

Distinguishing among Structural Ensembles of the GB1 Peptide: REMD Simulations and NMR Experiments

Daniel S. Weinstock, Chitra Narayanan, Anthony K. Felts, Michael Andrec, Ronald M. Levy,*
Kuen-Phon Wu, and Jean Baum*

*BIOMAPS Institute and Department of Chemistry & Chemical Biology, Rutgers University,
610 Taylor Road, Piscataway, New Jersey 08854-8087*

Received October 30, 2006; E-mail: levy@lutece.rutgers.edu; jean.baum@rutgers.edu

Unlike globular proteins with stable native structures, peptides in solution, partially folded proteins, and natively unfolded proteins are best characterized as conformational ensembles of rapidly interconverting structures. Structural characterization of these dynamic systems is critical to understand principles of protein folding and to understand the basis of protein misfolding that results in protein aggregation and disease. NMR provides a method for the structural description of peptides and proteins in solution; however, observables such as NOEs and chemical shifts reflect averages of the properties of individual conformations, and the structural characterization of disordered systems cannot be done using conventional static constraints typically used for globular proteins.

One approach to the characterization of these conformational ensembles is to integrate NMR and computational strategies.^{1a} A number of computational methods have been suggested based on molecular dynamics with time- or ensemble-averaged NMR restraints introduced as an additional term in the effective potential used for structure refinement.^{1a–c} If the effective potentials are accurate enough and the sampling is extensive enough, it should not be necessary to add NMR restraint terms to the potential function to generate structural ensembles which fit the NMR data. However, there have only been a few comparisons published of the results of “unbiased” MD simulations of peptide ensembles and corresponding NMR experiments.^{2a,b} We recently reported the results of replica exchange molecular dynamics (REMD) simulations of GB1 peptide,^{3a,b} in which 20 peptide ensembles were generated as a function of temperature, spanning a range from the most ordered ensemble at 270 K to the most disordered at 690 K. In this communication, we compare each of these structural ensembles with experimental NMR data we have recorded for the peptide. We show that the peptide ensembles in the middle range of temperatures near 400 K provide the best fit to the low temperature (278 K) experimental NMR data, thereby providing a set of models for visualizing the heterogeneity present in the experimental ensemble and also an approach that can be used to help calibrate the effective potential.

Ensembles of peptide conformations were generated using REMD,⁴ the OPLS-AA/AGBNP effective potential⁵ within the IMPACT⁶ molecular simulation package. The current generation of implicit solvent models, including OPLS-AA/AGBNP, is not parametrized to accurately model temperature effects and tends to predict overly structured molecules at low temperature,^{3a} thus leaving open the question of which of the ensembles best matches the experimental chemical shifts, *J*-couplings (Table S1), and NOE NMR parameters (Figure S1) obtained on the GB1 peptide at pH 7.0 and 278 K.

The C-terminal β -hairpin of the B1 domain of G protein is a small dynamic system. This peptide has been shown by NMR and fluorescence measurements to form a β -hairpin in solution.^{7a,b} One measure of the extent of hairpin formation within the peptide

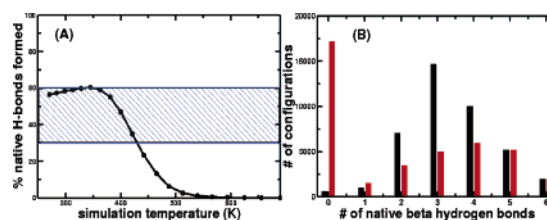


Figure 1. (A) Percentage of native β hydrogen bonds formed relative to the total number possible. (B) Distributions of configurations with 0–6 native β hydrogen bonds in the 270 K (black) and 421 K (red) ensembles.

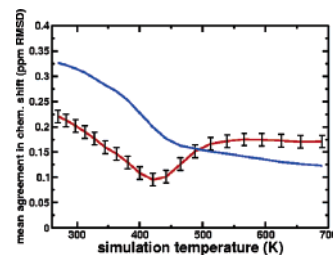


Figure 2. Mean agreement in HA chemical shifts for each simulation ensemble relative to the experimental shifts at 278 K (red curve) and random coil (blue curve). See Supporting Information for details.

ensemble is the percent of native H bonds formed, where native refers to the H bonds formed in the β -hairpin within the intact protein. The REMD ensembles, calculated as a function of temperature, indicate that the system is 35–60% β -hairpin over temperatures ranging from 270 to 421 K (Figure 1a). All of the peptide ensembles below 421 K are consistent with previous estimates of percent β -hairpin derived from NMR chemical shift analysis as well as computer simulations.^{7a,8a–c} However, detailed analysis of two of these ensembles (270 and 421 K) shows that the populations of the six native H bonds within the peptide have very different H-bond distributions (Figure 1b). In the low temperature 270 K ensemble, 99% of the structures have at least one native H bond formed. In contrast, the higher temperature ensemble shows a very heterogeneous distribution of H bond populations, with approximately 50% of the structures having no native H bonds.

To determine which ensembles within the range of temperatures from 270 to 690 K best fit the experimental data, comparisons are made between predicted and measured values of HA (Figure 2) and HN (Figure S2) chemical shifts and J_{HAHN} scalar couplings (Figure S3). At each simulation temperature, chemical shifts and scalar couplings are calculated for each of the residues in each of the 40 000 structures constituting an ensemble using SHIFTX⁹ for the chemical shifts and a Karplus relationship for the *J*-couplings.¹⁰ The HA and HN chemical shift and J_{HAHN} scalar coupling plots all have minima within the range of temperatures from 381 to 442 K, indicating that the simulation ensembles in this middle temperature range are the most similar to the ensemble represented by the NMR

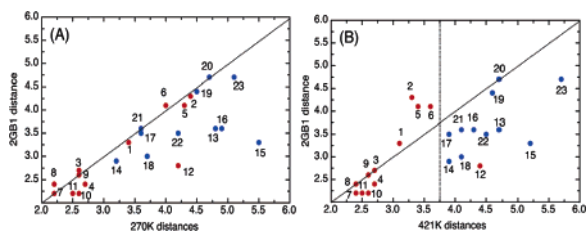


Figure 3. Backbone–backbone NOEs (obtained from the BMRB) observed in the G-protein which are either observed (red) or unobserved (blue) in the GB1 peptide, plotted with the distance in the protein (pdb structure 2gb1) versus the average distance in the low temperature (A, 270 K) and a representative middle range temperature (B, 421 K) simulation ensembles. NOE list is provided in Supporting Information.

data. Additionally, these ensembles are substantially different from those near 270 K and those near 690 K. The correspondence between the simulation ensembles in this middle range and the low temperature experimental results is further supported by comparing the temperature dependence of predicted and experimental HN chemical shifts.^{11a,b} The experimental trends are most closely reproduced by matching the predicted temperature-dependent HN chemical shifts for the ensembles in the middle of the simulation temperature range to the experimental low temperature values (Figure S4), supporting the correlation between the experimental system and the heterogeneous hydrogen bonding pattern described by the simulations.

Comparison of distances derived from NOEs provides another approach for selecting the ensembles which best fit the NMR data at low temperature. The NOEs that are included (Figure 3a,b) are those observed for residues 41–56 in the intact G-protein,¹² with the subset seen in the peptide NMR experiments labeled in red. Distances derived from these NOEs, taken from the NMR structure of the protein, are plotted as a control against predicted interproton distances from two ensembles: one at low temperature (270 K) and one from the middle simulation temperature range (421 K). Predicted NOEs are obtained from the simulations by assuming the members of the ensembles interconvert very rapidly and by averaging $\langle r^{-3} \rangle$ over the ensembles.^{13a,b} At low temperature, most of the points lie along the diagonal, indicating that the conformational ensemble at 270 K is similar to the NMR structure obtained for the β -hairpin region in the G-protein. However, the peptide experimental NOEs observed at 278 K are not consistent with either the NMR structure of the G-protein or the simulated ensemble at low temperature; certain short- and long-range NOEs are observed, but many expected short- and long-range NOEs are missing in an unsystematic way.

In contrast, at 421 K, there is a clear systematic separation of NOEs, with all those observed in the peptide at short ensemble-averaged distances. The NOEs that have moved most significantly relative to the 270 K ensemble are sequential HN–HN NOEs corresponding to points 6 (55HN–56HN), 5 (51HN–52HN), and 2 (45HN–46HN). Neighboring amide protons in the strand of a β -hairpin are constrained to be maximally separated, and any change in conformation will bring these protons closer together. The opposite effect takes place for $\text{HA}_i\text{--HN}_{i+1}$ protons in β -strands, which are constrained to be close. Therefore, the introduction of new non-hairpin populations will significantly affect the predicted distances of the sequential amide protons while not greatly changing the averaged distances between sequential HA–HN pairs due to the $\langle r^{-3} \rangle$ averaging, which is biased toward short distances.

That only short interproton distances are observed in the peptide is due to the fact that the small size of the peptide gives rise to a smaller rotational correlation time relative to the same residues in protein G leading to a reduction in the NOE intensity. This smaller NOE intensity in turn means that, for a given noise level, signals corresponding to long distances are lost in the noise. The one NOE that does not follow this pattern is a cross-strand NOE (54HA–44HN) which is observed even though the protons are separated on average by more than 4 Å in the high temperature ensemble. This reflects the fact that, apparently, the simulation does not align the strands of the hairpin as closely as is observed in the peptide.

In conclusion, detailed comparison of REMD simulations to numerous experimental NMR parameters, including HA, HN chemical shifts, J_{HAHN} scalar coupling, and NOEs has provided an approach for identifying the set of simulated ensembles that best match the low temperature experimental data. These ensembles allow the visualization of distinct populations within the conformational ensemble of the GB1 peptide, thereby providing access not only to averages of conformational properties, such as the number of hydrogen bonds, but also to the distributions which make up those averages. What is striking is that the ensembles that best match the experimental NMR data are very inhomogeneous and the fact that they include a large population of conformations with no native hydrogen bonds.

Acknowledgment. This work has been supported by grants from the NIH (GM30580 and GM45302). We thank David Wishart for providing the SHIFTX source code.

Supporting Information Available: NMR chemical shifts, J -couplings, and NOE parameters; computational details; full ref 6. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) (a) Lindorff-Larsen, K.; Kristjansdottir, S.; Teilmann, K.; Fieber, W.; Dobson, C. M.; Poulsen, F. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 3291. (b) Choy, W.-Y.; Forman-Kay, J. D. *J. Mol. Biol.* **2001**, *308*, 1011. (c) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289. (d) Best, R. B.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 8090. (e) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476.
- (2) (a) Satoh, D.; Shimizu, K.; Nakamura, S.; Terada, T. *FEBS Lett.* **2006**, *590*, 3422. (b) Zagrovic, B.; Sorin, E. J.; Pande, V. *J. Mol. Biol.* **2001**, *313*, 151.
- (3) (a) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins* **2004**, *56*, 310. (b) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6801.
- (4) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
- (5) Gallicchio, E.; Levy, R.; Levy, M. *J. Comput. Chem.* **2004**, *25*, 479.
- (6) Banks, J. L.; et al. *J. Comput. Chem.* **2005**, *26*, 1752.
- (7) (a) Blanco, F. J.; Rivas, G.; Serrano, L. *Struct. Biol.* **1994**, *1*, 584. (b) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
- (8) (a) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068. (b) Garcia, A. E.; Sonbonmatsu, K. Y. *Proteins* **2001**, *42*, 345. (c) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J. Am. Chem. Soc.* **2004**, *126*, 7238.
- (9) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215.
- (10) Lindorff-Larsen, K.; Best, R. B.; Vendruscolo, M. *J. Biomol. NMR* **2005**, *32*, 273.
- (11) (a) Case, D. A. *Curr. Opin. Struct. Biol.* **2000**, *10*, 197. (b) Baxter, N. J.; Williamson, M. P. *J. Biomol. NMR* **1997**, *9*, 359.
- (12) Gronenborn, A. M.; Fipula, D. R.; Essig, N. Z.; Archari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657.
- (13) (a) Olejniczak, E. T.; Dobson, C. M.; Karplus, M.; Levy, R. M. *J. Am. Chem. Soc.* **1984**, *106*, 1923. (b) Kessler, H.; Greisinger, C.; Lautz, J.; Muller, A.; van Gunsteren, W. F.; Berendsen, H. J. *J. Am. Chem. Soc.* **1988**, *110*, 3393.

JA0677517