Overlapping Ensemble Dynamics: A Method for Structure Calculations of Multiple Configurational Isomers

Dale F. Mierke

Contribution from the Department of Molecular Pharmacology, Physiology, & Biotechnology, Division of Biology & Medicine, and Department of Chemistry, Brown University, Providence, Rhode Island 02912

Received December 2, 1997. Revised Manuscript Received August 28, 1998

Abstract: A computational approach to utilize the overlapping NOE cross-peak intensities of multiple configurational isomers (e.g., cis/trans isomerization about substituted amide bonds in peptides or proteins) is illustrated. The method, overlap ensemble dynamics, utilizes an ensemble containing the appropriate ratio of the configurational isomers. The NOE intensities, containing contributions from different isomers, are then used as a constraining function to the entire ensemble. In this manner, NOE cross-peaks which before had to be ignored can now be readily used in the computation-based structural refinement. This method should find widespread use given the large number of small molecules of pharmaceutical and medicinal interest with substituted amides and multiple configurational isomers in their NMR spectra.

Introduction

In the transformation of natural peptide drug candidate to therapeutic agent, one of the first steps is alkylation, or similar modification, of the amide position. Peptoids, in which the amino acid side chain function is attached to the backbone nitrogen, are a good example.[1] This modification, carried out to improve the pharmacokinetic or -dynamic properties of the molecule, produces problems with the structural characterization of the molecule, important for establishing a structure−activity relationship. One problem in the NMR-based structural characterization of these molecules is the presence of different configurational isomers (e.g., cis/trans isomers about the substituted amide) leading to multiple sets of NMR resonances.

Traditionally, experimental restraints can only be generated from NOE cross-peaks that are unambiguously assigned[2] and well resolved from the signals of other configurational isomers. Cross-peaks resulting from more than one isomer cannot be utilized since the contribution of the different isomers to the cross-peak intensity is not known (of course, the distances between the protons giving rise to the cross-peak are also unknown). However, it is usually the case that the percent contribution of each of the different isomers is known, determined by careful integration of peaks in a well-resolved region of the ^{1}H NMR spectrum.

Here, an ensemble-based approach, in which an ensemble of molecules is created with the appropriate ratio of the different isomers, is described. During the simulation, forces are applied to the complete ensemble to reproduce the integrated intensity of the overlapping cross-peak. The correct ratio of isomers in the ensemble will account for the uncertainty in the relative contribution of each isomer to the cross-peak: only the distances between the protons of the different isomers, and therefore their contribution to the cross-peak volume, are variables.

This approach is illustrated for a model peptide in which a central proline is in trans and cis configurations. Ensemble calculations are carried out with and without use of the overlapping cross-peaks, and the conformational information that can be gained from the overlapping cross-peaks is highlighted. This approach is only applicable by utilization of an ensemble of molecules in the reproduction of the NMR observables.[3] Ensemble calculations replace the more standard procedure of using one structure for refinement and then repeating the calculation many times; the ensemble method is particularly important for peptide systems undergoing fast conformational dynamics on the NMR time scale.[4]

Experimental Methods

The overlap ensemble method is developed using simulated NMR data for Ac-Ala^1-Phe^2-Pro^3-Ala^4-Leu^5-Ala^6-NH_2. Two conformations were constructed: one containing a trans Phe-Pro amide bond (accounting for 60% of the population) and a cis configurational isomer, accounting for 40%. The two configurational isomers were energy minimized using the Discover force field within the InsightII program (Biosym/MSI, Inc.). The resulting conformation of the Phe^2 residue was −63°,69° for ϕ,ψ for the trans isomer and −70°,−74° for the cis isomer.

* Address correspondence to Department of Molecular Pharmacology, Division of Biology & Medicine, Brown University, Providence, RI 02912. Telephone: (401) 863-2139. Fax: (401) 863-1595. E-mail: dale_mierke@brown.edu.

(1) (a) Simon, R. J.; Kania, R. S.; Zuckermann, R. N.; Huebner, V. D.; Jewell, D. A.; Banville, S. N. S.; Wang, L.; Rosenberg, S.; Marlowe, C. K.; Spellmeyer, D. C.; Tan, R.; Frankel, A. D.; Santi, D. V.; Cohen, F. E.; Bartlett, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992,** *89*, 9367−9371. (b) Zuckermann, R. N.; Martin, F. J.; Spellmeter, D. C.; Stauber, G. B.; Shoemaker, K. R.; Kerr, J. M.; Figliozzi, G. M.; Goff, D. A.; Santi, M. A.; Simon, R. J.; Banville, S. C.; Brown, E. G.; Wang, L.; Richter, L. S.; Moos, W. H. *J. Med. Chem.* **1994**, *37*, 2678−2685.

(2) Progress in using ambiguously assigned NOEs in structure refinement has been reported by Nilges and co-workers. (a) Nilges, M.; Macias, M. J.; O'Donoghue, S. I.; Oschkinat, H. *J. Mol. Biol.* **1997,** *269*, 408−422. (b) Nilges, M. *J. Mol. Biol.* **1995,** *245*, 645−460.

(3) (a) Mierke, D. F.; Scheek, R. M.; Kessler, H. *Biopolymers* **1994**, *34*, 559−563. (b) Mierke, D. F.; Kurz, M.; Kessler, H. *J. Am. Chem. Soc.* **1994,** *116*, 1042−1049. (c) Kenmink, J., van Mierlo, C. P. M.; Scheek, R. M.; Creighton, T. E. *J. Mol. Biol.* **1993,** *230*, 312−322. (d) Scheek, R. M.; Torda, A. E.; Kenmink, J.; van Gunsteren, W. F. In *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance*; Plenum Press: New York, 1991; pp 209−217.

(4) (a) Pellegrini, M.; Mammi, S.; Peggion, E.; Mierke, D. F. *J. Med. Chem.* **1997**, *40*, 89−95. (b) Pellegrini, M.; Royo, M.; Chorev, M.; Mierke, D. F. *J. Peptide Res.* **1997,** *49*, 404−414. (c) Pellegrini, M.; Gobbo, M.; Rocchi, R.; Peggion, E.; Mammi, S.; Mierke, D. F. *Bioploymers* **1996,** *40*, 561−569. (d) Mierke, D. F.; Rudolf-Böhner, S.; Müller, G.; Moroder, L. *Biopolymers* **1995**, *36*, 811−828.

10.1021/ja974090s CCC: $15.00
Published on Web 09/30/1998

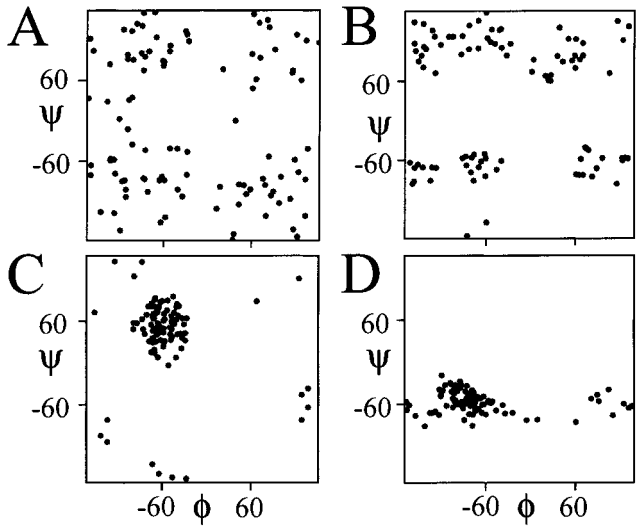

Figure 1. Ramachandran maps for Phe[2] of the peptide Ac-Ala[1]-Phe[2]-Pro[3]-Ala[4]-Leu[5]-Ala[6]-NH_2. The results from standard ensemble-based DG calculations for the trans and cis configurational isomers are shown in panels A and B, respectively. The results from the overlapping ensemble method, utilizing the overlapping cross-peaks of the configurational isomers, are shown in panels C and D, for the trans and cis isomers, respectively.

For the simulation of the NMR data, it was assumed that all of the ^{1}H NMR resonances preceding the proline were overlapping (one set of resonances for Ac-Ala[1]-Phe[2] of both isomers) and therefore all NOEs involving these residues are overlapping. For the remaining residues (Pro[3]-Ala[4]-Leu[5]-Ala[6]-NH_2), the resonances were assumed to be completely resolved and exact distances were calculated from the structures of the trans and cis configurational isomers resulting in 58 and 61 distances, respectively. The distances were adjusted by ±10% to create the upper and lower bounds of the "NOE data".

These simulated data were chosen to mimic realistic situations for peptide and nonpeptidic systems with substituted amides. Often, the amino acids preceding the residue with the substituted amino acid are overlapping, while other portions of the peptide are well resolved. From resolved NMR signals, it is possible to obtain accurate integrated signal intensities (and therefore the ratio of the two isomers). Often there are a significant number of NOEs between the resolved resonances and those which are overlapped.

A holonomeric matrix was calculated for both the trans and cis isomers following standard procedures.[5−7] A variation of 2% was used to allow for flexibility in the bond lengths and bond angles. To the holonomeric matrices was added the "NOE data" from the C-terminal portion of the peptide. Only those "NOEs" which are more restrictive than the holonomeric distances (based on the geometry of the molecule) were utilized: 35 distances (24 interresidue) and 38 distances (28 interresidue) were used for the trans and cis isomers, respectively. One hundred structures of each isomer were calculated using random metrization[7] and refined against these matrices following published procedures (the refinement was run for 5000 steps with a step size of 25 fs with tight coupling[8] to a temperature bath at 300 K, followed by 3000 steps with weak coupling to a temperature bath at 1 K).[3,9]

The overlapping NOEs observed for the N-terminus of the molecule (Ac-Ala-Phe-) would not normally be utilized since the percentage of the cross-peak intensity from the trans and cis isomer is not known. To use the information contained in the volume of the cross-peak, an ensemble of 100 molecules containing the correct ratio of isomers (60 trans and 40 cis) was created from the resulting structures, chosen randomly, from the standard DG methods described above.

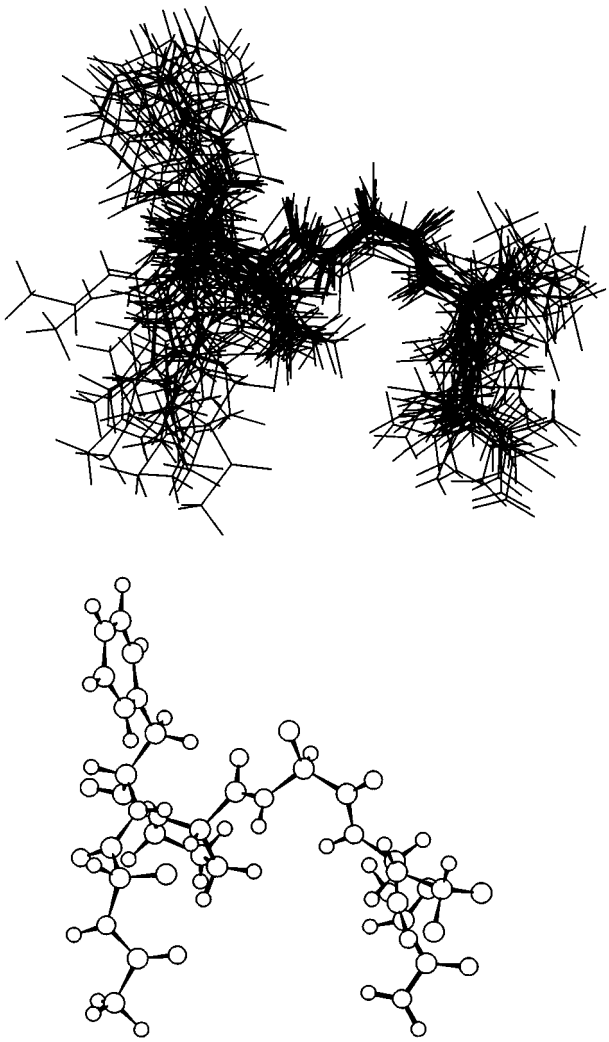

Figure 2. Resulting low-energy conformations from the overlap-ensemble calculations of the cis configurational isomer. The heavy backbone atoms of the Pro-Ala[4]-Leu[5]-Ala[6] have been superimposed. The original conformation used to simulate the NMR data (i.e., target structure) is shown below.

In the standard ensemble method, the NOEs are treated separately from the holonomeric distances. Each member of the ensemble must fulfill the holonomeric distances and, thereby, maintain the correct geometry (e.g., bond lengths, angles) of the molecule. For the NOE restraints, the average over the ensemble is utilized. If the ensemble-averaged distance is too long, a restraining force to shorten the distance is applied to the entire ensemble.[3,4] To these two functions was applied an additional penalty function to force the ensemble to reproduce the volume of the overlapping cross-peak. The ensemble-averaged distances for each isomer are converted to volumes using a reference cross-peak, between methylene protons, with a defined volume (using the isolated two-spin approximation):

$$\sigma_{ij} = \sigma_{\mathrm{ref}}(r_{ij}/r_{\mathrm{ref}})^6$$

A penalty function is then applied to minimize the differences in the ensemble-calculated volume and target volume.[10,11] Currently, a simple harmonic penalty term based on the difference in the volumes is utilized. The analytical function for the cross-peak volume with

(5) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; Research Studies Press Ltd.: Somerset, England; John Wily: New York, 1988.
(6) Havel, T. F. *Biopolymers* **1990,** *29*, 1565−1585.
(7) Havel, T. F. *Prog. Biophys. Mol. Biol.* **1991,** *56*, 43−78.
(8) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A., Haak, J. R. *J. Chem. Phys.* **1984,** *81*, 3684−3690.
(9) Kaptein, R.; Boelens, R.; Scheek, R. M.; van Gunsteren, W. F. *Biochemistry* **1988**, *27*, 5389−5395.
(10) Borgias, B. A.; James, T. L. *J. Magn. Reson*. **1988,** *79*, 493−512.
(11) Stawarz, B.; Genest, M.; Genest, D. *Biopolymers* **1992,** *32*, 633−642.

respect to the atomic coordinates will be incorporated soon.[12] A total of 57 overlapping NOE volumes were calculated for the model system used here, 39 of which were inter-residue.

All distance geometry and ensemble-based refinement calculations were carried out on a Silicon Graphics Indy (R5000, 180 MHz) computer. The refinement of an ensemble of 100 structures of the hexapeptide required approximately 2 CPU hours.

Results and Discussion

The results from the standard DG calculations for the C-terminal portion of the molecule, Pro^3-Ala^4-Leu^5-Ala^6-NH_2, display a tight cluster of conformations, with dihedral angle order parameters[6] for ϕ,ψ ranging in values from 0.8 to 0.95. The N-terminal portion of the molecule shows great conformational freedom with dihedral angle order parameters below 0.20. A Ramachandran map for Phe^2 from these simulations for the trans and cis configurational isomer is given in Figure 1 (panels A and B, respectively). These results are expected given the simulated NOEs for the C-terminus and the complete lack of experimental data, because of the overlapping cross-peaks, for the N-terminus.

The results from the overlapping ensemble procedure are displayed in panels C and D of Figure 1. There is a cluster about the target ϕ,ψ values (recall that the overlapping NOEs were calculated from conformations of $-63°,69°$, and $-70°,-74°$ for the trans and cis isomers, respectively). During the application of the overlapping constraints there is a convergence toward the target structure. In Figure 2, the results from the overlap-ensemble calculations for the cis isomer are illustrated. The general fold of the peptide of the target conformation is well reproduced by the ensemble of molecules.

By creating an ensemble of molecules with the correct ratio of configurational isomers, one can utilize cross-peaks with contributions from different isomers in structural refinement calculations. In this manner, NOE cross-peaks which before had to be ignored, despite containing structurally important information, can now be readily incorporated into structure calculations. This method should find widespread use given the large number of small molecules of pharmaceutical and medicinal interest with substituted amides. For systems in which spin diffusion may be prevalent, a full relaxation matrix for calculation of the cross-peak volumes[10,13−15] could be incorporated. An ensemble-based, full-relaxation refinement protocol has been illustrated in the literature.[16]

Acknowledgment. The author thanks Prof. Christian Griesinger (University of Frankfurt) for useful discussions concerning the ensemble method and problems concerning overlapping NOE signals developed here. Dr. Maria Pellegrini is acknowledged for careful reading of the manuscript and providing many helpful comments.

JA974090S

(12) Yip, P.; Case, D. A. *J. Magn. Reson.* **1989,** *83,* 643−648.

(13) Baleja, J. D.; Moult, J.; Sykes, B. D. *J. Magn. Reson.* **1990,** *87,* 375−384.

(14) Dellwo, M. J.; Wand, J. *J. Biomol. NMR* **1993,** *3,* 205−214.

(15) Bolens, R.; Koning, T. M. G.; van der Marel, G. A.; van Boom, J. H.; Kaptein, R. *J. Magn. Reson.* **1989,** *82,* 290−308.

(16) Bonvin, A. M. J. J.; Rullmann, J. A. C.; Lamerichs, R. M. J. N.; Roelens, R.; Kaptien, R. *Proteins: Struct., Funct. Genet.* **1993,** *15,* 385−400.