Dynamics of Unfolded Proteins: Incorporation of Distributions of Correlation Times in the Model Free Analysis of NMR Relaxation Data

Alexei V. Buevich and Jean Baum*

Department of Chemistry, Rutgers University
Piscataway, New Jersey 08854

Received April 1, 1999
Revised Manuscript Received August 4, 1999

Understanding the backbone dynamics of denatured states of proteins can provide valuable insight into protein function, protein diseases such as amyloidoses and prion disease, and the early stages of protein folding.[1] NMR spectroscopy has been widely used for studying protein dynamics of folded proteins on the picosecond to millisecond time scale,[2] and more recently has been used to interpret dynamics of unfolded proteins.[3] However, ^{15}N relaxation data of unfolded proteins are difficult to analyze using standard approaches. For example, it has been recognized that the assumptions of the "model free" analysis of Lipari and Szabo[4] are more suitable for folded proteins and may not be appropriate for unfolded proteins which are composed of an ensemble of states.[3c−f] While studying the backbone dynamics of the unfolded form of pro-peptide of subtilisin (PPS), it became clear that the ^{15}N NMR relaxation data, in particular the R_2 magnetic field independence, could not be fit by all the options provided by the "model free" approach as it stands.[5] It appears that the R_1 and NOE data dominate the minimization process causing large discrepancies between experimental and theoretical R_2 values. This communication presents a novel modification to the "model free" approach that now fits the PPS ^{15}N relaxation data by introducing a distribution of correlation times on the nanosecond time scale to individual sites in the protein.

Nuclear relaxation of backbone ^{15}N nuclei is directly related to the spectral density function, $J(\omega)$, which describes the contribution of different frequencies to the motion of a given nucleus.[6] In the "model free" approach,[4] the spectral density function can be factored into the sum of slow overall molecular reorientation and fast internal motion:

$$J(\omega) = 2/5\{S^2\tau_m/(1 + (\omega\tau_m)^2) + (1 - S^2)\tau/(1 + (\omega\tau)^2)\} \quad (1)$$

where S^2 is the generalized order parameter that reflects the spatial restriction of the N−H bond vector, τ_m is the overall rotational correlation time, $\tau^{-1} = \tau_e^{-1} + \tau_m^{-1}$, and τ_e is the effective correlation time for internal motion.

We propose a variation of the "model free" approach that incorporates the notion that unfolded proteins are composed of an ensemble of states and that characterizes the dynamics of individual sites by a distribution of correlation times on the nanosecond time scale. Specifically, an approach was adopted using the distribution function developed for analysis of R_1 and NOE data of polymers,[7] including Cole−Cole,[8] Fuoss−Kirkwood,[9] and (log−χ^2)[10] distributions. In this report we incorporate the simplest symmetrical distribution function, the Cole−Cole distribution function, into the "model free" approach (termed CC-MF). The Cole−Cole distribution function, $F(s)$, is given by:

$$F(s) = \frac{1}{2\pi}\frac{\sin(\epsilon\pi)}{\cosh(\epsilon s) + \cos(\epsilon\pi)} \quad (2)$$

where $s = \ln(\tau_c/\tau_0)$ and τ_0 and ϵ ($0 < \epsilon < 1$) are the mean value and the width of the distribution function, respectively (Figure 1, inset). The width at half-height of the Cole−Cole distribution is $\Delta s_{1/2} = (2/\epsilon)\cosh^{-1}(2 + \cos(\pi\epsilon))$, so a smaller value of ϵ corresponds to a wider distribution of correlation times and $\epsilon = 1$ corresponds to the single correlation time model characterized by τ_0. The Fourier transformation of the Cole−Cole distribution function gives the spectral density (Figure 1):

$$J(\omega) = \frac{1}{\omega}\frac{\cos\left[\frac{\pi}{2}(1-\epsilon)\right]}{\cosh[\epsilon\ln(\omega\tau_0)] + \sin\left[\frac{\pi}{2}(1-\epsilon)\right]} \quad (3)$$

To fit the R_1, R_2, and NOE ^{15}N relaxation data at three magnetic field strengths, the Cole−Cole spectral density function (eq 3) was incorporated into the "model free" approach resulting in the following functional form of the spectral density function:

$$J(\omega) = \frac{1}{5}\left[\frac{1}{\omega}\frac{S^2\cos\left[\frac{\pi}{2}(1-\epsilon)\right]}{\cosh[\epsilon\ln(\omega\tau_0)] + \sin\left[\frac{\pi}{2}(1-\epsilon)\right]} + \frac{1}{\omega}\frac{(1-S^2)\cos\left[\frac{\pi}{2}(1-\epsilon_e)\right]}{\cosh[\epsilon_e\ln(\omega\tau)] + \sin\left[\frac{\pi}{2}(1-\epsilon_e)\right]}\right] \quad (4)$$

where $\tau^{-1} = \tau_0^{-1} + \tau_e^{-1}$ can be reduced to $\tau = \tau_e$ when $\tau_0 \gg \tau_e$. Equation 4 has two terms and can be further simplified by assuming $\epsilon_e = 1$, indicating a distribution of local "overall" correlation times on the nanosecond time scale. This assumption is supported by the fact that use of the "extended model free approach"[11] with two distinct correlation times to describe the internal motions did not improve the fit of the relaxation data. In addition, no assumption has been made concerning the rate of interconversion within the ensemble of unfolded conformations.

PPS is a 77 residue sequence that is essential for the folding of active subtilisin.[12] The 77 residue intramolecular chaperone is unfolded under physiological conditions and is an example of an

* Corresponding author. Phone: 732/445-5666. FAX: 732/445-5312. E-mail: baum@rutchem.rutgers.edu.

(1) Shortle, D. R. *Curr. Opin. Struct. Biol.* **1996**, *6*, 24−30. Cohen, F. E.; Prusiner, S. B. *Annu. Rew. Biochem.* **1998**, 67, 793−819. Dyson H. J.; Wright P. E. *Nature Struct. Biol.* **1998**, Suppl. 5, 499−503. Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 868−893.

(2) Peng, J. W.; Wagner, G. *Methods Enzymol.* **1994**, 239, 563−596. Palmer, A. G.; Williams, J.; McDermott, A. *J. Phys. Chem.* **1996**, *100*, 13293−13310. Kay, L. E. *Nature Struct. Biol.* **1998**, Suppl. 5, 513−517.

(3) (a) van Mierlo, C. P. M.; Darby, N. J.; Keeler, J.; Neuhaus, D.; Creighton, T. E. *J. Mol. Biol.* **1993**, 229, 1125−1146. (b) Alexandrescu, A. T.; Shortle, D. *J. Mol. Biol.* **1994**, 242, 527−546. (c) Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1997**, *36*, 2390−2402. (d) Frank, M. K.; Clore, G. M.; Gronenborn, A. M. *Protein Sci.* **1995**, 4, 2605−2615. (e) Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *Biochemistry* **1997**, *36*, 13043−13053. (f) Penkett, C. J.; Redfield, C.; Jones, J. A.; Dodd, I.; Hubbard, J.; Smith, R. A. G.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1998**, *37*, 17054−17067.

(4) Lipari, G.; Szabo A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559. Lipari, G.; Szabo A. *J. Am. Chem. Soc.* **1982**, *104*, 4559−4570.

(5) Model free analysis was performed by the Model Free 3.1 program (A. G. Palmer, Columbia University, NY).

(6) Abragam, A. *Principles of Nuclear Magnetism*; Clarendon Press: Oxford, 1961.

(7) Bovey, F. A.; Mirau, P. A. *NMR of polymers*; Academic Press: San Diego, 1996; pp 353−378.

(8) Cole, K. S.; Cole, R. H. *J. Chem. Phys.* **1941**, *9*, 341−351.

(9) Fuoss, R. M.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1941**, *63*, 385−394.

(10) Schaefer, J. *Macromolecules* **1973**, *6*, 882−888.

(11) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1990**, *112*, 4989−4991.

(12) Ikemura, H.; Takagi, H.; Inouye, M. *J. Biol. Chem.* **1987**, 262, 7859−7864. Zhu, X.; Ohta, Y.; Jordan, F.; Inouye, M. *Nature* **1989**, *339*, 483−484.

10.1021/ja9910412 CCC: $18.00
Published on Web 08/31/1999

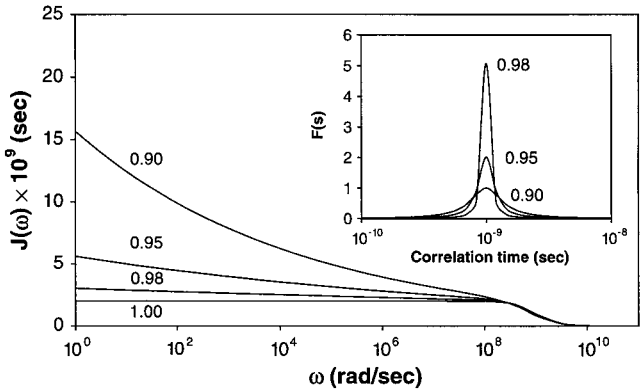

Figure 1. Plot of the Cole−Cole spectral density functions for $\tau_0 = 1$ ns arising from the Cole−Cole distribution function (inset) as a function of the distribution width parameter ϵ. The spectral density function $J(\omega)$ is sensitive to the width of the distribution function, in particular at low frequencies. Because $J(0)$ terms dominate R_2 relaxation rates, the use of distributions of correlation times may give a better fit to the R_2 rates.

important class of proteins that are functional in the unfolded form.[13] The R_1, R_2, and NOE data are quite typical of unfolded proteins with uniform R_1 values ranging from 0.8 to 2.0, more variable R_2 values from 1 to 6 s^{-1}, and NOE values from −3 to 0.2 (Supporting Information). Relaxation data were obtained at three different magnetic field strengths (400, 500, and 600 MHz), and R_2 values showed very little magnetic field dependence, although R_2 variations were significant in many regions of the protein. While analyzing the relaxation data by the CC-MF approach the parameter ϵ_e was kept equal to 1 and parameters τ_0, ϵ, S^2, τ_e, and R_{ex} were optimized separately for each residue by performing a nonlinear minimization of the theoretical R_1, R_2, and NOE values at three fields to the experimental data (Figure 2).

The results of the CC-MF analysis have implications for assessing dynamics in different regions of unfolded PPS. The parameter τ_0 represents the mean value of the correlation time of a specific residue and ϵ represents the width of the distribution of correlation times. The bell shape of the τ_0 curve shows that the terminal regions have smaller values relative to the central regions consistent with less restricted dynamics at the ends than in the center of the protein. The width of the distribution of correlation times, ϵ, arises from the diversity of motions that every residue participates in and reflects the heterogeneity of interactions with other residues in the sequence. As has been described for polymers,[10,14] the motions can be classified into rapid short-range processes and slow longer range processes, suggesting that the greater the disparity in correlation times, the wider the distribution. Overall, the shape of ϵ is anti-correlated with that of τ_0, and in addition, ϵ shows expected sequence-specific variations, with values ranging from 0.94 to 0.99. For example, two neighboring Gly residues (G34 and G35) which have less intramolecular interactions also have relatively large ϵ values of 0.97, corresponding to a narrower width of distribution functions. The average values of S^2 and τ_e are 0.57 ± 0.06 and 74 ± 11, respectively, and appear to be similar to values found in other unfolded proteins in which the traditional "model free" approach was used, suggesting that S^2 and τ_e are less sensitive to the distribution function than τ_0 and R_{ex}. Small R_{ex} terms are applied to residues that have a magnetic field dependence in the R_2 data

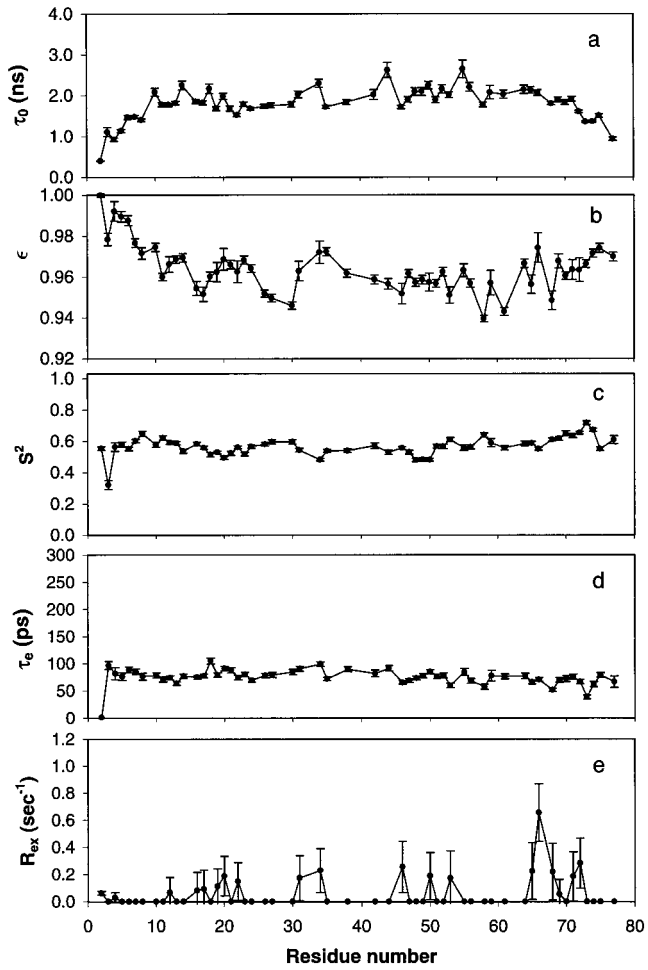

Figure 2. Analysis of ^{15}N relaxation data (R_1, R_2, NOE) of pro-peptide with the Cole−Cole model free approach. ^{15}N relaxation experiments were performed at 9 °C on recombinant, isotope enriched PPS, 0.7 mM, 50 mM sodium phosphate buffer, pH 6.0. The τ_0 correlation times (a), ϵ (b), S^2 (c), τ_e (d), and R_{ex} (e) are shown for each residue.

suggesting that the major component of the R_2 variability in unfolded proteins arises from a distribution of correlation times on the nanosecond time scale.

In conclusion, the analysis of the PPS unfolded data shows that a statistical distribution model of the "model free" approach can better explain the observed ^{15}N relaxation data. This method, which includes a distribution of correlation times on the nanosecond time scale, is likely to apply to other unfolded proteins which similarly show little R_2 dependence in magnetic field strength.[3d,e] Describing site-specific residues of unfolded proteins in terms of distributions of correlation times will help shed new light on the nature of the unfolded state by providing a new framework for thinking about the time scales of motions and the types of correlated motions that are prominent in unfolded proteins.

Acknowledgment. We would like to acknowledge helpful discussions with R. M. Levy and B. Brodsky. This work was supported by NIH Grant GM45304.

Supporting Information Available: A table containing ^{15}N R_1, R_2, and NOE data for PPS measured at 400, 500, and 600 MHz, a table with the Cole−Cole model free analysis of ^{15}N relaxation data, and a figure with experimental relaxation data and their fit obtained by the Cole−Cole model free analysis (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA9910412

(13) Plaxco, K. W.; Gross, M. *Nature* **1997**, *386*, 657−659. Daughdrill, G. W.; Hanely, L. J.; Dahlquist, F. W. *Biochemistry* **1998**, *37*, 1076−1082. Zhang O.; Forman-Kay, J. D. *Biochemistry* **1997**, *36*, 3959−3970. Penkett, C. J.; Redfield, C.; Dodd, I.; Hubbard, J.; McBay, D. L.; Mossakowska, D. E.; Smith, R. A. G.; Dobson, C. M.; Smith, L. J. *J. Mol. Biol.* **1997**, *274*, 152−159. Radhakrishnan, I.; Perez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Cell* **1997**, *91*, 741−752.

(14) Heatley, F.; Begum, A. *Polymer* **1976**, *17*, 399−408.