the gradients are less than 0.1 kcal/(mol·Å)] indicate the minimization of the combined energy–distance penalty function has essentially converged. In order to explore nearby minima in the energy–penalty function surface, three of the minimized structures were subjected to further refinement by constrained molecular dynamics methods. A cycle of heating to 1000 K followed by cooling to 1 K was applied to the structures. The results are listed in Table IIIB. In every case, the molecular dynamics refinement leads to structures with lower total energies and lower total distance violations than the corresponding minimized structures from which the molecular dynamics refinement was initiated. The lowest energy structure, which has the standard fold, has a total energy of −692 kcal/mol and a total distance violation of 1.05 Å. In contrast, the partial pseudo mirror image structure shown in Figure 5 has an energy of −623 kcal/mol after molecular dynamics refinement. While the energy of this structure has decreased by 200 kcal/mol after refinement by the more powerful molecular dynamics procedure, it is still higher in energy than the value obtained by applying the cycle of heating and cooling molecular dynamics to the structure with the standard fold. It should be noted that the strain energy of the bonds, angles, and torsions as well as the Lennard–Jones term are smaller for the structure with the standard fold and thus the total energy is smaller. The electrostatic interactions however, are somewhat more favorable for the alternative fold. In summary, the molecular mechanics studies are consistent with the suggestion that the standard fold is the correct fold, but the results are not clear because the lowest energy structure corresponding to the standard fold and that corresponding to the pseudo mirror image fold shown in Figure 5 differ by only ~10% after molecular dynamics refinement.

Distance Geometry Simulation of an Alternative Fold Target Structure. There have been a variety of studies indicating that the standard metric matrix distance geometry algorithm may not adequately search conformational space (de Vleig et al., 1988; Brünger et al., 1987; Metzler et al., 1989). These studies have primarily concentrated on questions concerning how well the local structure of a molecule (such as a loop in a peptide) is sampled by the algorithm. A more significant question which has not been as adequately addressed is how well the distance geometry algorithm samples all the various peptide folds consistent with the input distance data. To test whether the implementation of the distance geometry algorithm we are using will sample multiple folds consistent with the input distances, we performed distance geometry calculations with simulated input data to see how accurately the algorithm could reproduce one of the alternative folds (the partial pseudo mirror image structure MC-6, shown in Figure 5). This structure was used as the target molecule for the simulation, and the proton–proton experimental NOE interactions used in the original distance geometry calculation were entered with the same level of precision except that the distances were measured from the coordinates of the structure MC-6 (Figure 5). This distance data set was then entered in

the distance geometry program DSPACE (Hare Research, Inc.), and six embedded starting structures were generated in a manner similar to that previously described (Pardi et al., 1988). These embedded structures were then subjected to enough cycles of refinement so that the overall folding of the peptide backbone could be defined. None of these structures possessed the folding pattern of the target structure, but one structure had the standard fold and five structures had the pseudo mirror image of the standard fold (Figure 3).

Discussion

Although enormous progress has been achieved in the determination of solution structures of proteins using information derived from 2D NMR data, the field is still young. Concepts as basic as developing a working definition of the "resolution" of a solution structure remain imprecise. The algorithms that have been published to date which generate a protein structure from a set of structural constraints all have a statistical component to them. Several structures that approximately satisfy the constraints are generated and compared. When the rms deviation among the structures is small, the protein structure is said to be precisely determined. However, this presumes that the set of trial structures represents a complete sampling of all structures that satisfy the input constraints. Practically, it is difficult to determine whether a set of structures does constitute a representative sampling of all conformations that satisfy the constraints. In the present study, NP-5 structures have been generated from NMR data by a new Monte Carlo procedure, and these structures differ in significant respects from those previously reported by using a distance geometry structure generation algorithm. It is clear that the Monte Carlo algorithm is sampling a larger region of conformational space than the implementation of the distance geometry algorithm used to determine the NP-5 structure. Whereas all the distance geometry structures have essentially the same backbone topology, the Monte Carlo search generated structures with several different folds. In a qualitative sense, the folds can be related to each other by mirror plane operations on portions of the molecule. While the problem of distinguishing between symmetry-related structures has been alluded to previously in the context of modeling NMR-derived distance constraints (Braun et al., 1986; Wagner et al., 1987), on the basis of our present experience with NP-5, we believe the problem is potentially greater than previously appreciated. The reason is that there are many different ways domains can be assembled in which only a portion of the complete structure forms a pseudo mirror image. As the size of the protein increases, the combinatorial problem associated with assembling the domains increases dramatically. However, if the number of interdomain constraints also increases rapidly with size, the problem may be less severe. For NP-5, the following information is potentially useful in distinguishing among the various trial structures: (i) consistency of the structures with the measured NMR data, (ii) molecular mechanics energies, and (iii) comparisons of the structures with known folding patterns observed in protein crystal structures. These criteria are discussed below where the standard fold reported previously is shown to be most consistent with all the available information.

The reason it is possible to generate structures that have different folding topologies but small total distance constraint violations is that there are a relatively small number of constraints between domains of the molecule. We have examined the way in which the NOE distance constraints are distributed in NP-5 and the distribution of distance violations in the trial structures. Of the 94 distance constraints used in these calculations, 29 are long range; i.e., they involve protons separated by more than 10 residues in the amino acid sequence. These include NOEs from the N-terminal arm to the C-terminal arm (i.e., within domain 1) and several NOEs between residues in the β-hairpin (domain 3). When these intradomain long-range constraints are excluded from the list, there remain 13 distance constraints between pairs of protons in different domains which are important in determining the orientation of one domain relative to another. These constraints, the corresponding distance ranges, and the upper bounds used in the distance geometry calculations are listed in Table IVA. Table IVB lists the number of residual distance violations of these constraints for NP-5 structures corresponding to each of the different folds. The number of violations are listed corresponding to three different constraint range definitions: (1) pseudoatom distance constraints used in the distance geometry calculations, (2) proton distance constraints used in the Monte Carlo studies, and (3) distance constraints based on the nearest "equivalent" proton. Most reports of solution protein structures are based on methods that use pseudoatoms for which the NOE distance bounds are increased. As is apparent from an analysis of Table IVB, it is important to consider the distance bounds used when comparing the total residual distance violations among different structures. When pseudoatom-based distance bounds are used, total residual distance violations are extremely small for all of the structures corresponding to the different folds. When the tighter distance bounds are used, one of the alternative folds (MC-3 shown in Figure 6) has a large distance violation (1.03 Å) for an interdomain atom pair (T4 H^{α}–I22 H^{δ}). Thus while the total residual distance violation for this structure is only 1.9 Å, it does not consist of many small violations distributed throughout the molecule but rather is localized on an interdomain pair. Such a large violation for an interdomain atom pair provides strong evidence to exclude this fold. The data are not as conclusive for the other folds. The pseudo mirror image structure MC-5 shown in Figure 4 has two interdomain atom pair distance violations between 0.5 and 1.0 Å; however, some of the structures with the standard fold also have an interdomain distance violation this large.

We have also used the shape of the β-hairpin as a criterion to judge the alternative folds. The standard structure has a close to ideal β-hairpin with a right-handed twist and the correct orientation of side chains relative to the plane of the hairpin. The alternative fold corresponding to the structure MC-6 shown in Figure 5 also has a close to ideal β-hairpin. The structures (MC-3–5) corresponding to the three other alternative folds have distorted β-sheets. There are 11 long-range distance constraints involving the hairpin. We have analyzed the residual distance violations corresponding to these constraints. The results are presented in Table IVC. The alternative folds with distorted β-hairpins each have distance violations greater than 0.5 Å. Still it is interesting to note that the distance violations in the distorted hairpins are not as large as might be expected on the basis of visual inspection. In addition to increased distance violations, the side-chain orientations in the distorted structures deviate from patterns observed in crystal structures (Janin et al., 1978).

On the basis of an analysis of interdomain distance violations, the structure of the β-hairpin, and energetics, the most likely alternative topology for NP-5 is the partial pseudo mirror image fold shown in Figure 5 where domain 1 folds behind the β-hairpin in the figure. The "best" structure with this fold (MC-6) has four distance violations less than 0.2 Å and three distance violations between 0.2 and 0.5 Å (Table IVB). This

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

Table IV

(A) Thirteen Interdomain NOE Constraints

constraint	range	upper bound used in DG	constraint	range	upper bound used in DG	constraint	range	upper bound used in DG
16 H^{α} 28 H^{γ}	3.4–5.4	6.4	30 H^{α} 16 H^{α}	2.0–2.8	2.8	6 H^{α} 27 H^{ϵ}	2.6–4.1	4.1
17 HN 28 H^{γ}	3.2–4.7	5.7	4 H^{α} 22 H^{δ}	2.2–3.2	4.2	5 H^{β} 22 H^{δ}	2.3–3.8	6.6
17 H^{β} 29 H^{γ}	3.0–4.5	6.3	7 HN 27 H^{β}	2.6–4.1	5.9	12 H^{β} 29 H^{δ}	2.4–3.9	10.7
14 H^{β} 30 H^{β}	2.1–4.9	6.7	20 HN 4 H^{γ}	2.4–3.9	4.9	5 H^{β} 29 H^{δ}	2.3–3.8	9.6
14 H^{γ} 30 H^{β}	2.0–4.8	6.6						

(B) Residual NOE Distance Violations for 13 Interdomain Atom Pairs

structure	no. of violations per structure in range <0.2	0.2–0.5	0.5–1.0	>1.0	structure	no. of violations per structure in range <0.2	0.2–0.5	0.5–1.0	>1.0
Pseudoatom Distance Bounds									
DG(1–8)[a]	0	0.125	0	0	MC-5	0	0	0	0
DGREF(1–8)[a]	0	0.0	0	0	MC-2 and -6–8	0	0	0	0
DG-3 (best)	0	0	0	0	MC-13	2	0	0	0
MC-3	0	0	0	0	MC-6	0	0	0	0
MC-4	0	0	0	0					
Tighter Bounds/Distance Averaging									
DG(1–8)[a]	0.875	1.75	3.25	1.125	MC-5	5	3	2	0
DGREF(1–8)[a]	6	2	0.125	0	MC-2 and -6–8	4.75	2	3.25	0
DG-3 (best)	1	2	0	0	MC-13	7	0	2	0
MC-3	5	4	0	1	MC-6	4	3	0	0
MC-4	6	6	0	0					
Tighter Bounds/Nearest Proton									
DG(1–8)[a]	1.75	1.125	1.125	0.125	MC-5	1	2	1	0
DGREF(1–8)[a]	1.25	0.375	0	0	MC-2 and -6–8	1.25	0.125	1.5	0
DG-3 (best)	0	0	0	0	MC-13	2	0	0	0
MC-3	1	1	1	0	MC-6	1	0	0	0
MC-4	2	1	0	0					

(C) Residual Distance Violations for 11 Hairpin NOE Distance Constraints

structure	no. of violations per structure in range <0.2	0.2–0.5	0.5–1.0	>1.0	structure	no. of violations per structure in range <0.2	0.2–0.5	0.5–1.0	>1.0
Pseudoatom Distance Bounds									
DG(1–8)[a]	0.125	0	0	0	MC-5	2	0	1	0
DGREF(1–8)[a]	1.75	0.75	0	0	MC-2 and -6–8	2	1.25	0.5	0
DG-3 (best)	1	1	0	0	MC-13	2	1	0	0
MC-3	2	2	2	0	MC-6	3	1	0	0
MC-4	3	1	1	0					
Tighter Bounds/Distance Averaging									
DG(1–8)[a]	0.75	0.625	0.75	0	MC-5	4	1	1	0
DGREF(1–8)	3.25	0.75	0	0	MC-2 and -6–8	5.25	1.5	0.5	0
DG-3 (best)	2	1	0	0	MC-13	6	1	0	0
MC-3	4	3	2	0	MC-6	6	1	0	0
MC-4	5	2	1	0					
Tighter Bounds/Nearest Proton									
DG(1–8)[a]	0.5	0.375	0.125	0	MC-5	2	1	1	0
DGREF(1–8)[a]	1.5	0.875	0	0	MC-2 and -6–8	1.5	1.25	0.5	0
DG-3 (best)	1	1	0	0	MC-13	2	1	0	0
MC-3	2	2	2	0	MC-6	2	1	0	0
MC-4	4	1	1	0					

[a] DG(1–8) is the average of the eight distance geometry structures. DGREF(1–8) is the average of the eight refined distance geometry structures.

is to be compared with the standard fold for which the "best" structure has one violation less than 0.2 Å and two violations between 0.2 and 0.5 Å. It is difficult to rule out the partial pseudo mirror image fold corresponding to the MC-6 structure by using distance violations or energy criteria because the values are relatively close to values calculated for structures with the standard fold. However, when a careful comparison is made between the NOEs predicted from this structure with those actually observed, it is found that, for this pseudo mirror image fold, predicted NOEs which should be very strong are not observed experimentally. This is discussed further below.

Figure 7 shows stereo diagrams of a portion of the backbone and certain side-chain heavy atoms corresponding to the standard fold and the most likely alternative topology, MC-6 (the one for which domain 1 is a pseudo mirror image of the standard fold, see Figure 5). A significant difference between these structures is the location of the N-terminal residues, including residues 4–8, relative to the β-hairpin involving residues 19–29. These two regions of the peptide are covalently connected by the disulfide C5–C20. In both folds the backbone from residues 4 to 7 runs quasi-perpendicular to the long axis of the β-hairpin. Since residue 20, which is on one strand of the β-sheet, is covalently attached to residue 5, there are two possible orientations for the 4–7 backbone. In the orientation corresponding to the standard fold, residue 5 is directly above residue 20 and the portion of the backbone including residues 6 and 7 is directed toward and above residue 27. There is a reasonably strong NOE between the C^{α} proton of R6 and the C^{ϵ} proton of H27. In the alternative fold, for which domain 1 is located below the hairpin, the hairpin is rotated at the base to facilitate the NOE constraint between residues 6 and 27 (see Figure 5). However, in this alternative structure, residue 20 is located directly between residues 6 and 27. For example, the C^{β} proton on C20 is less than 2.5 Å away from both the

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

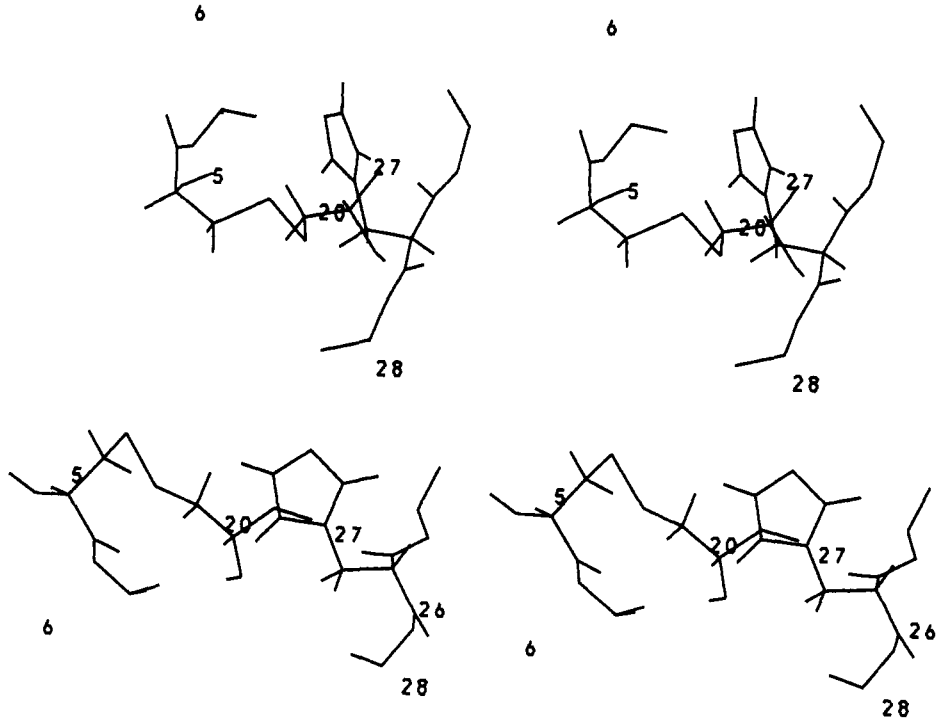

FIGURE 7: Stereoview of a portion of the peptide backbone corresponding to the standard fold (DG-3, top) and an alternate fold (MC-6, bottom). The side-chain atoms are shown for R6, C20, and H27.

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

C^{α} proton of R6 and the C^{ϵ} proton of H27. The C^{β} protons of C20 are well resolved. No NOEs are observed between C20 and either R6 or H27, whereas strong NOEs are predicted to both residues on the basis of the distances measured in the alternative structure.

There is a final piece of complementary information that supports the choice of the standard fold as the correct backbone topology. NMR assignments have been made and NOEs measured for two homologous peptides, rabbit NP-2 and human HNP-1 (Zhang, 1989). Only 11 of 33 residues are conserved in all three molecules, so the resonance assignments and solution structures were generated independently of the NP-5 structure. It is very likely that the overall folds of all the defensins are the same. There are NOEs observed in NP-2 and HNP-1 which are more consistent with the NP-5 standard fold than the most likely pseudo mirror image fold (MC-6). For example, in the human peptide HNP-1, in addition to the analogous NOE from the C^{α} proton of residue 6 to the side chain of W27, W27 also shows an NOE to the side chain of I7. It will be more difficult to satisfy this distance constraint in the pseudo mirror image structure because residue 7 would be further away from residue 27 than residue 6. In contrast, in structures with the standard fold, residue 7 is located right above residue 27 so that these NOE constraints are more easily accommodated.

Conclusions

In this paper we report additional NP-5 structures obtained by using a new structure generation algorithm—Monte Carlo search in torsion angle space. These structures have a large rms backbone deviation from the distance geometry structures (~5.0 Å). The backbone topologies differ in significant respects from the distance geometry structures and from each other. Structures are found that are pseudo mirror images of part or all of the fold corresponding to that first obtained with the distance geometry procedure. Some of the pseudo mirror image structures satisfy the original criteria used to evaluate the quality of the NP-5 solution structures; that is, the total residual distance violations are small, and the β-hairpin has the correct right-handed structure. However, when all the available information is carefully considered, including distance violations of the trial structures, energetics, comparison of predicted NOEs with observed NOEs, and NMR data from homologous peptides, it is concluded that the folding pattern originally reported is most consistent with this information. The present results demonstrate the need for more systematic sampling of conformational space in the structure generation process. From both the calculations reported here and other recent studies (de Vleig et al., 1988; Brünger et al., 1987; Metzler et al., 1989), it appears that while the distance geometry procedure may be computationally faster than alternative methods, it is possible to incorrectly conclude from distance geometry studies alone that a solution conformation has been uniquely determined. The general problem of what requirements the set of structural constraints must satisfy in order to uniquely determine the polypeptide fold, as well as higher resolution information, needs further study. In a forthcoming report, we analyze the accuracy with which the NP-5 side-chain conformations have been determined.

Acknowledgments

The calculations reported here were carried out on a Convex C210 minisupercomputer in the Chemistry Department at Rutgers and on computers at the John Von Neumann Supercomputer Center. We thank the National Allocation Committee of the John Von Neumann Supercomputer Center for the generous award of computer time. We thank Dr. Fumio Hirata for help with the molecular dynamics refinement of the NP-5 structures and Dorothea Kominos for help with the conformational analysis.

Registry No. NP-5, 116110-47-5.

References

Bach, A. C., Selsted, M. E., & Pardi, A. (1987) *Biochemistry 26*, 4389–4397.

Bassolino, D. A., Hirata, F., Kitchen, D., Kominos, D., Pardi,

A., & Levy, R. M. (1988) *Int. J. Supercomput. Appl. 2*, 41-61.
Braun, W., & Go, N. (1985) *J. Mol. Biol. 186*, 611-626.
Braun, W., Wagner, G., Wörgötter, E., Vasak, M., Kägi, J., Wüthrich, K. (1986) *J. Mol. Biol. 187*, 125-129.
Brünger, A. T., Clore, G. M., Gronenborn, A., & Karplus, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3801-3805.
Brünger, A. T., Clore, G. M., Gronenborn, A., & Karplus, M. (1987) *Protein Eng. 1*, 399-406.
Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1985) *J. Mol. Biol. 186*, 435-455.
Crippen, G. M. (1977) *J. Comput. Phys. 24*, 96-107.
de Vlieg, J., Scheek, R. M., van Gunsteren, W. F., Berendsen, H. J. C., Kaptein, R., & Thomason, J. (1988) *Proteins 3*, 209-218.
Havel, T. F., & Wüthrich K. (1985) *J. Mol. Biol. 182*, 281-294.
Janin, J., Wodak, S., Levitt, M., & Maigret, B. (1978) *J. Mol. Biol. 125*, 357-386.
Kline, A. D., Braun, W., & Wüthrich K. (1986) *J. Mol. Biol. 189*, 377-382.
Metzler, W. J., Hare, D. R., & Pardi, A. (1989) *Biochemistry 28*, 7045-7052.
Olejniczak, E. T., Dobson, C. M., Karplus, M., & Levy, R. M. (1984) *J. Am. Chem. Soc. 106*, 1923-1930.
Pardi, A., Hare, D. R., Selsted, M. E., Morrison, R. D., Bassolino, D. A., & Bach, A. C. (1988) *J. Mol. Biol. 201*, 625-636.
Richardson, J. (1981) *Adv. Protein Chem. 34*, 167-339.
Selsted, M. E., Brown, D. M., LeLange, R. T., Harwig, S. S., & Lehrer, R. I. (1985) *J. Biol. Chem. 260*, 4579-4584.
Sibanda, B. L., & Thornton, J. M. (1985) *Nature 316*, 170-174.
Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611-639.
Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295-315.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949-961.
Zhang, X. L. (1989) Ph.D. Thesis, Rutgers University.

Downloaded by RUTGERS UNIV on October 22, 2009 | http://pubs.acs.org
Publication Date: November 1, 1989 | doi: 10.1021/bi00450a017

Determination of the DNA Sugar Pucker Using ^{13}C NMR Spectroscopy†

Rodolfo A. Santos, Pei Tang, and Gerard S. Harbison*
Department of Chemistry, State University of New York, Stony Brook, New York 11794
Received June 6, 1989; Revised Manuscript Received July 18, 1989

ABSTRACT: Solid-state ^{13}C NMR spectroscopy of a series of crystalline nucleosides and nucleotides allows direct measurement of the effect of the deoxyribose ring conformation on the carbon chemical shift. It is found that 3′-endo conformers have 3′ and 5′ chemical shifts significantly (5–10 ppm) upfield of comparable 3′-exo and 2′-endo conformers. The latter two conformers may be distinguished by smaller but still significant differences in the carbon chemical shifts at the C-2′ and C-4′ positions. High-resolution solid-state NMR of three modifications of fibrous calf thymus DNA shows that these trends are maintained in high-molecular-weight DNA and confirms that the major ring pucker in A-DNA is 3′-endo, while both B-DNA and C-DNA are largely 2′-endo. The data show that ^{13}C NMR spectroscopy is a straightforward and useful probe of DNA ring pucker in both solution and the solid state.

Solid-state NMR has recently been added to the panoply of spectroscopic methods available for the determination of the structure of biological macromolecules. Application of solid-state NMR to these problems has usually followed one of two rather different approaches. The first approach uses cross-polarization/magic-angle spinning (CP-MAS) to eliminate the broadening effect of the anisotropic interactions that are intrinsic to NMR in the solid state, thus producing high-resolution spectra of rare spin-$^1/_2$ nuclei in solid or solidlike biological samples. This allows measurement of isotropic chemical shifts, which can then be related to molecular conformation. An example of the successful use of this strategy is the recent determination of the chromophore structure of the membrane protein bacteriorhodopsin largely using solid-state NMR (Harbison et al., 1983, 1984, 1985; deGroot et al., 1989). The second approach utilizes the *anisotropy* of the interactions observable in solid-state NMR spectra to determine the orientation of the principal axis system of the chemical shift, dipolar, or quadrupolar tensors relative to some sample reference frame; since the orientation of the principal axes of these NMR interactions can readily be related to the local molecular frame, this approach can be used directly to determine the orientation and structure of biological materials (Cross & Opella, 1985; Lewis et al., 1985). These two approaches have also been recently combined using a new two-dimensional CP-MAS experiment (Harbison & Spiess, 1986; Tang et al., 1989).

While it has been widely used to investigate other systems, application of solid-state NMR to DNA has been rather limited. ^{31}P studies have been used to observe the structure and dynamics of the phosphodiester backbone of DNA (Shindo et al., 1981, 1985; diVerdi & Opella, 1981; Nall et al., 1981). More recently, deuterium NMR has been used to probe specifically deuterated wet-spun DNA samples, allowing observation of the DNA bases (Vold et al., 1986; Brandes et al., 1988). However, high-resolution solid-state NMR methods have not extensively been applied to DNA. This is curious, given the great success of high-resolution solution NMR studies of DNA oligomers. Proton NMR studies, employing two-dimensional sequential assignment techniques (Feigon et

† Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research. Support was also obtained from the National Institutes of Health (GM-39071) and the National Science Foundation Materials Research Initiative (DMR-8706432).
* Author to whom correspondence should be addressed.

0006-2960/89/0428-9372$01.50/0 © 1989 American Chemical Society