NMR Analysis of Molecular Flexibility in Solution: A New Method for the Study of Complex Distributions of Rapidly Exchanging Conformations. Application to a 13-Residue Peptide with an 8-Residue Loop

D. O. Cicero, G. Barbato, and R. Bazzo*

Contribution from the Istituto di Ricerche di Biologia Molecolare (IRBM) P. Angeletti, via Pontina km 30,600, (00040) Pomezia, Rome, Italy

Received June 3, 1994[⊗]

Abstract: A new methodology, called NAMFIS (NMR analysis of molecular flexibility in solution), is described for the analysis of flexible molecules in solution. Once a complete set of conformations is generated and is able to encompass all the possible states of the molecule that are not *a priori* incompatible with the available experimental NMR evidence, NAMFIS allows for the examination of the occurrence and relevance of arbitrary elements of secondary structure, even when extensive conformational averaging defies a detailed experimental characterization. The analysis is based on the available experimental NMR data.

Introduction

The structure in solution, as determined by NMR,[1] of conformationally flexible molecules, particularly peptides, has been the object of several recent studies.[2−6] From an experimental (NMR) point of view, the problem is in most cases underdetermined, since structures and relative populations of several conformers in fast exchange can hardly be uniquely determined by NMR data only. Therefore, to a certain extent, one has to rely on computational methods. It should also be stressed that one should not expect the general problem to have a unique solution. In this respect, we would regard as a possible solution any given ensemble of molecular conformations, insofar as the latter proves to be compatible with the available experimental evidence. Ideally, an exhaustive analysis should somehow provide all possible solutions.

As a first step, any methodology has to consider the *a priori* generation of all conformations of the molecule that can be present with a population above a given minimum threshold. Theoretical approaches are possible, whereby one attempts an *ab initio* calculation, or else a calculation based on some empirical parameterization, of all molecular arrangements that correspond to relative energy minima.[2,3] One should appreciate that, if the computation of the energy were exact, these methods could indeed provide the unique solution to the problem, since the relative populations of the different conformations also could be directly extracted from their relative energy content. In practice, although many efforts are being spent in this field of research, the level of accuracy that can be presently obtained in such computations is generally considered not sufficient to allow a purely "theoretical" solution to be provided. Therefore, researchers employing a widely accepted view would rather reserve the estimation of relative populations (or molar fractions) of the conformations to a fitting procedure with the experimental data. For the generation of the conformations, the most-used techniques include conformational search procedures, Monte Carlo methods, and molecular dynamics (MD) simulations. In some cases the search is assisted by the application of NMR constraints, mostly derived from NOE data[4,5] and J couplings.[6]

In the end, however, one is usually confronted by a long list of different conformers that are to be considered as potential members of a complex distribution. The challenge of finding out which distribution is the case in the real system is addressed in this paper, and for this purpose we draw on the available NMR experimental data. Clearly, more than one distribution, or indeed a large set of distinct distributions, is likely to be compatible with the NMR data. This is not surprising, since a limited set of experimental measurements is required to somehow constrain the values of a large set of unknown quantities, namely the molar fractions of all the conformers that are potential candidates as members of the distribution. Different approaches to the problem are taken by different authors.[3,7,8] In one case, the theoretical model considers only two conformers in equilibrium, and only the best combination

[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1994.

(1) Abbreviations: ADC, absent distance constraint; Aib, aminoisobutyric acid; CD, circular dichroism; COSY, 2D correlation spectroscopy; cvff, coherent valence forcefield; DC, distance constraint; MD, molecular dynamics; MEDUSA, multiconformational evaluation of distance information using a stochastically constrained minimization algorithm; Mhe, 6-mercapto-1-aminohexanoic acid; NAMFIS, NMR analysis of molecular flexibility in solution; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, 2D nuclear Overhauser effect spectroscopy; QP, quadratic programming; rms, root mean square; SQP, sequential quadratic programming; TFE, trifluoroethanol; TOCSY, 2D total correlation spectroscopy.

(2) (a) Némethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 1883−1887. (b) Dueck, G.; Scheuer, T. *J. Comput. Phys.* **1990**, *90*, 161−175. (c) Morales, L. B.; Garduño-Juárez, R.; Romero, D. *J. Biomol. Str. Dyn.* **1991**, *8*, 721−735.

(3) Nikiforovich, G. V.; Prakash, O.; Gehrig, C. A.; Hruby, V. J. *J. Am. Chem. Soc.* **1993**, *115*, 3399−3406.

(4) (a) Fesik, S. W.; O'Donnell, J. O.; Gampe, R. T., Jr.; Olejniczak, E. T. *J. Am. Chem. Soc.* **1986**, *108*, 3165−3170. (b) Kessler, H.; Griesinger, C.; Lautz, J.; Müller, A.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393−3396. (c) Kim, Y.; Prestegard, J. H. *Biochemistry* **1989**, *28*, 8792−8797. (d) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223−235. (e) Mierke, D. F.; Kurz, M.; Kessler, H. *J. Am. Chem. Soc.* **1994**, *116*, 1042−1049.

(5) Brüschweiler, R.; Blackledge, M.; Ernst, R. R. *J. Biomol. NMR* **1991**, *1*, 3−11.

(6) (a) Torda, A. E.; Brunne, R. M.; Huber, T.; Kessler, H.; van Gunsteren, W. F. *J. Biomol. NMR* **1993**, *3*, 55−66. (b) Mierke, D. F.; Scheek, R. M.; Kessler, H. *Biopolymers* **1994**, *34*, 559−563.

(7) (a) Landis, C.; Allured, V. S. *J. Am. Chem. Soc.* **1991**, *113*, 9493−9499. (b) Yang, J.; Havel, T. F. *J. Biomol. NMR* **1993**, *3*, 355−360.

(8) Blackledge, M. J.; Brüschweiler, R.; Griesinger, C.; Schmidt, J. M.; Xu, P.; Ernst, R. R. *Biochemistry* **1993**, *32*, 10960−10974.

0002-7863/95/1517-1027$09.00/0

in terms of rms deviation is calculated.[8] In another approach, the search of all possible solutions is tackled using a stochastic sampling, but no systematic scan is tried.[3]

Our method for the NMR analysis of molecular flexibility in solution (NAMFIS) attempts to fill in this gap by providing tools that allow for the investigation of the probable population either of a single conformer or of a set of conformers sharing a common feature, like an element of secondary structure, even when extensive conformational averaging defies a detailed characterisation. NAMFIS analytically determines all possible values of the population of a given conformer that are in agreement with the experimental evidence, and performs a classification of such values with respect to the quality of the corresponding data reproduction. On the other hand, no restrictive assumption on the maximum number of conformers to be considered at the same time as significant members of the ensemble is made, since such assumption lacks, in our view, any experimental ground.

Three successive steps can be logically distinguished in NAMFIS: 1. The generation of the "complete" set of molecular conformations that are to be considered as potential members of the ensemble. 2. The determination of their relative populations. 3. The extraction of the structural information by the evaluation of the relative abundance of groups of conformations that share a common structural feature.

The extent to which these tasks can be accomplished by NAMFIS is here illustrated by the application to the conformational study of a 13-residue flexible peptide with an 8-residue cycle closed by a disulfide bridge (**1**). The conformational study is restricted to the backbone of the 8-residue cycle.

The scaffold of the peptide was designed in order to induce an α-helical conformation. The design was based on (i) the use of the strong helix inducer α-aminoisobutyric acid[9] and (ii) the formation of the $i-i+7$ S−S bridge between the L- and D-isomers of 2-amino-6-mercaptohexanoic acid.[10] Additional features include two lysine residues at the C-terminus to improve solubility and prevent intermolecular aggregation, and the capping of the N- and C-termini to prevent unfavorable interactions with the helical dipole.[11] As a feasibility test, the scaffold has been synthesized by introducing in the sequence a combination of residues (Ser,Tyr-Asn,Thr,Ser) that are all known to exhibit a very poor propensity for forming helices. However, a preliminary analysis by CD of peptide **1** in water showed a low helical content. As expected, the use of trifluoroethanol (TFE) as co-solvent (25%) induces an increase in the helical content (around 30%) as estimated by CD. Although the final goal of the design was not accomplished, we decided to study by NMR the backbone conformation of the eight residues within the cycle in the water−TFE mixture, in order to extract indications that may lead to an improvement in the design. For this purpose, the challenge for NMR was to attempt a detailed structural characterization at the atomic level, well beyond the qualitative indications provided by CD. And this research generated NAMFIS.

Ac-Asp-Aib-D-Mhe-Ser-Tyr-Aib-Asn-Thr-Ser-L-Mhe-Lys-Aib-Lys-COOH
S−S

1

Aib, aminoisobutyric acid; Mhe, 2-amino-6-mercaptohexanoic acid

Table 1. Experimental J Coupling Constants and Backbone Interproton Distances of **1** and Associated Errors

	J couplings		interproton distances[a]		
residue	J-(HNC$^{\alpha}$H), Hz	ΔJ-(HNC$^{\alpha}$H), Hz[b]	NOE type	distance (r), Å	Δr, Å[b]
D-Mhe-3	6.6	2.0	NH3−NH4	2.91	0.69
Ser-4	6.4	2.0	NH5−NH6	3.00	0.57
Tyr-5	7.2	2.0	NH6−NH7	3.64	0.87
Asn-7	6.3	2.0	NH9−NH10	3.33	0.81
Thr-8	6.2	2.0	αH3−NH3	2.62	0.42
Ser-9	6.4	2.0	αH3−NH4	2.85	0.36
L-Mhe-10	5.9	2.0	αH3−NH5	4.85	0.66
			αH4−NH4	2.90	0.39
			αH4−NH5	2.76	0.39
			αH4−NH6	4.52	0.57
			αH5−NH5	2.82	0.39
			αH5−NH6	2.56	0.33
			αH5−NH8	4.02	0.75
			αH7−NH7	2.78	0.42
			αH7−NH8	2.90	0.36
			αH8−NH8	2.65	0.39
			αH8−NH9	2.77	0.36
			αH9−NH9	3.12	0.54
			αH9−NH10	2.85	0.36
			αH10−NH10	2.68	0.45
			Har5−Har5(ref)	2.54	

[a] The interproton distances were calculated using the initial rate approximation according to a two-spin model. [b] Total associated error ($\Delta A_i^{exp} + \Delta A_i^{calc}$), see text.

Materials and Methods

Peptide **1** was prepared by Dr. Antonello Pessi (IRBM) using solid phase synthesis. The sample was dissolved in a mixture of water−trifluoroethanol-d_3 (75:25) at a concentration of 2.0 mM. No further increase of the helical content was detected by CD by adding more TFE.

The backbone proton magnetic resonances were assigned by standard COSY, TOCSY, and NOESY experiments on a Bruker AMX-500 MHz spectrometer. NOESY spectra were recorded at five mixing times (70, 100, 150, 180, and 200 ms) to check the linearity of the cross-relaxation buildup. Interproton distances were calculated using the initial rate approximation according to the internal calibration distance between the aromatic protons of Tyr-5. Water suppression was achieved by the application of the jump−return spin−echo scheme proposed by Sklenár and Bax.[12] In addition, four scrambling pulses were introduced after signal acquisition, in order to reproduce the same initial situation of the magnetization for all $t1$ values and decrease the $T1$ noise ridges.

The experimental NMR data are reported in Table 1. For convenience, interproton distances are reported instead of NOEs. Clearly, they represent "virtual" distances because they are determined by the averaging of the corresponding NOEs that are produced by the different conformers in rapid exchange equilibrium.[13]

Conformational Search. In order to generate all the conformations of the backbone that are compatible with the available experimental evidence, the MEDUSA (multiconformational evaluation of distance information using a stochastically constrained minimization algorithm) procedure was applied.[5] The MEDUSA algorithm was implemented as a FORTRAN program that uses the cvff force field of DISCOVER (Biosym). Five starting structures were generated by a 1-ns MD run performed using DISCOVER and subject to constraint-free energy minimization. All the absent distance constraints (ADCs) and the accepted distance constraints (DCs) were introduced as semiparabolic potentials, as described in ref 5. After the introduction of each NMR constraint, 500 cycles of energy minimization using a quasi-Newton−Raphson algorithm (va09a) were performed. The application of the MEDUSA conformational search procedure yielded 800 structures, which were clustered using an rms criterion for the deviations of the

(9) Karle, I. L.; Balaram, P. *Biochemistry* **1990**, *29*, 6747−6756.
(10) Jackson, D. Y.; King, D. S.; Chmielewsky, J.; Singh, S. *J. Am. Chem. Soc.* **1991**, *113*, 9391−9392.
(11) Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 536−537.
(12) Sklenár, V.; Bax, A. *J. Magn. Reson.* **1987**, *74*, 469−479.
(13) Neuhaus, D.; Williamson, M. P. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*; VCH Publishers, Inc.: New York, 1989; pp 170−175.

backbone angles[14] ϕ and ψ with a $f_{threshold} = 5°$. In this way, 71 classes or conformational families were obtained.

Feasible Solution and Feasible Space. The analysis that follows is entirely based on the available experimental parameters A_i^{exp} ($i = 1, n_p$, where n_p is the number of parameters) derived from the experimental NMR data (20 interproton distances and 7 J coupling values, reported in Table 1) and constitutes the NAMFIS method.

From any arbitrary distribution of the molecular population among the short-listed representative conformers, the averaged expected values of NMR parameters like coupling constants and NOEs can be calculated and compared with the corresponding experimental values. For convenience, the averaged NOEs are directly converted into virtual distances via the above-mentioned calibration[13] (see also Table 1).

In order to perform a meaningful comparison, we assign to each parameter A_i (see Table 1) its corresponding error (ΔA_i), by considering not only the experimental source of inaccuracy (ΔA_i^{exp}) like the actual quantitative estimation of NOEs and couplings from the spectra, but also the inherent approximations in the Karplus-type equations,[15] uncertainties in the estimation of internuclear distances and dihedral angles due to the structure simulation algorithm, *etc.* (ΔA_i^{calc}). In Table 1 the global values of the estimated maximum errors ($\Delta A_i = \Delta A_i^{exp} + \Delta A_i^{calc}$) are reported. The value $\Delta J_i = 2.0$ Hz represents the maximum uncertainty in the prediction of the J coupling from the value of the corresponding dihedral angle. Two factors contribute to the values of Δr_i (between 0.33 and 0.87 Å): (i) the experimental error in the determination of the distance from the buildup curves and (ii) the inaccuracy of the calculated distance as a consequence of the uncertainty in the real values of bond angles and bond lengths. Both contributions are different for each distance and depend on the intensity of the NOEs, the number of intervening bonds between the two protons, *etc.* All these contributions have been separately estimated and included in the global Δr_i listed in Table 1.

According to NAMFIS, a given set of values of the molar fractions is considered a *feasible solution* if condition 1 is fulfilled for all detected experimental parameters (virtual distances and coupling constants):

$$A_i^{exp} - \Delta A_i \le A_i^{calc} \le A_i^{exp} + \Delta A_i \qquad i = 1, n_p \tag{1}$$

The ensemble of all feasible sets of molar fractions constitutes what we call the *feasible space* of the variables.

This concept was already introduced by Nikiforovich *et al.*[3] They attempted a statistical exploration of the feasible space by generating a very large number of feasible sets of molar fractions. In our approach, since A_i^{calc} ultimately depends on the molar fractions x_j, conditions 1 are used as a series of constraints for the variables x_j. Two additional linear constraints 2, 3 are contained in their definition:

$$0 \le x_j \qquad j = 1, n_c \tag{2}$$

$$\sum_j x_j = 1.0 \qquad j = 1, n_c \tag{3}$$

where n_c is the number of conformers (in our case $n_c = 71$). Therefore, by minimizing or maximizing a function $F = x_k$ within the constraints 1−3, it is possible to calculate the upper and lower limits of the feasible domain for each molar fraction. Then the complete set of the feasible domains for the molar fractions constitutes an analytical solution for the feasible space, because it encloses all the possible distributions of the conformers that are compatible with the available experimental evidence. Mathematically the problem can be stated as:

$$\text{minimize } F(x) \text{ subject to: } \text{ll} \le \begin{Bmatrix} x \\ \text{LC}x \\ c(x) \end{Bmatrix} \le \text{ul} \tag{4}$$

where LC is an n_L by n constant matrix of linear constraints, and $c(x)$ is an n_c element vector of nonlinear constraints functions and ll and ul are the lower and upper limits, respectively. In our case, $n_L = 1$, LC-$(1,j) = 1$ ($j = 1, n_c$) with ll = ul = 1 (condition 3), and $c_i(x) = A_i^{calc}$ with ll $= A_i^{exp} - \Delta A_i$ and ul $= A_i^{exp} + \Delta A_i$. For this purpose an augmented Lagrangian merit function is defined as:

$$L(x,\lambda,s) = F(x) - \sum_i \lambda_i(c_i(x) - s_i) + {}^1/_2 \sum_i \varrho_i(c_i(x) - s_i)^2 \qquad i = 1, n_p \tag{5}$$

where the summation terms involve all the nonlinear constraints 1. The vector λ is an estimate of the Lagrange multipliers for the nonlinear constraints, and s_i are nonnegative slack variables introduced to allow nonlinear inequality constraints to be treated without introducing discontinuities. The nonnegative vector ϱ constitutes the penalty parameter. It is increased whenever necessary to ensure descent for the merit function. In order to perform this calculation, the E04UCF NAG FORTRAN Library Routine was used. E04UCF applies a sequential quadratic programming (SQP) algorithm in which the search direction is the solution of a quadratic programming (QP) problem. The solution of a QP subproblem provides a vector triple that serves as a direction of search for the three sets of variables (x, λ, s).

Degree of Matching with the Data. In some cases it is only reliable to consider the quantities ($A_i^{exp} - \Delta A_i$) and ($A_i^{exp} + \Delta A_i$) rather than A_i^{exp} itself. This is the case, for example, if we are dealing with a semiquantitative classification into strong/medium/weak NOEs. Whenever more quantitative measurements of interproton distances and coupling constants are accessible for all or some of the parameters, a further characterization of the feasible space is possible. For this purpose, we define the degree of *matching* with the data (M) as follows:

$$M = \prod_i e^{-\{1/2[(A_i^{exp} - A_i^{calc})/\Delta A_i)]^2\}} \qquad i = 1, n_q \tag{6}$$

where n_q is the number of parameters for which a quantitative estimation exists. In a real case it is quite possible that $n_q < n_p$, in which case the total n_p parameters are used to define the constraints 1 but only n_q are used to calculate M. In our case $n_q = n_p$.

Clearly, maximizing M is equivalent to minimizing the sum of the square differences between A_i^{exp} and A_i^{calc}, weighted according to the accuracy of each particular parameter. The matching M is calculated by an algorithm similar to the one described for the calculation of the feasible space. The variables are constrained always within the feasible space by applying the constraints 1−3. Each value of a given molar fraction, within its feasible domain, can then be characterized in terms of the best degree of matching with the data that can be correspondingly obtained. The matching curves reported in the present work are calculated by maximizing M while different values of the molar fraction under study are sequentially assumed and all other molar fractions are treated as variables, within their corresponding feasible domains. When the variable under study is the total molar fraction of a given set of conformations, then the procedure simply introduces a second linear constraint for the corresponding sum of populations. In this case, $n_L = 2$, LC$(2,j) = 1$ (if j belongs to the set under study) and ll = ul = the value of the scanning.

The solution that exhibits the highest value of M (the "best fit" solution) is unique and corresponds to the maximum of all curves. We arbitrarily assigned 100% of matching to this solution. It should be stressed that this solution only represents a single point in the n_c-dimensional feasible space that is defined by the feasible domains of the n_c molar fractions. The complete matching curve can be used to inspect the range of values for each given molar fraction that allows a good reproduction of the data.

Results

From the application of the MEDUSA algorithm to our set of experimental constraints on interproton distances derived from NOEs we obtained a large set of possible conformations (800). After clustering using a geometrical criterion,[14] we ended up with a list of 71 members. This set was assumed to be complete, namely to represent all possible states for the backbone of the eight residues of the cycle under investigation.

(14) Karpen, M. E.; de Haseth, P. L.; Neet, K. E. *Proteins* **1989**, *6*, 155–167.

(15) Pardi, A.; Wagner, G.; Wüthrich, K. *J. Mol. Biol.* **1984**, *180*, 741–751.

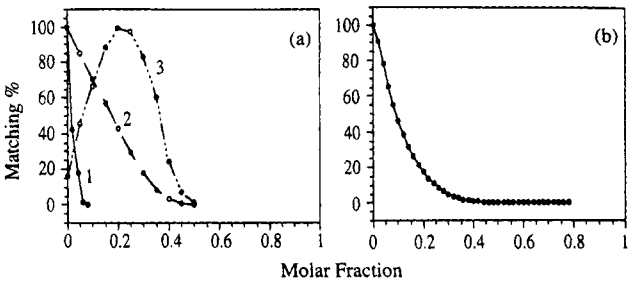

Figure 1. (a) Matching curve types obtained for the 71 molar fractions of conformers of **1**. (b) Matching curve for the sum of the molar fractions of conformers showing type 1 curves.

For the population of each of these 71 conformers, NAMFIS calculated the feasible domain and the corresponding matching function. Three types of matching curves were obtained, as reported in Figure 1a. They illustrate three types of conformers with regard to their abundance in the mixture. Conformers that show curves of type 1 and 2 both present a vanishing molar fraction at the "best fit" solution (the maximum). However, the analysis of the complete matching curve allows us to distinguish two families of conformers. Type 2 conformers (unless type 1) are allowed to have a significant population in solution without dramatically impairing the degree of matching with the data. Notice that the function M represents the best level of matching obtainable for each given value of the population of the particular conformer under examination. Moreover, each curve completely describes the allowed domain (feasible interval) for the corresponding molar fraction. Conformers that show a matching curve of type 3 are clearly among those expected to exhibit a more significant population, although a vanishingly small population could not be ruled out by the analysis. The differentiation among the different types of conformers is self-evident from a simple inspection of the graphs.

Any arbitrary property (or feature) of the molecule can now be examined as regard to its occurrence and relevance, by plotting the matching curve for the sum of the molar fractions of the conformers that have that property in common. Let us, for example, consider the total molar fraction of the conformers whose individual matching curve is of type 1. This is reported in Figure 1b. This set of conformers (54 out of 71) most likely represents less than about 20% of the total. It can be envisaged as a sort of background that is clearly not accessible to detailed experimental investigation.

Figures 2 and 3 summarize the structures and relative populations of the set of 17 conformers that exhibited matching curves of type 2 or 3. The structures were classified and grouped for practical convenience according to the conformation of the stretch of the first four amino acids in the cycle (residues 3−6).

A key criterion is then the definition of a particular structural feature. For instance, we consider a given stretch of residues to be in a helical-type conformation when the corresponding backbone dihedral angles are within the following ranges: (−30°, −90°) for Φ and (−10°, −70°) for ψ. We can then examine the helical content of a given portion of our molecule by plotting the function representing the best matching with the data as a function of the total molar fraction of the corresponding conformers that show in that region a helical conformation, according to the above defined criterion.

At the very first level, we can consider the single residues separately as to their propensity to be helical. In Figure 4 we report the corresponding matching curves for the eight residues within the loop. Clearly the first (D-Mhe-3) and the fourth residue (Aib-6) show the greatest propensity to be helical.

Then, at a more complex structural level, Figure 5 reports the matching curves for the total molar fraction of the conformers with the first and last four residues of the loop, respectively, in a helical-type conformation. The plots show a greater

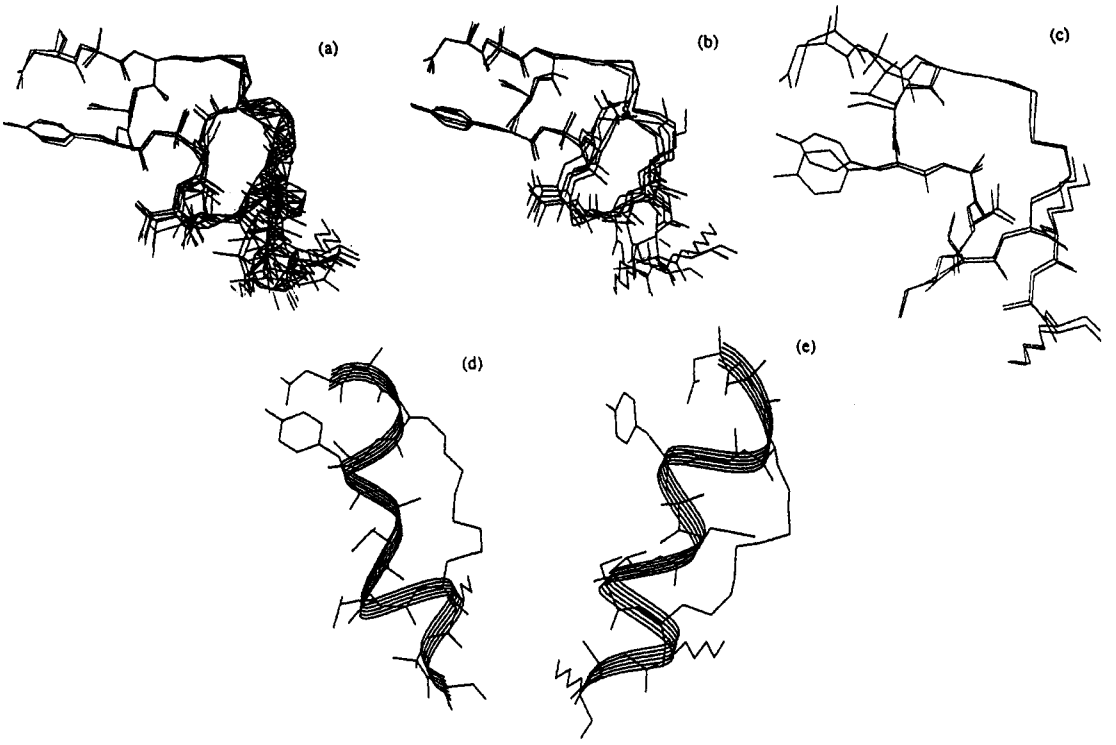

Figure 2. Graphic pictures of the structures of the 17 conformers showing type 2 or 3 matching curves grouped according to the conformation of the D-Mhe-Ser-Tyr-Aib stretch: (a) complete stretch in a helical conformation, 8 conformers; (b) Tyr-5 nonhelical and the rest helical, 5 conformers; (c) Ser-4 and Tyr-5 nonhelical, D-Mhe and Aib, helical, 2 conformers; (d) Ser-4 and Tyr-5 nonhelical, different from (c), D-Mhe and Aib helical, 1 conformer; (e) "extended" helical conformation, 1 conformer. See text for details.

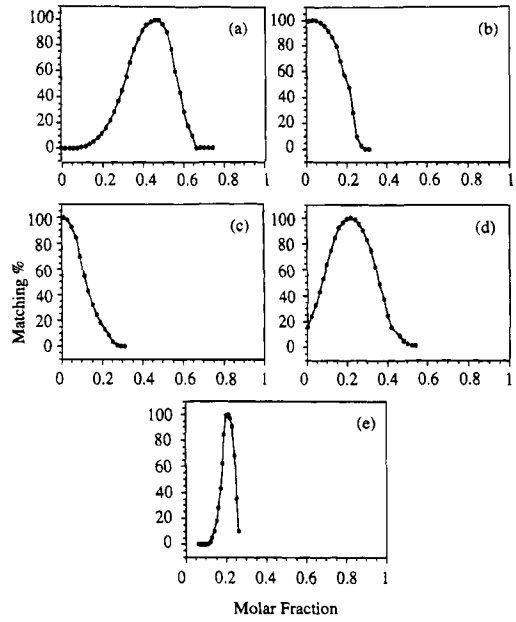

Figure 3. Matching curves for the sum of molar fractions of conformers showing type 2 or 3 curves grouped according to the conformation of the D-Mhe-Ser-Tyr-Aib stretch. The corresponding structures are depicted in Figure 2.

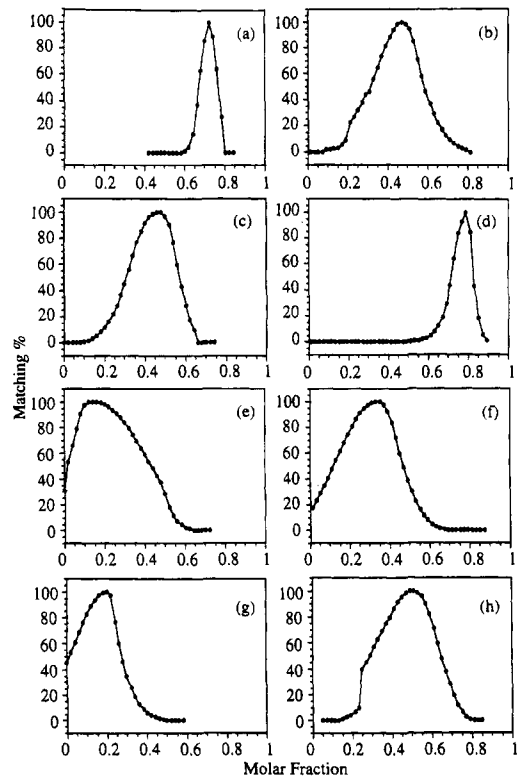

Figure 4. Matching curves for the sum of molar fractions of conformers showing a given amino acid in a helical conformation: (a) D-Mhe-3, (b) Ser-4, (c) Tyr-5, (d) Aib-6, (e) Asn-7, (f) Thr-8, (g) Ser-9, and (h) L-Mhe-10.

propensity of the first four residues to be helical (in about 30–60% of the molecules). The relative abundance of the entire helical conformation can be analogously estimated (Figure 6), and is shown to be less than about 20%.

Discussion

Conformational Analysis of Peptide 1. The matching curves in Figures 1 and 3 show the evidence that peptide **1** exists

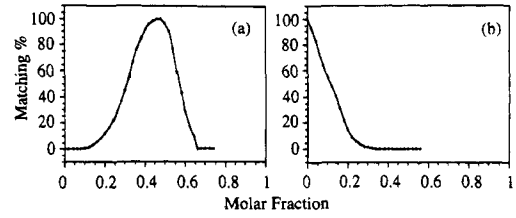

Figure 5. Matching curves for the sum of molar fractions of conformers showing (a) D-Mhe-Ser-Tyr-Aib or (b) Asn-Thr-Ser-L-Mhe in a helical conformation.

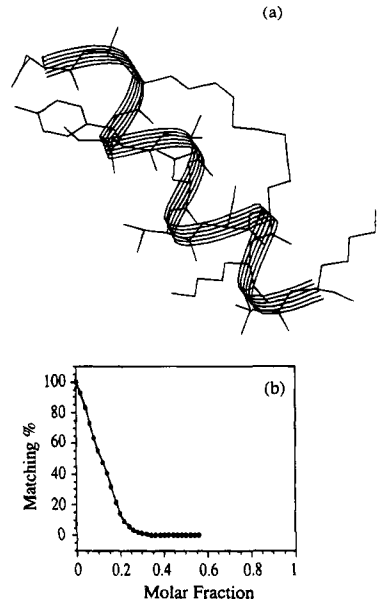

Figure 6. (a) Graphic picture and (b) matching curve for the molar fraction of the conformer that presents all eight amino acids of the cycle in a helical conformation.

in solution as a complex mixture of conformers. From a total of 71 conformers considered, 17 have a higher probability of being significantly populated in solution. These 17 conformers, whose structures are shown in Figure 2, can be combined in many different ways, all providing solutions with high M values. This fact states the limit of any accurate description of the system based on NMR data. However, a comparative analysis of the populations can provide additional information (Figure 3). About half of the molecules in the mixture contain the first part of the cycle (residues 3–6) in a helical conformation (Figure 2a). The other 50% are divided among conformers that present Ser-4 and/or Tyr-5 in nonhelical conformations (Figure 2b, c, and d), along with an extended helical conformer (Figure 2e). The latter is, curiously enough, the only conformer for which the molar fraction is not allowed by the data to be zero. Its population is likely to be between 15 and 25%, but it must be above 5%.

Figure 4 shows the tendency of each single residue to adopt a helical-type geometry. D-Mhe-3 and Aib-6 present the highest helical content. This result is not surprising, as residues 3 and 6 have been introduced in the scaffold as helix inducer residues. Asn-7 and Ser-9 seem to exhibit the least propensity to be in a helix. This conclusion is also experimentally supported by the qualitative NMR exchange data reported in Figure 7. The amide protons of Asn-7 and Ser-9 are clearly the most exposed to the solvent among those in the cycle and therefore the least involved in hydrogen bonds.

Inspection of Figure 5 reveals that the first half of the cycle is more structured than the second half. Moreover, by comparing the matching curves of Figures 4 and 5 one can conclude

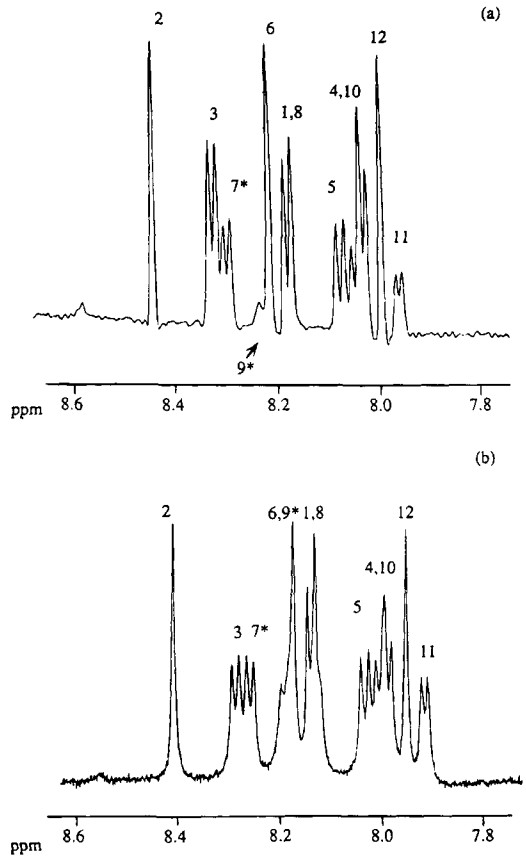

Figure 7. Selected regions of 1D spectra of peptide **1**. (a) Experiment with standard water presaturation. (b) Experiment with jump−return water nonexcitation. The two amide protons most affected by saturation transfer (Asn-7 and Ser-9) are indicated by asterisks.

that the first half of the cycle has an overall helical tendency that is equal to that of the residue within the fragment (Tyr-5) that exhibits the lowest propensity to be helical (see Figure 4c). This description is in agreement with the characteristic cooperativity of the process of helix formation: when Tyr-5 is helical, the whole stretch of the first four residues is likely to be helical. On the other hand, the second half of the cycle, as a whole, exhibits lower tendency to form a helix than the residue that individually shows the lowest tendency to be helical (Ser-9) (see Figure 4g). This result suggests that residues in the second part of the cycle behave independently from one another. Thus the structure turns out to be poorly defined. As a consequence, the conformer that has all 8 amino acids of the cycle in a α-helix conformation can be present only with a population of less (presumably much less) than 20% of the total (Figure 6).

The NAMFIS Methodology. The basic requirement of the NAMFIS analysis is the availability of a set of conformations able to scan all the molecular arrangements that can significantly occur in the real system. Such a requirement is common to any kind of conformational analysis based on the experimental data. However, the novelty of NAMFIS is the capability, through a systematic comparative analysis of the data, of disclosing the "complete" informative content of the experimental measurements. Such information is translated in a series of diagrams, the matching curves, and made available for examination. Yet, the generation of this complete set of molecular conformations remains in the general case a nontrivial problem. For this purpose, different approaches are possible. At the present time, the MEDUSA methodology[5] is, in our view, sufficiently reliable for molecules of size comparable to the one examined in this paper, provided sufficient computer time can be dedicated to an adequate statistical sampling of the conformational space. According to MEDUSA, any given molecular conformation is required to be in agreement with only subsets of NMR distance constraints in order to be considered as a potential candidate. A given "positive" distance constraint (derived from a measured NOE value) is considered to be satisfied if the corresponding distance in the conformer under examination is equal or shorter than the virtual distance (as determined from the NOE). All "negative" constraints, corresponding to vanishingly small or zero NOEs, are converted into lower bounds for the distances and are required to be satisfied by all "allowed" conformations. The rationale behind MEDUSA is to let arbitrary subsets of measured NOEs drive the molecular folding of the possible conformers, by introducing the NOEs randomly one at a time, while checking the consistency of each newly introduced constraint with the previously accepted ones. Clearly, each time MEDUSA needs to allow the molecular energy to relax below the given threshold. This fact raises two problems: one is the demand on computer time, which tends to be very high; the second is more substantial and has to do with the value of the energy threshold. This may not be chosen to be too high so the risk of introducing seriously distorted conformations can be avoided. Such conformations could already represent averaged structures and impair the reliability of the search algorithm. Any positive development in this field would obviously add to the general reliability of NAMFIS itself.

NAMFIS then considers how the total molecular population can be distributed among the calculated conformations. All distributions that are compatible with the data are considered as possible solutions and are examined by NAMFIS as to their capabilities of reproducing the data. It should be noticed that no restrictive assumption is made on the number of conformations that are considered to be present at any one time. The experimental evidence does not support any such *a priori* assumption. The degree of matching with the data is the only criterion, according to NAMFIS, for ranking the different possible distributions, irrespectively from their complexities. In this respect, the argument that the inclusion of more conformations in the distribution is bound to improve the matching with the data clearly does not apply. This is proved by the simple observation (Figures 1 and 4) that the population of most conformations (except only a few out of 71) turns out to be zero in the best fit solution.

The analytical determination of the so-called "feasible space" is a first asset of NAMFIS, in that it provides a simple manner to enclose all the possible distributions of populations. Strictly speaking, such a feasible space is characterized by projecting onto the dimension of each particular molar fraction the multidimensional hypersurface that contains all the allowed combinations of populations. A more detailed characterization in terms of such combinations, although conceivable, could only be obtained with a nontrivial effort *via* statistical methods.[2] However, this is rather impractical due to the conceivably extremely high number of combinations that may be put together without significantly violating the data.

The matching function (M) serves the purpose of measuring the obtainable degree of data reproduction as the population of each conformer is varied within its established limits (feasible domain). In our view, the quantity M may be treated, with some caution, as a probability index, in that it describes the interval that we would consider to enclose, most reliably, the actual value of a given population. It does not seem unreasonable to consider this interval to be narrower than the whole feasible interval,

clearly including the best fit value corresponding to the maximum of the matching curve. How much narrower the interval is, if at all, will depend on the shape and width of the particular matching curve. The undeniable occurrence of some degree of subjectivity in this kind of evaluation demands the application of a corresponding degree of caution.

The results shown in the previous chapter constitute, in our opinion, the maximum precision obtainable in the experimental description of a molecule undergoing extensive conformational averaging. In fact, it is impossible to determine the precise structure of each conformer in equilibrium by relying on averaged experimental information. Clearly, we hope a much better description can be achieved when a single structure proves to exist in solution. For complex mixtures of conformers, it is appropriate to ask questions about, for example, the helical content of the molecule or the propensity of a given residue or stretch of residues to adopt a given predefined conformation. In such cases, NAMFIS constitutes a simple tool for extracting the information encoded in the experimental data set.

On the other hand, we believe that the application of NAMFIS can be most fruitful for the study of systems of intermediate complexity. In the present example, we did not study the conformation of **1** in pure water, as the number of NMR constraints was very low. In this situation the peptide seems to behave as a random coil, in agreement with the CD result. Upon addition of TFE, the values of some coupling constants decreased, and the number of observed NOEs increased. This indicated a certain degree of structure for the peptide, although it remains a very flexible system.

Conclusion

The results shown in this paper prove that a fairly detailed conformational study of the molecule is indeed possible even when a close inspection of the available data set does not seem to provide any relevant detail except for the evidence of a significant conformational averaging (this is, for instance, clearly indicated by the values of the coupling constants, typical of a situation of relative flexibility). The application of NAMFIS, however, can really shed light onto an otherwise extremely vague picture of the molecular structure. It should be appreciated that the picture provided by NAMFIS intentionally remains an experimental picture, only one where all reasonable structural possibilities are considered and classified with regard to their relative capability of reproducing the experimental data. Clearly any further consideration may follow.

Acknowledgment. We are greatly indebted to Dr. Antonello Pessi, Dr. Maria Nicotra, Dr. Elisabetta Bianchi, and Dr. Aaron Garzon for providing peptide **1** and to Dr. Angelo Fontane for preliminary work. We are grateful to Dr. Anna Tramontano for help with computer automatization procedures.

JA941739+