Paramagnetic relaxation enhancements in unfolded proteins: Theory and application to drkN SH3 domain

Yi Xue, Ivan S. Podkorytov, D. Krishna Rao, Nathan Benjamin, Honglei Sun, and Nikolai R. Skrynnikov*

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

Received 28 January 2009; Revised 7 April 2009; Accepted 25 April 2009
DOI: 10.1002/pro.153
Published online 6 May 2009 proteinscience.org

Abstract: Site-directed spin labeling in combination with paramagnetic relaxation enhancement (PRE) measurements is one of the most promising techniques for studying unfolded proteins. Since the pioneering work of Gillespie and Shortle (J Mol Biol 1997;268:158), PRE data from unfolded proteins have been interpreted using the theory that was originally developed for *rotational* spin relaxation. At the same time, it can be readily recognized that the relative motion of the paramagnetic tag attached to the peptide chain and the reporter spin such as $^1H^N$ is best described as a *translation*. With this notion in mind, we developed a number of models for the PRE effect in unfolded proteins: (i) mutual diffusion of the two tethered spheres, (ii) mutual diffusion of the two tethered spheres subject to a harmonic potential, (iii) mutual diffusion of the two tethered spheres subject to a simulated mean-force potential (Smoluchowski equation); (iv) explicit-atom molecular dynamics simulation. The new models were used to predict the dependences of the PRE rates on the $^1H^N$ residue number and static magnetic field strength; the results are appreciably different from the Gillespie–Shortle model. At the same time, the Gillespie–Shortle approach is expected to be generally adequate if the goal is to reconstruct the distance distributions between $^1H^N$ spins and the paramagnetic center (provided that the characteristic correlation time is known with a reasonable accuracy). The theory has been tested by measuring the PRE rates in three spin-labeled mutants of the drkN SH3 domain in 2*M* guanidinium chloride. Two modifications introduced into the measurement scheme—using a reference compound to calibrate the signals from the two samples (oxidized and reduced) and using peak volumes instead of intensities to determine the PRE rates—lead to a substantial improvement in the quality of data. The PRE data from the denatured drkN SH3 are mostly consistent with the model of moderately expanded random-coil protein, although part of the data point toward a more compact structure (local hydrophobic cluster). At the same time, the radius of gyration reported by Choy *et al*. (J Mol Biol 2002;316:101) suggests that the protein is highly expanded. This seemingly contradictory evidence can be reconciled if one assumes that denatured drkN SH3 forms a conformational ensemble that is dominated by extended conformations, yet also contains compact (collapsed) species. Such behavior is apparently more complex than predicted by the model of a random-coil protein in good solvent/poor solvent.

Additional Supporting Information may be found in the online version of this article.

Abbreviations: drkN SH3, N-terminal SH3 domain of the *Drosophila* adapter protein drk; EDTA, ethylenediaminetetraacetic acid; ESR, electron spin resonance; FRET, fluorescence resonance energy transfer; MTSL, methanethiosulfonate spin label; NOE, nuclear Overhauser effect; NAG, *N*-acetyl-glycine.

Yi Xue and Ivan Podkorytov contributed equally to this work

Grant sponsor: NSF CAREER Grant; Grant number: 044563.

**Correspondence to:* Nikolai R. Skrynnikov, Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47907-2084. E-mail: nikolai@purdue.edu

Keywords: site-directed spin labeling; paramagnetic relaxation enhancement; unfolded proteins; segmental diffusion; translational relaxation; Smoluchowski equation; molecular dynamics; drkN SH3; ubiquitin

Introduction

All proteins begin their life cycle in the disordered state. The majority is subsequently folded in a chaperone-aided folding process. A substantial fraction remains disordered and fulfills its biological function in disordered state.[1,2] In both situations, correct operation of the cell machinery depends on its ability to differentiate between different unfolded proteins. This ability, in turn, depends on the distinctive structural features of specific disordered proteins.

For disordered proteins, the concept of structure is understood in terms of probabilistic preferences. The disordered protein can be viewed as a broad ensemble of interconverting conformational species, where some of the species occur with higher probability than the others. Often, the discussion is framed in terms of "residual structure."[3–5] In defining the residual structure, one makes a tacit reference to the idealized random coil—it is assumed that random coil possesses no structure; hence, any conformational preferences that distinguish the unfolded protein from the random coil are classified as residual structure.

Among many potential conformational preferences, most prominent are the tendencies to form hydrophobic clusters and prototypical secondary structure elements (the two tendencies appear to go hand-in-hand).[6,7] The convergence of these tentatively structured regions eventually leads to a formation of a molten globule, a state with an increased amount of structural order.[8] It should be emphasized however that a broad class of unfolded proteins behave essentially as random-coil polymers with only modest propensity for residual structure. This is put into evidence, for example, by the data on hydrodynamic radii and radii of gyration of chemically denatured proteins.[9–11] The random-coil-like behavior of unfolded proteins has been sometimes underappreciated, since much (deserved) attention has been given to the investigation of residual structure. It is this category of proteins, where disorder prevails over order, which is the focus of the present article.

Spectroscopic methods that are best suited for structural characterization of folded proteins tend to falter in the case of unfolded proteins. Clearly, high-resolution X-ray crystallography is impossible in the intrinsically disordered systems. NMR spectroscopy fares much better, as proton exchange,[12] chemical shifts,[13] heteronuclear relaxation rates,[14] scalar couplings,[15] and residual dipolar couplings[16] have all been used to obtain valuable information about unfolded proteins. However, NOE experiments, which provide the bulk of structural information in NMR studies of folded proteins, prove to be of limited value. Indeed, very few long-range contacts can be identified in unfolded proteins.[17–20] The fundamental reason for this is that in highly fluid systems the contacts between distal protons are too short-lived to produce a substantial NOE transfer. In this situation, to detect the elements of the long-range order, one has to rely on interactions that are stronger than $^1H—^1H$ magnetic dipolar interaction.

A number of experimental techniques exist that can produce long-range site-specific constraints for disordered protein ensembles: electron transfer,[21] disulfide bond formation,[22] FRET,[23] ESR,[24] and NMR paramagnetic relaxation enhancement (PRE) experiments.[25] The PRE technique is particularly attractive: (i) it provides a multitude of constraints from a single protein sample, (ii) relatively compact paramagnetic labels can be placed at different sites along the polypeptide chain with minimum perturbation to the system, and (iii) the results can be quantitatively interpreted since the process of nuclear spin relaxation is generally well understood.

In PRE studies of unfolded proteins, site mutagenesis is typically used to produce a protein with a single cysteine residue in a desired position. This cysteine is then tagged with thiol-reactive nitroxide label. In principle, there are many other ways to introduce paramagnetic centers into the system—through native or engineered metal-binding sites, peptide synthesis involving nonnatural amino acids, contrast agent molecules added to the solvent, and so forth. In this article, however, we concentrate on the most widely used thiol-reactive methanethiosulfonate spin label (MTSL).[25–27] The nitroxide label causes paramagnetic dipolar relaxation of nuclear spins; the corresponding PRE values can be translated into the effective distances between the paramagnetic center and the reporter spins (typically, amide protons). In this manner, one can detect the tendency of the peptide chain to form certain specific contacts. By scanning a series of samples with different placement of the MTSL label, a fairly detailed picture of the residual structure can be obtained.[25,28–32]

Although PRE is arguably the most valuable experimental technique for structural characterization of unfolded proteins, the proper theoretical treatment of this effect is conspicuously lacking. To understand the origin of PRE in the unfolded protein, it is necessary to consider various sources of stochastic dynamics in the system. The dipolar interaction between the unpaired electron spin of MTSL (S) and the backbone amide $^1H^N$ spin (I) is modulated by: (i) reorientation

of the *IS* vector in the laboratory frame of reference, (ii) variation of the distance r_{IS}, and (iii) stochastic transitions of the electron spin *S*. The latter process is much slower than the first two. Specifically, the correlation time relevant for the standard PRE experiment is the electron spin-lattice relaxation time $T_{1,S}$.[33] The ESR studies of the protein-bound MTSL label conducted at room temperature and static magnetic field 0.3 T report $T_{1,S}$ values on the order of several microseconds.[34,35] At higher magnetic field, $T_{1,S}$ becomes even longer.[36,37] This process, therefore, is several orders of magnitude slower than the segmental motion of the unfolded polypeptide chain (motional modes (i) and (ii), characteristic time ~1 ns). Consequently, electron spin relaxation has no bearing on paramagnetic relaxation of nuclear spins.

Under these circumstances, the PRE effect is determined by the motional modes (i) and (ii). The theoretical treatment of this situation was first formulated in the pioneering work by Gillespie and Shortle.[25,28] In effect, their treatment is equivalent to the model of a folded protein undergoing slow conformational exchange (while maintaining roughly the same overall shape). In this model, paramagnetic relaxation is controlled by the overall tumbling correlation time, τ_{rot}, which is uniform across the protein. On the other hand, the distances r_{IS} vary on a time scale that is much longer than τ_{rot} and, therefore, enter in the expressions for the paramagnetic relaxation rates as $\langle r_{IS}^{-6}\rangle$. Following the work by Gillespie and Shortle, much effort was put into modeling of $\langle r_{IS}^{-6}\rangle$ for structural ensembles representing unfolded proteins and predicting the relevant probability distributions.[30,38–40]

While the Gillespie–Shortle model is, perhaps, relevant for the case of the compact molten globule with reduced internal mobility (approaching the folded protein limit), it does not fit very well the case where the protein is strongly disordered and behaves essentially as a random coil. Indeed, in this case, the relative motion of the paramagnetic center and the amide $^1H^N$ is best described as relative translation and the conventional formulas for rotational relaxation do not apply. As the polypeptide chain twists and turns, both the orientation of the dipolar vector and its length fluctuate. In this situation, both motional modes (i) and (ii) are relaxation-active. Clearly, the relative translational motion of the two spins is subject to constraints—they cannot come closer than allowed by the van der Waals forces and cannot move farther apart than the length of the chain that connects them. Generally speaking, there is a probability distribution associated with the *IS* vector, $P(r_{IS})$, that can be taken into consideration by introducing the appropriate mean-force potential. Thus, to properly model the PRE effect in the unfolded proteins, one needs to solve the problem of translational diffusion in restricting potential. This problem is closely related to the one encountered in the studies of intermolecular paramagnetic relaxation.[41,42]

Gillespie and Shortle acknowledged the tentative character of their model, pointing out that "its application to a denatured protein ... raises a number of technical issues and dilemmas that cannot be fully resolved."[25] Furthermore, they allowed for a generous uncertainty margin, ±5 Å, for the effective distances $\langle r_{IS}^{-6}\rangle^{-1/6}$ extracted from the experimental PRE data. A margin of that size is likely to absorb the error arising from the crudeness of the model. The loose character of the PRE-derived restraints was also stressed by the other investigators.[38] Nevertheless, it is clearly desirable to construct a theory that would provide a more sound physical and mathematical description of the PRE effect in the unfolded proteins. Such theory can improve the accuracy of the PRE-derived restraints and clarify the limits of applicability of the Gillespie–Shortle model. In this article, we seek to develop this kind of theoretical framework.

The article is structured as follows. In the "Theory" part, we first present an overview of the PRE effect in unfolded proteins. In the second section, the Gillespie–Shortle model is briefly reviewed. In the third section, we present the analytical results for the model of two spheres diffusing in a square-well potential. In the fourth section, semianalytical expressions for the spheres diffusing in a harmonic potential are given. In the fifth section, the procedure for evaluating the PREs by numerically solving the equation of diffusion in an arbitrary potential (i.e. Smoluchowski equation) is discussed. Finally, in the sixth section, molecular dynamics (MD) simulations aimed at the modeling of the PRE effects are described. In the "Theoretical Predictions" section, we consider the dependences of the PREs on the relative placement of the two spins in the polypeptide chain and on the static magnetic field strength. In addition, we use the MD simulation results to test the validity of the Gillespie–Shortle model. Finally, in the "Experimental Results" section, we discuss the experimental procedures used to measure the PRE rates in MTSL-tagged proteins. The original experimental data from denatured samples of *Drosophila* drk N-terminal SH3 domain (drkN SH3) and ubiquitin are reported and discussed in relation with the developed theoretical models.

Theory

Overview

Paramagnetic dipolar relaxation of the nuclear spin $I = 1/2$ by the electron spin $S = 1/2$ is described by:[43,44]

$$R_2^{pmag} = \mathrm{PRE} = \frac{1}{20} D_{IS}^2 (4J(0) + 3J(\omega_I)) \qquad (1.1)$$

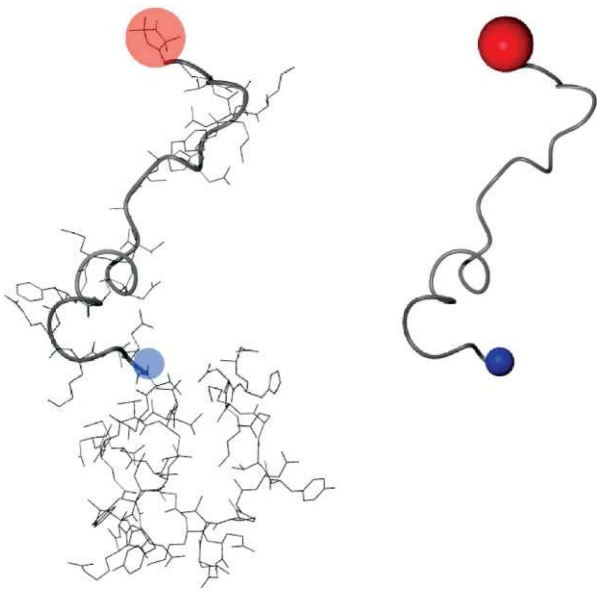

Figure 1. Model of the two spherical particles on a string. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

$$R_1^{\text{pmag}} = \frac{1}{10} D_{IS}^2 (3J(\omega_I)) \tag{1.2}$$

$$J(\omega) = \int_0^\infty g(\tau)\cos(\omega\tau)\mathrm{d}\tau \tag{2}$$

$$g(\tau) = 4\pi \left\langle \frac{Y_{20}^*(\Omega_{IS}(t))}{r_{IS}^3(t)} \frac{Y_{20}(\Omega_{IS}(t+\tau))}{r_{IS}^3(t+\tau)} \right\rangle \tag{3}$$

where R_2^{pmag} and R_1^{pmag} are the transverse and longitudinal paramagnetic relaxation rates, respectively, $J(\omega)$ is the spectral density function, $g(\tau)$ is the dipolar correlation function, $Y_{2m}(\Omega)$ are spherical harmonics, $\Omega_{IS}(t)$ are the time-dependent polar angles of the vector connecting the two interacting spins, and $r_{IS}(t)$ is this vector variable length. The interaction constant in Eq. (1) is $D_{IS} = (\mu_o/4\pi)\mu_B g_S \gamma_I$, where μ_o is the magnetic permeability of vacuum, μ_B is the Bohr magneton, γ_I is the nuclear spin gyromagnetic ratio, and g_S is the electron spin g factor. For the MTSL label $g_S = 2.0058$,[45] which leads to $D_{IS} = 4.9764 \times 10^{-22}$ m^3 s^{-1} for dipolar interaction between the MTSL and $^1H^N$. Normally, the PRE experiments target the quantity R_2^{pmag}; we shall refer to it as the PRE rate.

Consider a simplified model where $^1H^N$ and the MTSL nitroxyl moiety are represented as a pair of spheres connected by a string, Figure 1. The distance of closest approach for the two spheres is equal to the sum of their radii, denoted by d_o. The distance of maximum separation corresponds to the fully extended string and is denoted by L, Figure 2(a). The relative motion of the two spheres can be modeled as a diffusion process (segmental diffusion in a polymer chain).

The vector r_{IS} connecting the centers of the two spheres is characterized by the probability distribution, $P(r_{IS}, \Omega_{IS})$. Since in isotropic solution the orientational dependence is trivial, this distribution is reduced to $P(r_{IS})$. The pair correlation function (pcf) $P(r_{IS})$ in an unfolded protein is, generally speaking, nontrivial; the specific form of $P(r_{IS})$ can be emulated by introducing the mean-force potential, $U(r_{IS}) = -k_B T \ln P(r_{IS})$, where k_B is the Boltzmann constant and T is temperature. By doing so, the problem is cast in terms of mutual diffusion of the two particles subject to a restraining potential. For instance, Figure 2(a,b) illustrates the case where one of the particles is placed at the origin, whereas the other diffuses in the restricted region of space between the two concentric spheres, of radii d_o and L. This specific motion is termed "diffusion in a square-well potential" in accordance with the shape of

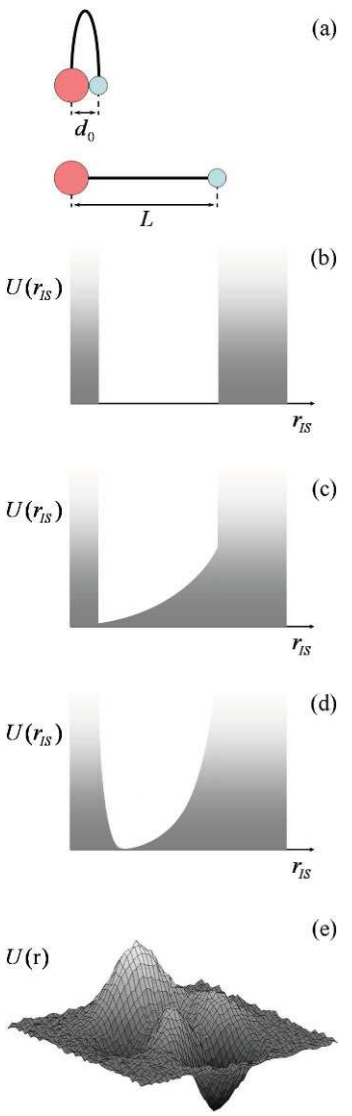

Figure 2. Schematic representation of the restraining potentials in different diffusion models: (a,b) reflecting-wall boundary conditions and the corresponding square-well potential; (c) harmonic potential; (d) numerically defined potential; (e) force-field potential as employed in the MD simulations. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

$U(r_{IS})$. Other forms of $U(r_{IS})$ investigated in this study are illustrated in Figure 2(c–e).

Note that in reality the system diffuses over a complex energy landscape, such as schematically illustrated in Figure 2(e).[46] The one-dimensional function $U(r_{IS})$, Figure 2(b–d), represents a "collapsed" version of this complex surface. In the one-dimensional model, it is also assumed that the diffusion coefficient D_{tr} is constant and does not depend on the coordinates of the system in the phase space.

Model 1: Slow segmental diffusion (Gillespie–Shortle model)

In their pioneering work, Gillespie and Shortle suggested that Eqs. (2) and (3) for the PRE spectral densities can be evaluated as[25,28]:

$$J(\omega) = \left\langle \frac{1}{r_{IS}^6} \right\rangle \frac{\tau_{rot}}{1 + \omega^2 \tau_{rot}^2} \tag{4}$$

where τ_{rot} is the correlation time of the overall molecular tumbling. Strictly speaking, this model is valid if segmental diffusion in the polypeptide chain is much slower than the overall tumbling, that is, in the case of dense molten globule, approaching the limit of the folded protein. The attraction of this model is that the effects of the structure and dynamics are separated (albeit in an ad hoc manner). The structural variability manifests itself only in the $\langle r_{IS}^{-6} \rangle$ part, which can be readily calculated using any given model pcf, $P(r_{IS})$. For the sake of illustration, we consider the simplest form of $P(r_{IS})$ consistent with extensive conformational exchange in unfolded protein [cf. Fig. 2(a,b)]:

$$P(r_{IS}) = \begin{bmatrix} 0 & r_{IS} < d_0 \quad \text{and} \quad r_{IS} > L \\ \left(\frac{4\pi}{3}(L^3 - d_0^3)\right)^{-1} & d_0 < r_{IS} < L \end{bmatrix} \tag{5}$$

which immediately yields

$$J(\omega) = \frac{1}{d_0^3 L^3} \frac{\tau_{rot}}{1 + \omega^2 \tau_{rot}^2} \tag{6}$$

This result suggests that with the increasing separation L between the two spin sites in the polypeptide chain the PRE rates drop as L^{-3}. The segmental diffusion in Gillespie–Shortle model is assumed to be slow on the time scale of the overall tumbling, $\tau_{tr} \gg \tau_{rot}$. At the same time, it is reasonable to expect that segmental diffusion is fast on the time scale of chemical shift variations, $\tau_{tr} \ll \Delta\omega_H^{-1}$, so that the observations are not hampered by exchange broadening (which, in fact, happens in many unfolded proteins[25]). The first condition, $\tau_{tr} \gg \tau_{rot}$, is highly restrictive; it will be lifted in the following sections.

Model 2: Diffusion in a square-well potential

In the unfolded protein with random-coil-like behavior, the relative motion of the $^1H^N$ spin and the MTSL spin is best described in terms of mutual diffusion. The simple model illustrated in Figure 1 remains the most obvious choice. There is a single characteristic time to the problem, that of segmental (translational) diffusion, $\tau_{tr} = d_0^2/D_{tr}$. Unlike in the Gillespie–Shortle model, no assumptions are made about the length of τ_{tr} (there is also no need to invoke the correlation time τ_{rot}, since this measure is not particularly meaningful for a random-coil polymer). In the radial direction, the diffusion is bounded by the two reflecting walls: at $r_{IS} = d_0$, corresponding to the distance of closest approach between the particles, and at $r_{IS} = L$, corresponding to the fully stretched chain [see Fig. 2(a,b) and Eq. (5)]. In the lateral direction, the diffusion is free. The solution to this problem is summarized below (complete derivation can be found in the Supporting Information):

$$J(\omega) = \frac{1}{d_0^3} \frac{3}{L^3 - d_0^3} \sum_{k=1}^{\infty} a_k \frac{\tau_{tr}^{(k)}}{1 + \left(\omega \tau_{tr}^{(k)}\right)^2} \tag{7.1}$$

$$a_k = \left(\int_1^{\lambda} x^{-1} \rho_k(x) dx \right)^2 \Big/ \left(\int_1^{\lambda} x^2 \rho_k^2(x) dx \right) \tag{7.2}$$

$$\lambda = L/d_0 \tag{7.3}$$

$$\rho_k(x) = n_2'(\beta_k) j_2(\beta_k x) - j_2'(\beta_k) n_2(\beta_k x). \tag{7.4}$$

Here $j_2(x)$ and $n_2(x)$ are spherical Bessel functions of the first and the second kind, respectively,

$$j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x}\right) \sin x - \frac{3}{x^2} \cos x \tag{7.5}$$

$$n_2(x) = -\left(\frac{3}{x^3} - \frac{1}{x}\right) \cos x - \frac{3}{x^2} \sin x, \tag{7.6}$$

the prime attached to j_2 and n_2 denotes the derivative:

$$f'(a) = \left. \frac{df(x)}{dx} \right|_{x=a}, \tag{7.7}$$

and β_k stands for the kth root of the following equation:

$$j_2'(\beta) n_2'(\lambda\beta) - n_2'(\beta) j_2'(\lambda\beta) = 0. \tag{7.8}$$

Finally, the multiple correlation times $\tau_{tr}^{(k)}$ are expressed as

$$\tau_{tr}^{(k)} = \tau_{tr}/\beta_k^2. \tag{7.9}$$

Multi-Lorentzian spectral density function, such as the one given by Eq. (7.1), represents a general solution for a stationary Markovian process that obeys detailed balance conditions.[47] Various examples of multi-Lorentzian spectral densities have been surveyed by Halle *et al.*[48]

The results Eqs. (7.1)–(7.9) are equivalent to the ones obtained by Bertil Halle in his analysis of intermolecular relaxation, although the mathematical

formulation is different.[49] Halle treated a system where spherical molecules bearing spin S are confined to the surface layer of a (large) spherical molecule bearing spin I. It should be noted that numerical implementation of Halle's formulas requires certain precautions because of the singularities that occur therein. We also find it advantageous that Eq. (7.1) puts into evidence the multi-Lorentzian character of the spectral density. Aside from that, the two sets of formulas are equally suitable for the purpose of numeric calculations and produce identical results.

The result Eqs. (7.1)–(7.9) reproduces two familiar limiting cases. The case $\lambda = 1$ corresponds to a rigid dumbbell-shaped molecule, where two spheres are in direct contact with each other. The spectral density is then reduced to

$$J(\omega) = (1/d_0^6)\tau_{rot}/(1 + \omega^2\tau_{rot}^2) \tag{8}$$

where $\tau_{rot} = d_0^2/6D_{tr}$. On the other hand, in the limit $\lambda = \infty$, the model represents two spherical particles connected with an infinitely long string, that is, effectively a pair of freely diffusing particles. In this case, the expression for spectral density asymptotically converges to the formula for translational relaxation[41,50]:

$$J(\omega) = \frac{4}{9}\frac{1}{d_0^3 L^3}\tau_{tr}\left(\frac{1 + \frac{5\sqrt{2}}{8}(\omega\tau_{tr})^{1/2} + \frac{1}{4}\omega\tau_{tr}}{1 + \sqrt{2}(\omega\tau_{tr})^{1/2} + \omega\tau_{tr} + \frac{\sqrt{2}}{3}(\omega\tau_{tr})^{3/2} + \frac{16}{81}(\omega\tau_{tr})^2 + \frac{4\sqrt{2}}{81}(\omega\tau_{tr})^{5/2} + \frac{1}{81}(\omega\tau_{tr})^3}\right) \tag{9}$$

Note the distinctive behavior at low magnetic fields: for $\lambda = \infty$ the spectral density declines as $J(0) - \text{const} \cdot \sqrt{\omega}$, whereas for $\lambda = 1$ the low-field asymptote is $J(0) - \text{const} \cdot \omega^2$.[51]

Finally, Eqs. (7.1)–(7.9) can be dramatically simplified when the spectral density at zero frequency, $J(0)$, is evaluated. Using the symbolic computation module of MATLAB,[52] we obtained

$$J(0) = \frac{1}{d_0^3 L^3}\tau_{tr}\varphi(\lambda) \tag{10.1}$$

$$\varphi(\lambda) = \frac{\lambda^2(32\lambda^4 + 37\lambda^3 + 42\lambda^2 + 37\lambda + 32)}{72(\lambda^2 + \lambda + 1)(\lambda^4 + \lambda^3 + \lambda^2 + \lambda + 1)} \tag{10.2}$$

In the limiting cases of $\lambda = 1$ and $\lambda = \infty$, this expression reduces to $(1/d_0^3L^3)(1/6)\tau_{tr}$ and $(1/d_0^3L^3)(4/9)\tau_{tr}$, respectively.

The latter result suggests that the effect of dynamics on the residues proximal to the MTSL label (i.e., separated by a short segment of a chain) and the distant residues (separated by a long segment of a chain) is appreciably different. Note however that experimental observation of the PREs close to the MTSL site is impossible using conventional techniques (such measurements would require specialized experiments with ^{13}C detection[53,54]). Furthermore, the steep distance dependence tends to mask the difference between the motional factors such as $(1/6)\tau_{tr}$ and $(4/9)\tau_{tr}$.

The biggest weakness of the above approach is that the rectangular pair-correlation function, Eq. (5), is greatly oversimplified. It is clear, for instance, that finding a polypeptide chain in its most extended conformation is a rare happenstance, and the probability $P(r_{IS} = L)$ should be, therefore, negligibly small. This is not what is predicted by Eq. (5). The treatments utilizing more realistic pcfs are described in the following three sections.

Model 3: Diffusion in harmonic potential

The problem of the dipolar relaxation of the two spins diffusing in the harmonic potential has been treated by Ullman.[55] It should be noted, however, that this treatment is deficient in two aspects: (i) the conditional probability $P(r_o, \Omega_o, 0|r, \Omega, \tau)$[41] does not converge to the equilibrium distribution $P(r, \Omega)$ at $\tau \to \infty$, and (ii) the boundary conditions are implemented incorrectly as the two particles can diffuse through each other (although the dipolar interaction is effectively "switched off" for a period of time when the distance between the two spins falls below d_o). The same inconsistencies appear in the Abragam's treatment of translational relaxation[43]; as pointed out by Hwang and Freed, correcting these problems leads to significant changes in the predicted relaxation rates.[41,56] Finally, it should be noted that Ullman's result for $J(\omega)$ takes a form of a triple integral which cannot be easily evaluated because of severe singularities.

In this work, we present a rigorous result for the spectral density at zero frequency, $J(0)$, for the model involving harmonic potential. While a general result for $J(\omega)$ can be produced by means of the numeric scheme (see next section), it is useful to have a computationally efficient and relatively compact expression for $J(0)$ that dominates the PRE rates. In our model, one of the particles is placed at the origin while the other diffuses in the space enclosed between two concentric spheres, of radii d_o and L. The second particle is also subjected to the effect of the restraining potential [Fig. 2(c)]:

$$U(r_{IS}) = k_B T\frac{r_{IS}^2}{2\sigma^2} \tag{11}$$

where the significance of parameter σ is discussed later. Note that the harmonic potential is not dictated by the actual atomic forces—rather it is a ruse to

reproduce the Gaussian distribution $P(r_{IS})$ characteristic of the random-flight polymers.

In our theoretical treatment, we follow the recipe developed by Szabo *et al.*[57] In brief, these authors showed how the time integral of the correlation function (i.e., $J(0)$) can be obtained without recourse to explicit solution of the Smoluchowski equation. Instead, the problem has been reduced to an inhomogeneous differential equation that can be readily integrated. We summarize the obtained results below (the calculation details can be found in the Supporting Information):

$$J(0) = \frac{\tau_{tr}}{3d_0^2\sigma^4}\frac{1}{(I_2(L/\sigma) - I_2(d_0/\sigma))} \times \int_{d_0/\sigma}^{L/\sigma} z^{-4}(a_1 + a_2 I_4(z) + I_1(z))dz \quad (12.1)$$

$$I_1(z) = 1 - \exp\left[-\frac{z^2}{2}\right] \quad (12.2)$$

$$I_2(z) = \sqrt{\frac{\pi}{2}}\mathrm{erf}\left[\frac{z}{\sqrt{2}}\right] - z\exp\left[-\frac{z^2}{2}\right] \quad (12.3)$$

$$I_4(z) = 3\sqrt{\frac{\pi}{2}}\mathrm{erf}\left[\frac{z}{\sqrt{2}}\right] - (z^3 + 3z)\exp\left[-\frac{z^2}{2}\right] \quad (12.4)$$

$$a_1 = \frac{R_2'(L)R_3'(d_0) - R_2'(d_0)R_3'(L)}{R_1'(L)R_2'(d_0) - R_1'(d_0)R_2'(L)} \quad (12.5)$$

$$a_2 = \frac{-R_1'(L)R_3'(d_0) + R_1'(d_0)R_3'(L)}{R_1'(L)R_2'(d_0) - R_1'(d_0)R_2'(L)} \quad (12.6)$$

$$R_1(r) = \left(\frac{\sigma}{r}\right)^3 \exp\left[\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right] \quad (12.7)$$

$$R_2(r) = I_4\left(\frac{r}{\sigma}\right)R_1(r) \quad (12.8)$$

$$R_3(r) = I_1\left(\frac{r}{\sigma}\right)R_1(r) \quad (12.9)$$

The prime in Eqs. (12.5) and (12.6) denotes the derivative with respect to r, Eq. (7.7).

The question remains as to how to select the parameter σ, Eq. (11). The choice is largely empirical—low σ value corresponds to a more compact chain, while high σ describes an expanded chain. For ideal polymer chain $\sigma = l/\sqrt{3}$, where l is the root-mean-square distance between the ends of the chain (in our case, the distance between the paramagnetic label and the selected H^N atom). Importantly, l can be related to the number of segments in the chain which connects the two particles:[58]

$$l = b(|n_{HN} - n_{MTSL}| + n_{tag})^{1/2} \quad (13)$$

Here b is the effective length of the monomer (i.e., amino acid), n_{HN} and n_{MTSL} are the residue numbers of the amino acids carrying the $^1H^N$ reporter spin and the MTSL tag, respectively, and n_{tag} accounts for the length of the tag per se. The relationship Eq. (13) supplements the results Eqs. (12.1)–(12.9) in one important way: it introduces the explicit dependence on the residue number. Strictly speaking, however, this relationship is not a part of the model leading to Eqs. (12.1)–(12.9).

Two familiar limiting cases can be recovered from the results Eq. (12.1)–(12.9). The limit $\sigma \to 0$ corresponds to a pair of spherical particles connected with an infinitely stiff spring. The relative translation of the two particles in this situation is quenched and $J(0)$ turns into purely rotational spectral density, Eq. (8). On the other hand, if the spring is infinitely soft, $\sigma \to \infty$, then the present model reproduces the case of the square-well potential. Under these conditions, the results Eqs. (12.1)–(12.9) for $J(0)$ are equivalent to Eqs. (10.1)–(10.2). While harmonic restraining potential, Eq. (11), is adequate for modeling of the generalized polymer coil, it has a number of obvious shortcomings. For example, it fails to take into consideration the finite size of the polypeptide chain.[59] It is clearly desirable to use a more realistic potential representative of a specific primary sequence. Such an improved treatment is presented in the next section.

Model 4: Diffusion in numerically defined realistic potential

Below, we outline the "hybrid" approach to the calculation of the PRE in unfolded proteins. In this approach, the pcf $P(r_{IS})$ is constructed using molecular modeling/MD techniques. The potential of the mean force, which controls the relative position of the two spins, is then defined in a standard fashion as $U(r_{IS}) = -k_B T \ln P(r_{IS})$. On the other hand, the relative motion of the spins is treated as simple diffusion process, same as in the previous models. The result is the Smoluchowski equation for diffusion in the (numerically defined) potential $U(r_{IS})$. The algorithm for solving this equation, adapted for the problem of intermolecular nuclear spin relaxation, has been described by Hwang and Freed.[41] This algorithm, with several minor modifications, is employed in our study.

To produce the realistic pcf $P(r_{IS})$, we generated a large ensemble of the MTSL-tagged protein structures. Consider, for example, the specific example of the A3C mutant of the protein drkN SH3.

First, the random conformation of drkN SH3 A3C was generated using the program TraDES (FOLDTRAJ).[60] The input parameters in TraDES were set as recommended by Forman-Kay and coworkers.[61] It has been demonstrated that this program provides a reasonable sampling of the conformational space sampled by unfolded proteins.[61,62]

As a next step, an MTSL tag was attached to the protein molecule *in silico*. The topology of the MTSL, as coded by Battiste,[63] was taken from the Xplor-NIH

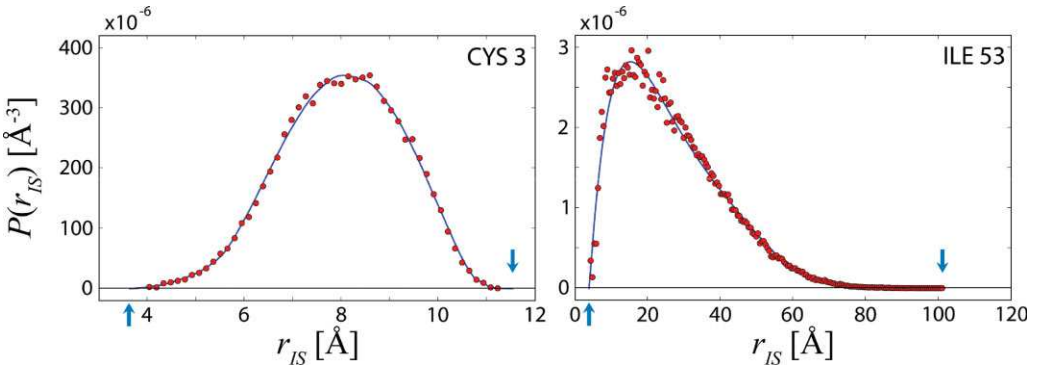

Figure 3. Selected pair correlation functions $P(r_{IS})$ from ensemble simulations of drkN SH3 A3C-MTSL. Represent the distance distribution between the paramagnetic center and the amide protons from residues Cys 3 and Ile 53. The data shown (red circles) are derived from the histograms representative of 50,000 random protein conformations generated with the help of the programs TraDES and Xplor-NIH. The data were partially smoothened by fitting all points on the histogram with seventh order polynomial, and then interpolated using piecewise cubic Hermite polynomials[67] (blue lines). The arrows mark the distance of closest approach, d_0, and the distance of the maximum separation, L. Note that programs other than TraDES have also been successfully used to generate conformational ensembles for disordered proteins.[4] [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

package.[64] The conformation of the linker connecting the 5-membered ring of MTSL to the protein backbone was initially randomized (the torsional angles χ_1 – χ_5 were set to random values).

The resulting construct, drkN SH3 A3C-MTSL, was optimized in the Xplor-NIH (1 ps of MD in a vacuum at a temperature of 800 K, followed by 200 steps of Powell energy minimization; CHARMM22 force field). The point charges for MTSL (previously reported[65] and recalculated by us using Gaussian[66]) were not included in the final series of calculations. Indeed, it was found that including the charges leads to some instances where the MTSL ring becomes tightly packed against the protein chain (essentially, an artefact from the structure calculations *in vacuo*).

Using this protocol, an ensemble comprised of 50,000 random conformations of the MTSL-tagged A3C drkN SH3 has been formed. This ensemble was used to generate distance histograms for distances between the unpaired electron, S, and the amide proton from nth residue, I. In doing so, it was assumed that the unpaired electron is localized at the nitrogen atom of the MTSL ring.[68] The histogram entries were then normalized and divided by $4\pi r_{IS}^2 dr_{IS}$ to arrive at the desired pcf, $P(r_{IS})$. Figure 3 shows the examples of $P(r_{IS})$ for the amide proximal to the paramagnetic center, as well as the distant amide (residues Cys 3 and Ile 53, respectively).

The histogram-derived $P(r_{IS})$ data (indicated with circles in Fig. (3)) were smoothened and interpolated, leading to continuous representation (solid line in Fig. 3). The points where $P(r_{IS})$ turns into zero (marked with arrows in the plot) are taken to be the distance of closest approach, d_0, and the distance of maximum separation, L. The smoothened pcf has been used to evaluate the effective force that enters into the Smoluchowski equation, as described below.

The Smoluchowski equation for diffusion of the particle in the potential $U(r)$ has been discussed in some detail by Hwang and Freed.[41] It is convenient to cast this equation in terms of the dimensionless distance, $x = r/d_0$, and then discretize it on a uniform grid, $x_0 = 1$, $x_1 = 1 + \Delta$, $x_2 = 1 + 2\Delta$, ..., $x_N = \lambda$. Following a series of transformations, this leads to the system of linear equations:

$$[-W + B - i(\omega\tau_{tr})E]Q = R \tag{14}$$

that needs to be solved with regard to the column vector Q (which is in turn used to calculate $J(\omega)$, see below). The column vector R is composed of the elements $P(d_0x_i)/x_i^2$ with $i = 0, 1, 2, ..., N$. The transition matrix W is defined as follows (only nonzero elements are listed):

$$W_{i,i-1} = \frac{1}{\Delta^2} + \frac{F(x_i)}{2\Delta} \tag{15.1}$$

$$W_{i,i} = -\frac{2}{\Delta^2} - \frac{F(x_{i+1})x_{i+1} - F(x_{i-1})x_{i-1}}{2x_i\Delta} \tag{15.2}$$

$$W_{i,i+1} = \frac{1}{\Delta^2} - \frac{F(x_i)}{2\Delta} \tag{15.3}$$

This definition holds for all elements from the three main diagonals of the matrix, except those for which $i = 0$ and $i = N$:

$$W_{0,0} = \left(\frac{x_1}{x_0}\right)\left\{-\frac{2}{\Delta^2} - \frac{F(x_1)}{\Delta}\right\} \tag{15.4}$$

$$W_{0,1} = \frac{2}{\Delta^2} - \frac{F(x_0)}{\Delta} \tag{15.5}$$

$$W_{N,N-1} = \frac{2}{\Delta^2} + \frac{F(x_N)}{\Delta} \tag{15.6}$$

$$W_{N,N} = \left(\frac{x_{N-1}}{x_N}\right)\left\{-\frac{2}{\Delta^2} + \frac{F(x_{N-1})}{\Delta}\right\} \qquad (15.7)$$

The definitions Eqs. (15.4)–(15.7) correspond to the reflecting wall boundary conditions at $r = d_o$ and $r = L$. The elements $W_{i,j}$ so defined satisfy the probability conservation condition:

$$\sum_{i=0}^{N} c_i x_i W_{i,j} = 0 \quad (j = 0, 1, 2, \ldots, N) \qquad (16)$$

where c_i is the integration weighting coefficient, $c_i = 1$ for all i except the end points, $c_o = c_N = 0.5$. The force that enters into the transition probability matrix W is

$$F(x_j) = -\frac{1}{k_B T}\frac{\mathrm{d}U(d_o x)}{\mathrm{d}x}\bigg|_{x=x_j} = \frac{\mathrm{d}\ln P(d_o x)}{\mathrm{d}x}\bigg|_{x=x_j} \qquad (17)$$

Finally, Eq. (14) includes the diagonal matrix B, $B_{i,i} = 6/x_i^2$, and the term $(\omega\tau_{tr})E$, where E is the unity matrix, ω is a Larmor frequency chosen to sample the spectral density, $J(\omega)$, and τ_{tr} is equal to d_o^2/D_{tr}. It can be readily noted that the matrix $-W + B - i(\omega\tau_{tr})E$ in the Eq. (14) is tridiagonal. The equation can be, therefore, efficiently solved using the Thomas algorithm.[69] The column vector Q obtained from solving Eq. (14) is then used to construct the spectral density $J(\omega)$:

$$J(\omega) = 2\pi\frac{\Delta}{D_{tr} d_o}\sum_{i=0}^{N} c_i \frac{\mathrm{Re}\{Q_i\}}{x_i^2} \qquad (18)$$

To test the performance of this algorithm, we conducted the calculations using the uniform pcf, Eq. (5). The results proved to be in perfect agreement with those obtained from the analytical treatment, Eqs. (7.1)–(7.9).

The procedure presented in this section constitutes a major improvement over the simplistic model involving the square-well effective potential. Nevertheless, it is important to realize that the described formalism has some serious limitations. Indeed, real proteins evolve in a phase space of very high dimension, diffusing over the rugged potential energy surface.[46] In the current treatment, this complex potential surface is "collapsed" and reduced to the one-dimensional $U(r_{IS})$. This approach involves a great deal of averaging and results in $U(r_{IS})$ taking the appearance of a smooth function. In an attempt to develop a more realistic model for the relative motion of the two spins, we resort to MD simulations.

Model 5: MD simulations of the PRE effect

It is difficult to simulate the dynamics of an unfolded protein in a fully realistic fashion with sufficiently good statistics. A very large water box is needed to accommodate the unfolded protein, which sharply raises the cost of computations. The phase space accessible to the unfolded protein is much larger than the one for a folded protein. It is difficult to expect that this phase space can be sufficiently well sampled over the length of the typical MD trajectory, assuming that the simulations are conducted at around room temperature.

Faced with these difficulties, we opted for the simplified MD protocol where the simulations are conducted (i) in vacuum and (ii) at an elevated temperature. The results cannot be, therefore, viewed as quantitatively accurate. The simulations are rather intended to capture the general trends in the dynamics of the random-coil protein and the resulting PRE parameters. Note also that our MD simulations are in no way related to the attempts to fold proteins *in silico*.[70] Instead, we monitor protein motions on the far periphery of the folding funnel, which appears to be a relatively straightforward task.

The starting coordinates of the MTSL-tagged drkN SH3 have been generated as described in the previous section. These coordinates were energy-minimized prior to the MD production run; the first 200 ps of the MD trajectory were considered an equilibration phase and excluded from the data analyses. The trajectories were recorded in vacuum, using torsional angle MD (IVM module[71] of Xplor-NIH). Bond, angle, improper, and van der Waals energy items have been included, in accordance with the standard IVM protocol.[72,73] The translational motion of the center of mass has been subtracted out. The reorientational motion, however, was retained. In the case of the unfolded protein, any attempt to separate the reorientational dynamics from the internal motions would be (i) rather meaningless and (ii) not helpful from the point of view of the PRE predictions.

The *in vacuo* simulation protocol proved to be sufficiently robust, with the production rate of about 10 ns per day on 3 GHz dual-core Xeon processor. A total of nine trajectories has been generated for three MTSL-tagged drkN SH3 mutants (A3C, D32C, and D59C) at three target temperatures, see Table I. At the time of processing, the trajectories ranged from 200 to 278 ns in length. From each MD trajectory, we extracted the dipolar vectors connecting the paramagnetic center with individual H^N atoms and evaluated the temporal correlation functions $g(\tau)$, Eq. (3).[74] Note that the time dependence in $g(\tau)$ arises from both angular and radial degrees of freedom, r_{IS} and Ω_{IS} (as can be appreciated from the analyses in the previous sections, these degrees of freedom are coupled and cannot be easily separated). The MD-based correlation functions $g(\tau)$ were subsequently used to evaluate the spectral density and predict the PRE rates.

The high-temperature MD trajectories of MTSL-tagged drkN SH3 show extensive conformational dynamics (two excerpts from the trajectories, rendered in the form of .mpg movies, can be found in the Supporting Information). Accordingly, the correlation

Table I. *Radii of Gyration (Å) for Various Structural Ensembles Representing Unfolded drkN SH3–MTSL*

	TraDES/Xplor ensemble[a]	Xplor trajectories		
		700 K[b]	1000 K[b]	1400 K[b]
A3C	18.4	13.4	16.5	21.0
D32C	18.2	13.0	17.0	20.4
D59C	18.4	13.2	18.8	22.9

[a] The TraDES input parameters as recommended by Marsh *et al.*[61] and used in this work lead to most compact ensembles. If TraDES is used in default configuration, R_g is increased by ca. 1.5 Å. If secondary structure preferences[96] are omitted, R_g is further increased by ca. 1 Å.

[b] During the MD simulations, the temperature fluctuated around the target values. The actual readings were 716 ± 39, 681 ± 37, and 702 ± 38 K for three low-temperature trajectories, 963 ± 46, 968 ± 45, and 976 ± 45 K for medium-temperature trajectories, and 1416 ± 54, 1400 ± 57, and 1389 ± 53 K for high-temperature trajectories. Note that comparable degree of protein expansion can be obtained in room-temperature simulations—but only when the solvent is included.

functions $g(\tau)$ are reasonably well converged (verified by a number of tests involving duplicate trajectories and trajectories of varying length).* The characteristic decay times of $g(\tau)$ are on the order of 10 ps (see Fig. 4), which is much shorter than the value $\tau_{tr} = 2$ ns based on experimental estimates. Thus, the use of high-temperature torsional angle dynamics greatly improves the sampling of the conformational space. It can be suggested that in this respect our 200-ns long trajectories are equivalent to tens of microsecond simulations at room temperature.

To relate the high-temperature MD data to the results of the experimental measurements taken at room temperature, we "reset" the time axis in the MD simulations. For this purpose, all correlation functions from each trajectory (see Fig. 4) have been fitted to a single exponential and the effective decay times τ_c^{eff} were determined. Next, the MD time axis was rescaled such as to equate the average τ_c^{eff} with the preselected value $\tau_{tr} = 2$ ns. The described procedure is akin to adjustment of the translational diffusion coefficient D_{tr} in the diffusion models discussed previously. Note that in all cases there is only one adjustable parameter involved, whereas the experimental PRE data sets consist of multiple data points.

It should be reiterated that the presented MD model offers, at best, a semiquantitative insight into the PRE effect in unfolded proteins. While there is evidence that MD simulations at elevated temperature adequately represent the unfolding events,[76–78] it is clearly desirable to conduct the simulations under milder conditions. Furthermore, introducing solvent and electrostatic interactions is key to more realistic modeling of the PRE effect (as discussed later, however, our vacuum simulations reproduced certain essential features that are normally associated with protein solvation). Our attempts to employ Langevin dynamics[79] were unsuccessful as the simulations failed to reproduce large-amplitude motions in the extended protein chain. The use of the implicit solvent,[80–83] on other hand, appears to be a viable option. While the computations involving implicit solvent are slower by approximately a factor of 5, they should still produce a satisfactory statistics.

In general, long MD trajectories of unfolded proteins can be viewed as an extension of static conformational ensembles.[17,38,84] Particularly attractive is the possibility to directly relate these trajectories to dynamic data such as PREs, as well as ^{15}N relaxation rates. The time-scale information obtained in this fashion is unique and adds a new dimension to the structural modeling.

Theoretical Predictions

Dependence of PRE on residue number

In this section, we consider the theoretical predictions for PRE rates as a function of the residue number, that is, the position of the $^1H^N$ reporter spin in the polypeptide chain (the discussion of the experimental data is deferred until later). We begin with the square-well potential model, Eqs. (7.1)–(7.9).

There are a number of parameters that need to be estimated prior to calculations. Assuming that the unpaired electron is localized on the nitrogen atom of the MTSL ring, we estimate $d_o \approx 4$ Å (cf. Fig. 3). The maximum separation distance is calculated according to $L = (|n_{HN} - n_{MTSL}| \cdot 2.55 + 12)$ Å, where the value 2.55 Å/residue refers to the maximally extended polypeptide chain and 12 Å is the estimated length of the extended MTSL chain. The calculations also require the knowledge of the effective translational diffusion coefficient, D_{tr}, that characterizes the relative motion of the MTSL label and $^1H^N$ reporter spin. The values of D_{tr}, both experimental and theoretical, can be found in the literature discussing the fluorescence quenching and electron transfer in unfolded proteins. Although the results vary, the consensus estimate appears to be $D_{tr} \sim 10^{-6}$ cm^2 s^{-1} for a pair of sites separated by a sufficiently long stretch of the chain (>ca.10 residues) at room temperature.[85–91] This corresponds to the characteristic motional time scale $\tau_{tr} = d_o^2/D_{tr} \sim 2$ ns.

The results of the computations employing the above set of parameters, d_o, L, and τ_{tr}, are presented in Figure 5 (red bars). In the same plot, we have also shown the predictions from the Gillespie–Shortle model integrated with the rectangular pcf, Eq. (6) (blue bars). In evaluating Eq. (6), the correlation time was set to $\tau_{rot} = (4/9)\tau_{tr}$, that is, to 0.9 ns.* This choice ensures that the results from the Gillespie–

*For instance, two independently started 200 ns trajectories of drkN SH3 D59C – MTSL at 1400 K produced highly correlated PRE data, $r = 0.99$ (as always in this work, the comparison is limited to the PRE rates that are lower than 100 s^{-1}).

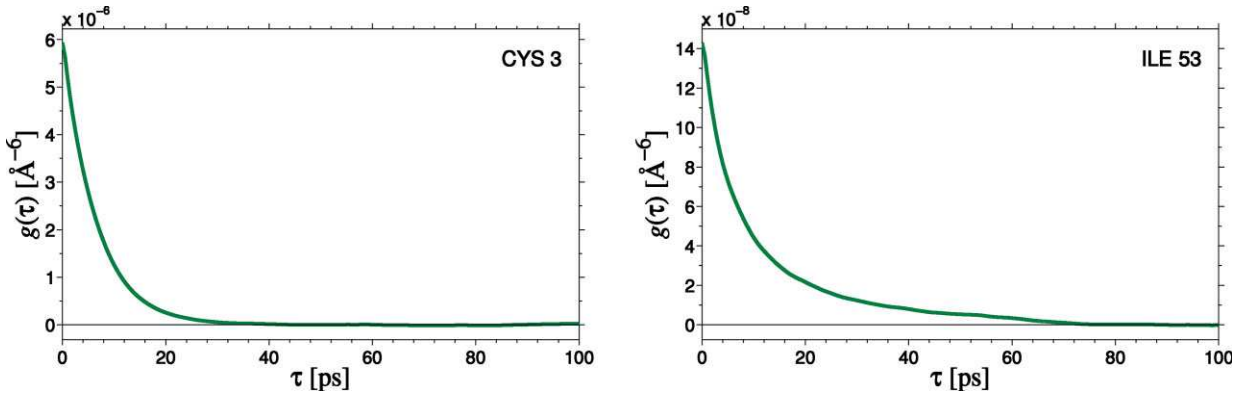

Figure 4. Time correlation functions $g(\tau)$ for dipolar interactions between the paramagnetic center and H^N atoms from residues Cys 3 and Ile 53 in drkN SH3 A3C-MTSL. Extracted from 242-ns MD trajectory (target temperature 1000 K, sampling step 0.5 ps). Processing of the MD data, including the optimized sampling and smoothing of $g(\tau)$, is the same as described in our recent work.[75] Note that the correlation for Cys 3 decays noticeably faster than that for Ile 53 (effective decay time 6.8 ps vs. 10.7 ps). Indeed, for two nearby particles mutual diffusion leads to rapid modulation of the dipolar interaction; conversely, for two distant particles, the modulation is slow. A similar trend is predicted in the square-well potential model, Eq. (10). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

separated by a sufficiently long stretch of the chain (>ca.10 residues) at room temperature.[85–91] This corresponds to the characteristic motional time scale $\tau_{tr} = d_0^2/D_{tr} \sim 2$ ns.

The results of the computations employing the above set of parameters, d_o, L, and τ_{tr}, are presented in Figure 5 (red bars). In the same plot, we have also shown the predictions from the Gillespie–Shortle model integrated with the rectangular pcf, Eq. (6) (blue bars). In evaluating Eq. (6), the correlation time was set to $\tau_{rot} = (4/9)\tau_{tr}$, that is, to 0.9 ns.* This choice ensures that the results from the Gillespie–Shortle model coincide with those from the diffusion in a square well model in the limit of long thread L. Figure 5 shows that the two models indeed predict nearly identical rates for the separations of 10 residues or more. For smaller separations, however, a sizeable difference is observed. This difference reflects the dependence of the effective correlation time on $\lambda = L/d_o$, Eq. (10).

Generally, Figure 5 demonstrates a steep dependence on the length of the tether L, which is dominated by the factor L^{-3}, Eqs. (6) and (7.1). This dependence reflects a crude nature of rectangular pcf—it greatly exaggerates the volume occupied by a random-coil protein and hence leads to underestimated PRE rates for residues with large L. Replacing the rectangular pcf with fundamentally much more sound Gaussian-type pcf leads to more satisfactory results.

As a next step, we consider the predictions obtained from the model involving diffusion in a harmonic potential. There is one extra input parameter that is required to evaluate the spectral density function in this model, namely, the effective monomer length b, Eq. (13).[92] This parameter is often expressed as $b = \sqrt{C_\infty} b_o$, where b_o is the distance between α carbons, 3.8 Å, and C_∞ is the constant that characterizes the degree of expansion of polypeptide chain. For a Gaussian chain, b is related to the radius of gyration R_g according to $R_g = (b/\sqrt{6})N^{0.5}$, where N is the number of residues in the protein. As it turns out, this simple formula holds fairly well for denatured proteins: the extensive experimental evidence suggests $R_g \sim N^{0.6}$, in agreement with the theoretical analyses taking into consideration excluded volume effects.[10,93]

Referring to the specific protein system investigated in this study, 59-residue drkN SH3 domain in solution of 2*M* GuHCl, we note that the radius of gyration for this system has been measured experimentally, $R_g = 21.9$ Å.[94] The effective monomer length is, therefore, estimated to be $b = 7$ Å.

We further related the length of the MTSL tag to b. For this purpose, we used the ensemble of 50,000 protein conformations (cf. Fig. 3) and calculated the root-mean-square distance between the nitrogen in the 5-membered MTSL ring and the H^N proton from the MTSL-tagged cysteine residue. The obtained value, 8 Å, is approximately equivalent to the root-mean-square amide–amide distance across three residues. We therefore assumed that $n_{tag} = 3$.

Note also that the PRE rates calculated from the harmonic potential model are limited to the term

*It is not quite clear what value of the correlation time is appropriate for use with the Gillespie-Shortle model. In the original work, the R_2^{pmag}, R_1^{pmag} data from the denatured protein Δ131Δ were used to determine τ_c (= τ_{rot}) according to Eqs. (1.1), (1.2) and (4). Strictly speaking, however, these equations are not suitable for random-coil proteins and, therefore, this analysis may not be fully self-consistent. Furthermore, it appears that a trivial computational error has been made in the article so that the extracted correlation times were overestimated by a factor of 3: the reported average correlation time at 32°C was 4.1 ns, the actual value is 1.3 ns (confirmed by D. Shortle). This error has percolated into subsequent body of work by other investigators,[30,38,40] as 4 ns was taken to be a typical value for any unfolded protein.

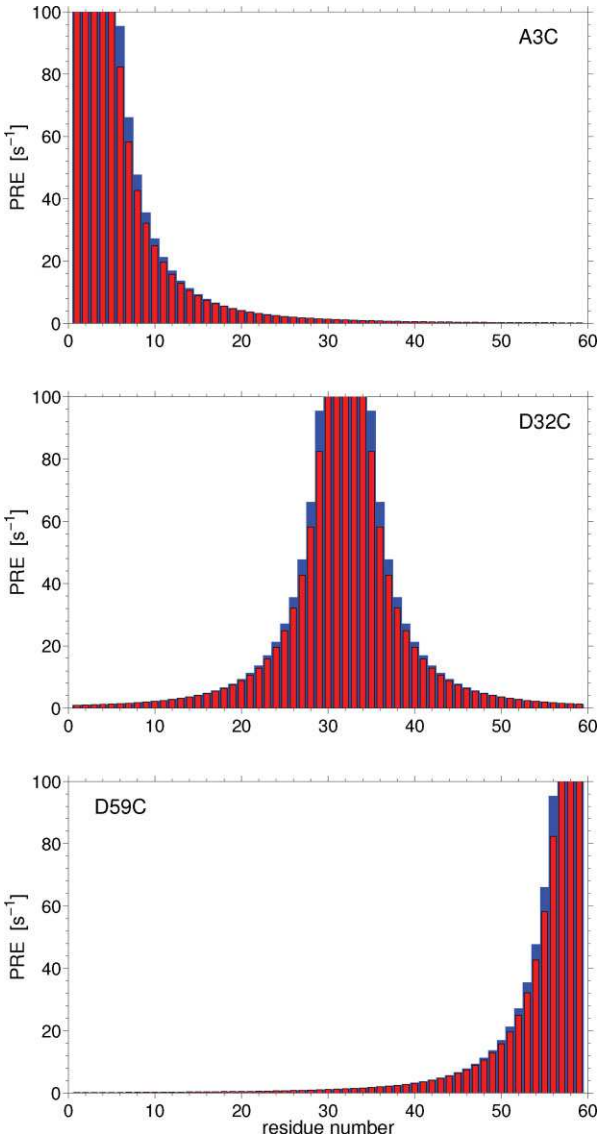

Figure 5. PRE profiles as predicted in the square-well potential model (red bars) and Gillespie–Shortle model with rectangular pcf (blue bars). The calculation seeks to mimic three MTSL-tagged mutants of the unfolded protein drkN SH3: A3C, D32C, and D59C. The simulation parameters are $d_0 = 4$ Å, $\tau_{tr} = d_0^2/D_{tr} = 2$ ns (0.9 ns for Gillespie–Shortle model), $\omega_H/2\pi = 600$ MHz.

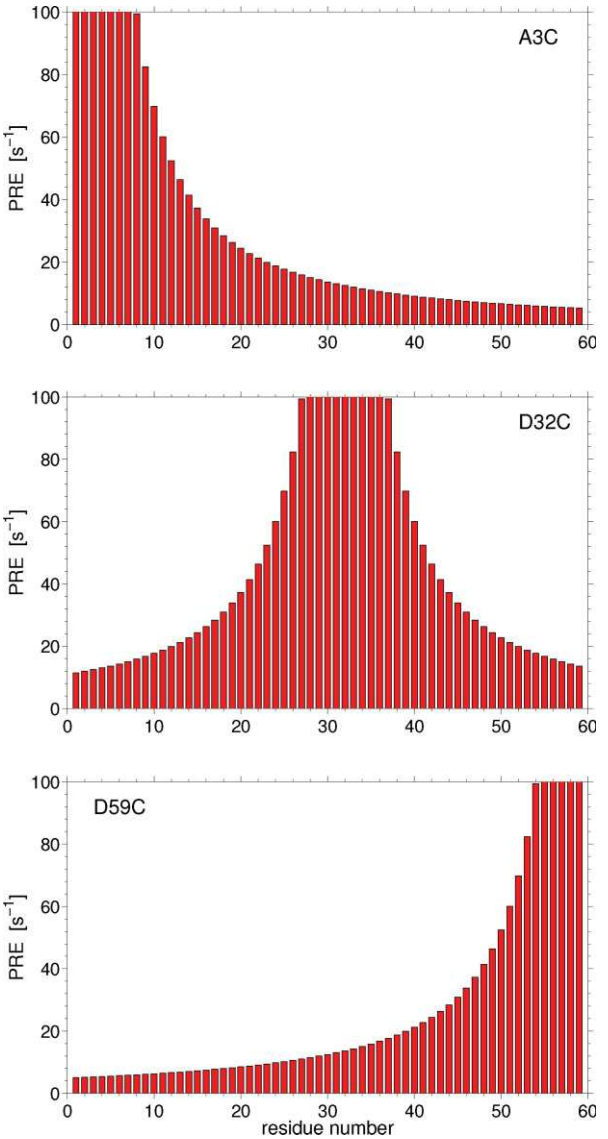

Figure 6. PRE profiles as predicted in the harmonic potential model. The rates are restricted to the (leading) term $J(0)$, Eqs. (1.1) and (12.1)–(12.9). The simulation parameters are $d_0 = 4$ Å, $\tau_{tr} = d_0^2/D_{tr} = 2$ ns, $b = 7$ Å.

containing spectral density function at zero frequency, $J(0)$, Eq. (12.1). Our subsequent calculations using numerically defined potentials (see later) indicate that omission of the $J(\omega_H)$ term translates into 4.5% average error in the calculated rates. Given the approximate character of the present model, the deviation of this magnitude is of little importance.

The calculated rates shown in Figure 6 in a good approximation follow the simple dependence PRE $\sim l^{-3} \sim (|n_{NH} - n_{MTSL}| + n_{tag})^{-3/2}$. The dropoff is obviously less steep than in the previous model, Figure 5, where the asymptotic dependence was $\sim L^{-3} \sim (|n_{NH} - n_{MTSL}| + n_{tag})^{-3}$. Thus, it is simply the fundamental property of the Gaussian chain, where the linear size increases as a square root of the number of residues, which dictates the shape of the PRE profile.

As a next step, we simulated the PRE rates in drkN SH3 using the numeric solution of the Smoluchowski equation for the particle diffusing in potential $U(r)$. As explained in the Theory section, the mean-force potential $U(r)$ was constructed on a basis of a large ensemble of random protein conformations generated by means of the programs TraDES and Xplor-NIH. In calculating the protein PRE rates, the diffusion coefficient was set to a constant, $D_{tr} = 0.8 \cdot 10^{-6}$ cm^2 s^{-1}. The correlation time τ_{tr}, Eq. (14), was calculated for each residue individually, $\tau_{tr} = d_0^2/D_{tr}$, using d_o values derived from the residue-specific pcfs, Figure

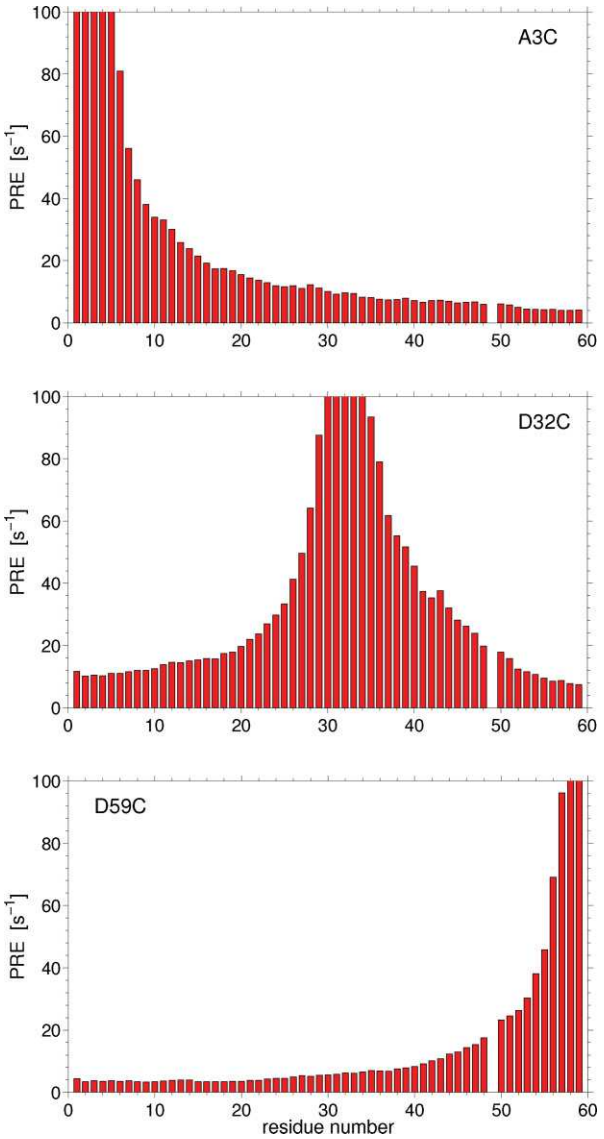

Figure 7. PRE profiles as predicted in the model involving numerically defined potentials. The pcf functions used in the calculations are as illustrated in Figure 3, $D_{tr} = 0.8 \cdot 10^{-6}$ cm^{2}s^{-1}, $\omega_H/2\pi = 600$ MHz. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

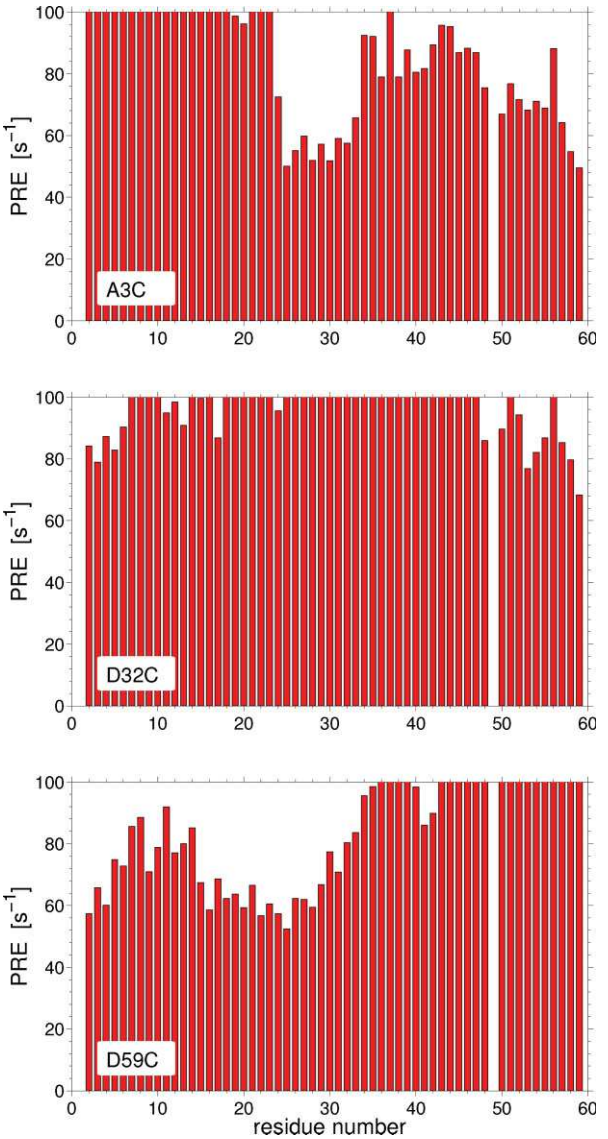

Figure 8. PRE profiles as predicted on the basis of the MD simulations at 700 K. The time scale of all MD trajectories is redefined such that the average decay time of dipolar correlation functions is equal to 2 ns (see text). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

3 (for d_o equal to 4 Å the correlation time is 2 ns, consistent with our previous calculations).

The results of the simulations using numeric solution of the Smoluchowski equation are shown in Figure 7. Interestingly, the simulated PRE profiles are rather smooth, suggesting that on a local level the polypeptide chain behaves similar to a homogeneous polymer, Figure 6. On the other hand, the profiles are clearly asymmetric. For instance, in the case of D32C, the residues upstream of the paramagnetic label are relaxed more efficiently: 35.3 s^{-1} in Asp 42 versus 23.7 s^{-1} in Thr 22. This asymmetry indicates the presence of subtle structural preferences that manifest themselves over a longer distance range. Of course, these observations are meaningful only insofar as the TraDES/Xplor-built ensemble can be considered realistic.

Finally, we discuss the PRE profiles generated on the basis of the MD simulations. Interestingly, the recorded trajectories reproduced different types of dynamic behavior covering the range from a collapsed chain to a highly expanded chain. At a lower target temperature, 700 K, the protein behavior is reminiscent of a molten globule, that is, the chain "jiggles and wiggles" while at the same time maintaining fairly compact near-spherical overall shape (visualized in Supporting Information). Normally, this kind of behavior would be attributed to a hydrophobic effect that forces the protein to minimize its solvent-exposed surface area. In our vacuum simulation, however, aggregation of the peptide

chain is caused by favorable van der Waals interactions. As the MD simulation temperature is raised, the chain adopts more extended conformations. At the 1400 K target temperature, it is highly expanded and shows the dynamics expected of a random-coil polymer in good solvent (see Supporting Information).

The degree of expansion of protein chain is commonly characterized by a radius of gyration. The R_g values calculated for our simulated trajectories (ensembles) are listed in Table I. These data can be compared with the experimental result by Choy *et al.*, who reported the effective radii of gyration 21.9 ± 0.5 Å for drkN SH3 in 2*M* guanidinium chloride.[94] Bear in mind that the experimental R_g value may include 1–2 Å contribution from the protein hydration shell, that is, the actual size of the protein could be somewhat smaller.[95] In the context of Table I, it appears that the experimental R_g value is best reproduced by the high-temperature (1400 K) MD simulations, corresponding to a highly expanded protein structure.

Before the MD data can be used to predict the PRE rates, the time scale of each trajectory needs to be redefined. For instance, in the case of drkN SH3 A3C–MTSL, we assume that the time interval between two consecutive snapshots is 86.2, 116.2, and 158.8 ps for trajectories with target temperatures of 700, 1000, and 1400 K, respectively. In this manner, it is ensured that the average decay time of the dipolar correlation functions $g(\tau)$ in each of the three trajectories is equal to 2 ns. Thus, all trajectories are rendered equivalent with respect to their "global" motional time scale; the individual correlation functions, however, remain different.

As it turns out, in the case of the molten globule simulations (700 K trajectories), the paramagnetic relaxation is very efficient, Figure 8. This can be readily rationalized as MTSL pyrrolinyl ring spends most of the time inside the compact globule, in proximity to multiple amide sites and with direct access to all of them. Note that this situation is markedly different from the one that is observed in a folded protein of comparable size: in a molten globule, the paramagnetic center can approach and efficiently relax all residues, whereas in a folded protein only a portion of the structure experiences a strong PRE effect.

As the simulation temperature is raised, the ensemble shifts toward more extended conformations. Consider, for instance, the data from drkN SH3 A3C-MTSL trajectory at 1000 K. As evidenced by gyration radius of this trajectory, 16.5 Å, the protein is moderately expanded. The PRE profile displays some nontrivial features, such as certain amount of "ruggedness" and a plateau extending toward C-terminus (top panel in Fig. 9). Relatively large PRE rates, $>15\ s^{-1}$, are observed even for residues located far away from the MTSL tag.

When the simulation temperature is raised even further, the PRE profile becomes more "generic" in appearance. The extreme example is drkN SH3 D59C-MTSL at 1400 K, which is highly expanded ($R_g = 22.9$ Å). The PRE profile in this case takes the appearance of a smooth, rapidly descending curve (bottom panel in Fig. 10). The PRE rates at the far terminus are small (under 2 s^{-1} for the first 20 residues). These observations are further discussed in the context of our experimental measurements.

Dependence of PRE on static magnetic field strength

In principle, the dependence of relaxation rates on magnetic field strength B_o provides the best test for distinguishing between various motional models. In this section, we provide some limited simulation data illustrating this dependence. For this purpose, we focus on the drkN SH3 A3C-MTSL construct and select several representative amide sites: Cys 3, which lies in close proximity to the paramagnetic center, Ser 10, which is removed far enough to allow for experimental PRE measurements, and the distant residue Ile 53, where the PRE effect is small. The spectral density profiles $J(\omega)$ for these three residues calculated on the basis of the square-well potential model, Eqs. (7.1)–(7.9), are shown in Figure 11 (red curves). In the same plot, we display the standard Lorentzian spectral densities generated according to Eq. (6) (blue curves). The parameters of the Lorentzian spectral density, τ_{rot} and d_oL, are adjusted in such a manner as to match the values of $J(0)$ and $J(\omega_H)$ at a proton frequency $\omega_H/2\pi = 600$ MHz.

Despite the crude character of the models, the pattern observed in Figure 11 is rather revealing. Recall that at low frequencies $J(\omega)$ is dominated by the conformational species where the two spins are separated by long distances r_{IS}. Indeed, for such distant spin pairs, mutual diffusion of the particles leads only to a slow modulation of dipolar interactions. Notice further that in the case of large separation L between the two particles, we recover the limiting case of translational relaxation with its characteristic low-field asymptote, $J(\omega) = J(0) - \text{const} \cdot \sqrt{\omega}$.[51] This distinctive asymptotic behavior is especially evident in the case of Ile 53, which is far removed from the paramagnetic label (bottom panel in Fig. 11). At the same time, even those amides that are close to the labeling site, Cys 3 and Ser 10, give rise to $J(\omega)$ profiles that are distinctly different from a single Lorentzian (two upper panels in Fig. 11).

While at low frequencies $J(\omega)$ behave in a peculiar fashion, at high frequencies, they follow the standard Lorentzian dependence, $J(\omega) = \text{const} \cdot \omega^{-2}$. The high-frequency behavior is decided by protein conformations where the two spins lie in close proximity to each other. For such proximal pairs, relative diffusion leads to a rapid modulation of dipolar interactions. Considering closely spaced pairs of particles, the distinction between rotational and translational

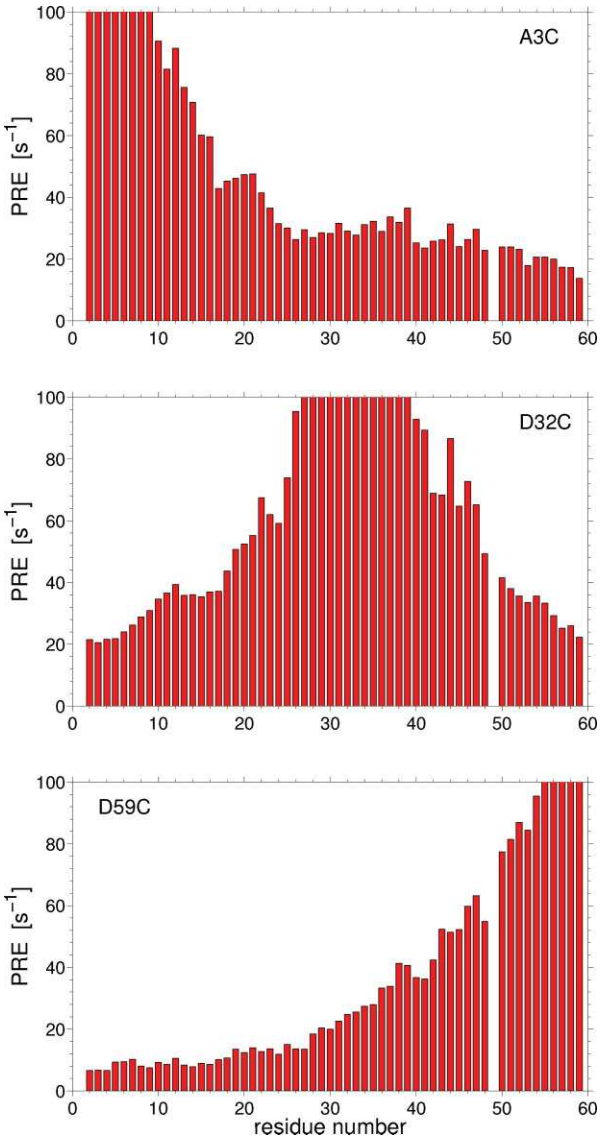

Figure 9. PRE profiles as predicted on the basis of the MD simulations at 1000 K. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

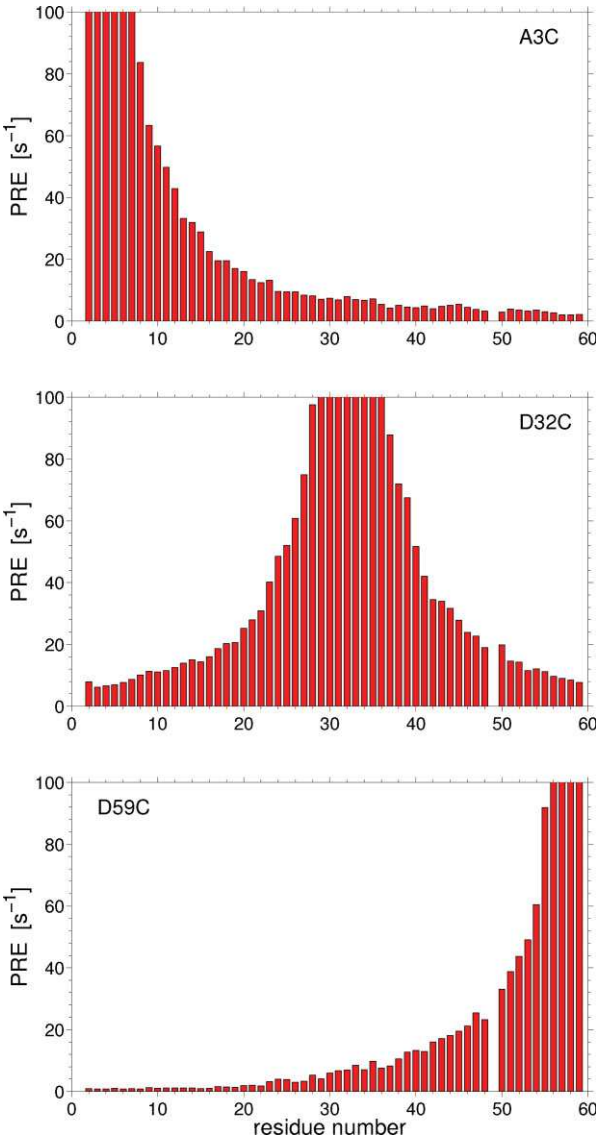

Figure 10. PRE profiles as predicted on the basis of the MD simulations at 1400 K. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

relaxation is lost, with both mechanisms predicting ω^{-2} high-field dependence. This Lorentzian dependence is observed at the tails of all spectral density profiles shown in Figure 11.

Consider now a more realistic model based on MD simulations data, Figure 12. In this case, the spectral density profile of Cys 3 is indistinguishable from a single Lorentzian. Two other residues, however, show significant deviations from Lorentzian behavior. For example, in the case of Ser 10, the maximum deviation between the two curves is 51%, reached at 120 MHz (Fig. 12, middle panel). This is similar to the square-well potential model, where the maximum deviation amounts to 63%, reached at 140 MHz (Fig. 11, middle panel).

While the non-Lorentzian character of $J(\omega)$ appears to be rather prominent, the experimental detection of this effect is likely to be extremely difficult. In principle, R_1^{pmag} measurements using a spectrometer with field-cycling capabilities can provide the answer.[97] The current state of technology, however, does not allow for high-resolution 2D spectra to be acquired in this fashion. Furthermore, proton spin diffusion can complicate the data analysis. On the other hand, an attempt to access the dispersion profiles $J(\omega)$ through heteronuclear measurements is likely to be faced with its own difficulties.[98] While we acknowledge the fundamental interest of relaxation dispersion studies, no attempt has been made to further investigate this aspect of the problem.

Applicability of the Gillespie–Shortle model

Already the simple square-well potential model suggests that there is little variation in the effective

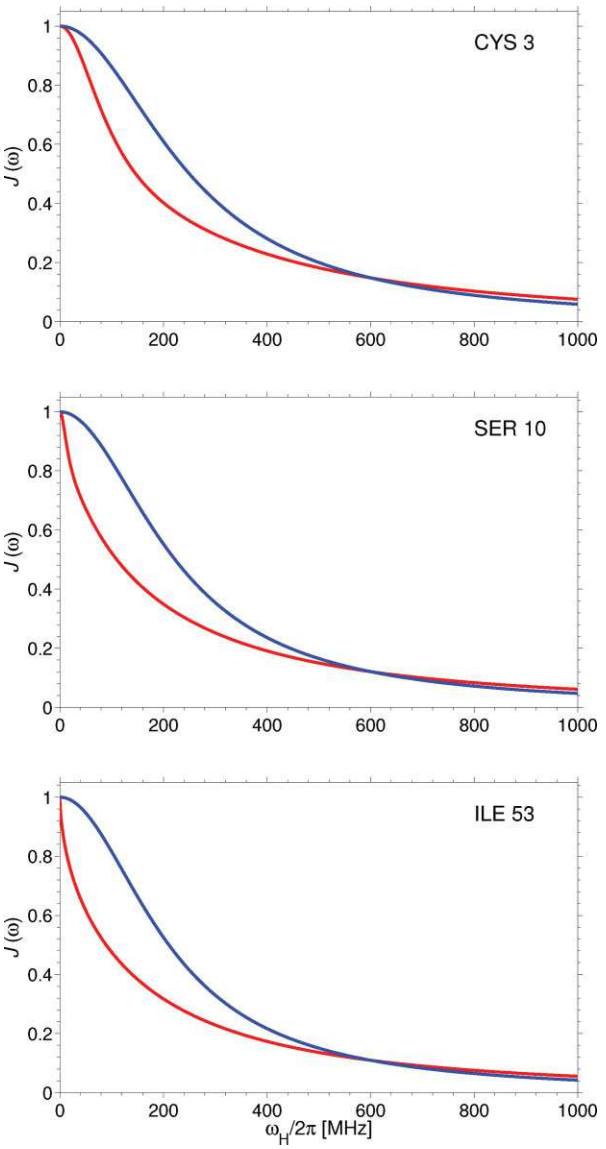

Figure 11. $J(\omega)$ profiles calculated on the basis of the square-well potential model for three H^N atoms in drkN SH3 A3C-MTSL: Cys 3, Ser 10, and Ile 53 (red curves). The simulation parameters are the same as in Figure 5. Also shown are Lorentzian spectral density profiles that reproduce the $J(\omega)$ values at 0 and 600 MHz (blue curves).

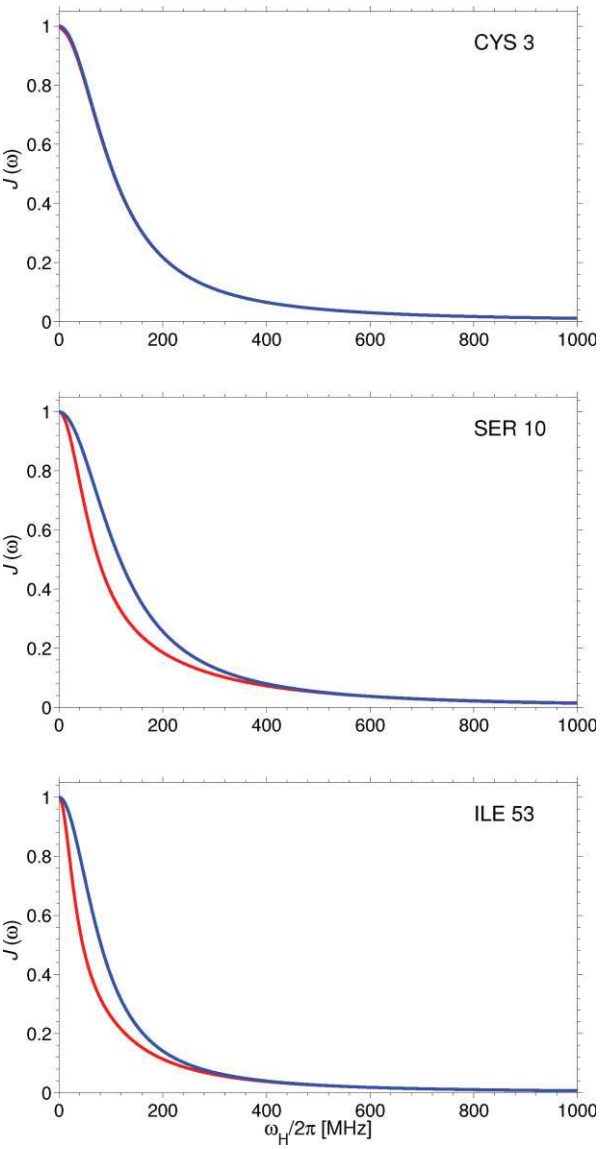

Figure 12. $J(\omega)$ profiles derived from MD trajectory of drkN SH3 A3C-MTSL (target temperature 1000 K) involving H^N atoms from Cys 3, Ser 10, and Ile 53 (red curves). Also shown are Lorentzian spectral density profiles that reproduce the $J(\omega)$ values at 0 and 600 MHz (blue curves).

correlation times along the polypeptide chain. Specifically, Eq. (10) predicts that the correlation time varies from $(1/6)\tau_{tr}$ for amide sites immediately next to the paramagnetic tag to $(4/9)\tau_{tr}$ for distant amide sites. In practice, this range is considerably more narrow since the rotational limit, $(1/6)\tau_{tr}$, is never reached (no amides are attached directly to the pyrrolinyl ring). Assuming that there is little variability in the correlation times, one may ask whether the Gillespie–Shortle model provides a reasonable empirical description of the PRE effect in unfolded proteins.

To address this question, we analyzed the results from the most realistic of our four theoretical models, that is, the MD model. The outcome is presented in Figure 13. The left part of the plot illustrates the variation of τ_c^{eff} along the peptide chain in drkN SH3 A3C-MTSL. As expected, the correlation times are lowest near the MTSL tag, 1.4 ns. Away from the N-terminus, τ_c^{eff} rises to more than 2 ns. Similar pattern is observed in all other trajectories.

Proximity to the MTSL labeling site is one factor that has (moderate) influence on the magnitude of τ_c^{eff}. In addition to that, the details of primary sequence are apparently important. The longest τ_c^{eff}, up to 2.8 ns, are observed in the vicinity of residue Lys 26. In this region the stretch of aminoacids ILKIL displays the highest Kyte-Doolittle hydropathy score of the entire sequence. The resulting clustering effect likely leads to a slowdown in segmental diffusion and consequently to long τ_c^{eff}. On the other hand, short correlation times

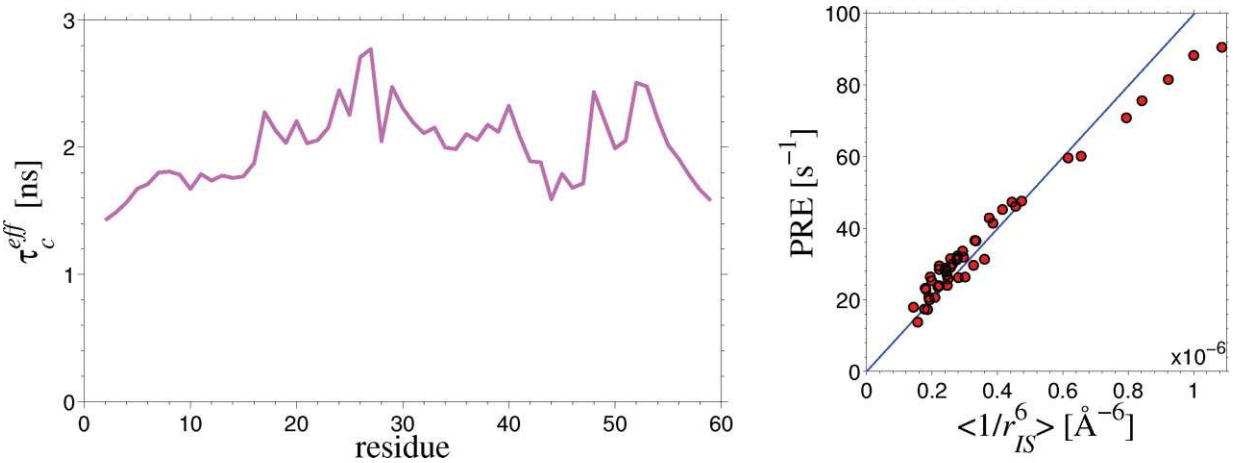

Figure 13. Left part: effective decay times of MD-derived correlation functions $g(\tau)$ in drkN SH3 A3C-MTSL at target temperature 1000 K. Obtained from single-exponential fitting of the correlation functions such as shown in Figure 4. The time scale is reset as discussed in the text so that the average τ_c^{eff} is rendered equal to 2 ns. Right part: correlation between the predicted PRE rates and $\langle 1/r_{IS}^6 \rangle$, where both parameters are extracted from the MD trajectory of drkN SH3 A3C-MTSL at 1000 K. Only those PRE rates that are lower than 100 s^{-1} are retained. The blue line corresponds to the Gillespie–Shortle model, Eqs. (1.1) and (4) with the global correlation time τ_{rot} set to 2 ns. The biggest deviation between the actual $\langle 1/r_{IS}^6 \rangle$ value and the one extracted from the Gillespie–Shortle model occurs in Lys 26 and amounts to 36%. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

on the order of 1.6 ns are found near the residue Lys 44. This is likely caused by the elevated mobility of this segment, GKEG, which is flanked by two glycine residues (these happen to be the only glycines in the sequence). Likewise, short τ_c^{eff} are found in the flexible terminal segments. These tendencies are consistently reproduced in all MD trajectories (data not shown).

While it is instructive to identify the sources of variability in τ_c^{eff}, the fact of the matter remains that the range of τ_c^{eff} is relatively narrow. Under these circumstances, it is interesting to probe the correlation between the predicted PRE rates and the distance factor $\langle 1/r_{IS}^6 \rangle$ as appears in the Gillespie–Shortle model, Eq. (4). The right part of Figure 13 shows the example of such correlation involving the PRE rates and $\langle 1/r_{IS}^6 \rangle$ extracted from the MD trajectory of drkN SH3 A3C-MTSL at 1000 K. Clearly, there is a nearly linear relationship between the two quantities, suggesting that the Gillespie–Shortle model provides an adequate empirical description for the PRE effect in unfolded proteins.

In fact, if the characteristic correlation time is known, then the Gillespie–Shortle model can be used to convert the PRE data into $\langle 1/r_{IS}^6 \rangle$. This is illustrated in Figure 13, where the blue line represents the Gillespie–Shortle formula with the global correlation time set to $\tau_{rot} = 2$ ns. Using this simple dependence, the effective distances $\langle 1/r_{IS}^6 \rangle^{-1/6}$ can be recovered with a very good accuracy. Note, however, that calculating $\langle 1/r_{IS}^6 \rangle^{-1/6}$ is not very meaningful in the case of an unfolded protein—instead, the data should be used to reconstruct the pcfs $P(r_{IS})$. The cost of the error in $\langle 1/r_{IS}^6 \rangle$ can be higher in this situation. Generally, however, we believe that the Gillespie–Shortle model is adequate for the purpose of estimating $P(r_{IS})$ (provided that the correlation time τ_{rot} is known with a reasonable accuracy).

Experimental Results

For the purpose of this study, we prepared three single-cysteine mutants of the N-terminal SH3 domain of the *Drosophila* adapter protein drk (drkN SH3): A3C, D32C, and D59C. The details of the sample preparation procedure are described in the Supporting Information. The sample conditions for drkN SH3 are 0.33 m*M* protein, 2*M* guanidinium chloride (GuHCl), 50 m*M* sodium phosphate, 90% H_2O–10% D_2O, pH 6.0. Chemical shift titration suggests that the protein is almost fully denatured in the presence of 2*M* GuHCl.[99] This conclusion is supported by the chevron plot (even in the presence of Na_2SO_4, which has a stabilizing effect on the structure).[100] All NMR measurements were conducted at 600 MHz Varian Inova spectrometer at a temperature of 5°C, where the conditions for observation were found to be optimal. The spectral assignment was obtained from the publication of Zhang and Forman-Kay[101] and confirmed by the CBCANH experiment.[102] The spectra of oxidized and reduced drkN SH3 A3C-MTSL are shown in the Supporting Information.

To assess the effect of intermolecular paramagnetic relaxation, we prepared an additional sample (A3C) with doubled protein concentration, 0.7 m*M*. The PRE rates measured in this concentrated sample are in good agreement with those obtained using the dilute sample; the contributions from increased intermolecular paramagnetic relaxation are in the range

0–2 s^{-1}, with the average of 1 s^{-1}. Given that the PRE measurements suffer from other more significant sources of error (see below), this small amount of bias can be disregarded.

Initially, we measured the PRE rates using the sequence by Donaldson *et al.*,[103] as well as the variant of this sequence published by Iwahara *et al.*[104] Both experiments are $^1H^N$ spin-echo relaxation experiments that record the magnetization decay profiles. The difference is that in the former case the effect of $^3J_{HN,H\alpha}$ coupling is eliminated by means of the $^1H^N$-selective pulse, whereas in the latter case it is eliminated via point-by-point division of the two decay curves (from paramagnetic and diamagnetic samples). The two experiments proved to be in excellent agreement and highly reproducible; the collected decay curves are high quality and can be nicely fitted with a single exponential (with the exception of several very weak peaks that belong to the residues proximal to the MTSL tag—these peaks are significantly affected by noise and occasionally show evidence of biexponential relaxation).

At the same time, we determined the PRE rates by a more conventional method based on comparison of the two HSQC maps (from paramagnetic and diamagnetic samples). This method normally relies on the following formula[63]:

$$\frac{I_{ox}}{I_{red}} = \frac{R_2^{red}}{R_2^{red} + PRE} \exp(-PRE \cdot 2\tau_{INEPT}) \qquad (19)$$

The left-hand side of this expression represents the ratio of peak intensities (i.e., peak heights) in the oxidized and reduced samples; on the right-hand side the first factor accounts for broadening of the spectral line in the direct (proton) dimension, and the second factor estimates the loss of magnetization during the two INEPT periods (duration τ_{INEPT}) when the proton magnetization is transverse.

This expression appears to be problematic in more than one way. First, one cannot assume that the line-broadening is equivalent to the relaxation rate. It would be more appropriate to use the factor $\Delta\omega_H^{red}/\Delta\omega_H^{ox}$ (i.e., the ratio of the experimentally observed proton linewidths). Indeed, $\Delta\omega_H^{red}/\Delta\omega_H^{ox}$ is different from $R_2^{red}/(R_2^{red} + PRE)$ as it contains the contributions from static magnetic field inhomogeneity, unresolved scalar couplings, or even a window function applied during the spectral processing. Second, one cannot rule out the presence of the paramagnetic line-broadening in the indirect ^{15}N domain, because of the effect of delocalized electron density,[98] which in principle should be accounted for via $\Delta\omega_N^{red}/\Delta\omega_N^{ox}$ factor. Third, the use of Eq. (19) requires the knowledge of R_2^{red}. In unfolded proteins, R_2^{red} rates show considerable variability along the peptide chain; accurate measurement of R_2^{red} requires a full-scale relaxation experiment.[103]

As an alternative to Eq. (19), we propose a simpler expression:

$$\frac{V_{ox}}{V_{red}} = \exp(-PRE \cdot 2\tau_{INEPT}) \qquad (20)$$

where V_{ox} and V_{red} are the volumes of the two respective peaks. To obtain the volumes, we used the program nlinLS[105] that fits the peaks with 2D Lorentzian or Gaussian shapes; with this approach, the peak volumes are determined with superior precision.

Note that Eq. (20) employs the ratio of the peak volumes obtained from two different samples. In this situation, one has to exercise special care to avoid the occurrence of bias. For instance, if a Shigemi tube is used in the measurements, then 5% difference in the volume of solution under the plunger translates into 5 s^{-1} error in the determined PRE rates. We have taken a number of steps to minimize potential sources of bias. In particular, a Hamilton syringe was used to add a small amount of ascorbate solution directly into the NMR tube. The volume of the sample under the plunger before and after the reduction was carefully matched. After the reduced sample was placed back into the spectrometer, it was verified that shimming, matching, and tuning of the probe, as well as the length of rf pulses, require little or no adjustment. The HSQC sequence used to quantify V_{ox} and V_{red} was fashioned from the experiment of Kay and coworkers.[103] to eliminate the effect of $^3J_{HN,H\alpha}$ coupling during the INEPT periods. The spectra were acquired using a long recycling delay, 5 s; shorter recovery delays lead to an unsatisfactory situation when certain signals recover fully in paramagnetic samples, but only partially in the reduced sample.

Finally, as an ultimate solution, an internal reference has been added to the samples. The assumption was that the signal from the reference compound should be the same in the two spectra, oxidized and reduced. For the role of the reference compound we selected *N*-acetyl-glycine (NAG), $CH_3CO^{15}NHCH_2$—COOH, which gives rise to an isolated HSQC peak (see Supporting Information). This small polar compound, added in 2:1 molar ratio with drkN SH3, does not interfere with the conformational equilibrium of the protein in the presence of a large quantity of denaturant (as confirmed by the invariance of chemical shifts). During data processing, the ratio of the NAG peak volumes was used to calculate the correction factor that was subsequently applied to all peaks in the spectra. These correction factors, $V_{ox}^{NAG}/V_{red}^{NAG}$, were found to be 1.09, 0.98, and 1.04 for A3C, D32C, and D59C samples, respectively. Clearly, even though our measurements have been conducted with extra care, the desired accuracy could not be achieved without the internal reference.

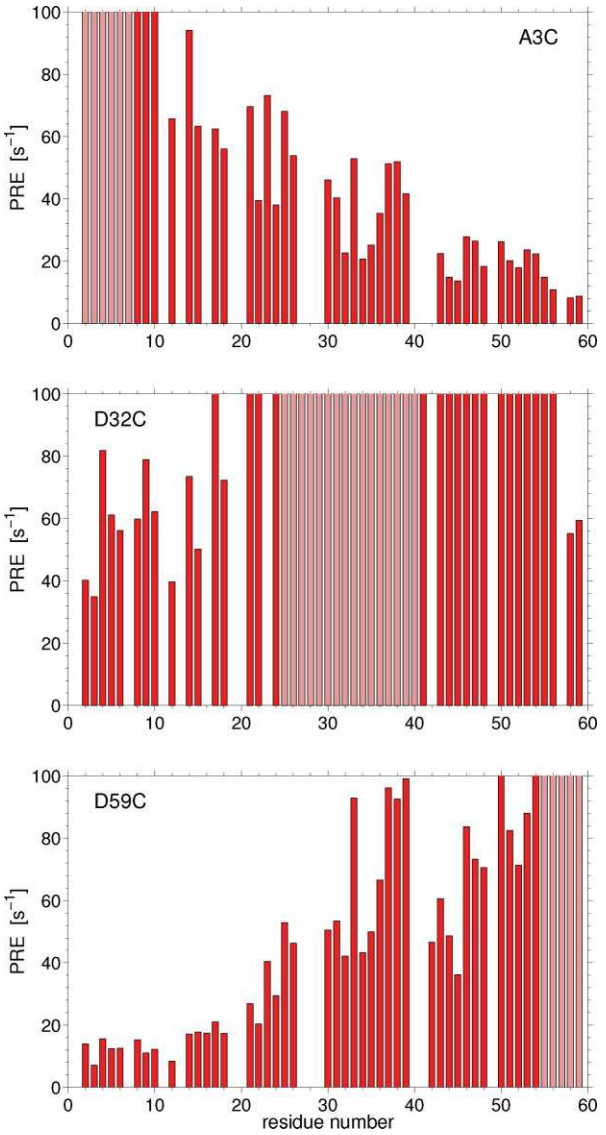

Figure 14. PRE rates as experimentally measured in the three MTSL-tagged mutants of drkN SH3 in the presence of 2*M* GuHCl. The measurements are based on peak volume ratios, Eq. (20). Pale-colored bars correspond to unobservable or extremely weak peaks that cannot be quantitatively integrated. The data in the middle panel point toward the presence of the hydrophobic cluster around residue W36.[99] The persistence of hydrophobic clusters at high denaturant concentration has been well documented.[3,111,112]

In the end, two series of experiments have been recorded: (i) spin-echo relaxation experiments yielding PRE = $R_2^{ox} - R_2^{red}$ and (ii) additional HSQC-type experiments allowing for determination of the PRE rates according to Eq. (20). How do these results compare? As it turns out, the two sets of PRE rates are fairly well correlated. For instance, the correlation coefficient $r = 0.94$ is obtained for drkN SH3 A3C-MTSL after three weakest peaks are removed from consideration. Crucially, however, the spin-echo experiment underestimates the PRE rates by approximately a factor of 2—or even worse in the case of the weakest peaks.

The reasons for this failure have been mentioned earlier.[104] As it happens, the paramagnetic sample contains a small fraction (about 10%) of diamagnetic species. As a result, the spectral peaks of the residues that are proximal to the MTSL tag are comprised of the two components: the broad, rapidly decaying component associated with paramagnetic species and the sharp, long-lived component corresponding to diamagnetic species. Although the initial amplitude of the sharp component is small, it nevertheless dominates over the strongly attenuated broad component. In this situation, the apparent relaxation rates R_2^{ox} are largely dictated by the diamagnetic "impurity," which results in dramatic underestimation of the PREs. In contrast, the volume ratio is relatively insensitive to the presence of the diamagnetic species. Indeed, even when V_{ox} represents nothing but diamagnetic impurity, the ratio V_{ox}/V_{red} turns out to be small due to the low content of diamagnetic species in the oxidized sample. As a result, the PRE rate is correctly determined to be large.

The in-depth discussion of this effect, including numeric simulations, can be found in the Supporting Information. Here we would like to summarize several observations that can be of practical interest. Our samples were presumably close to 100% MTSL-labeled, as can be judged from the chemical shifts of the tagged residues. The emergence of diamagnetic species was apparently due to partial reduction of the MTSL. It can be speculated that the reduction occurs during the labeling process (e.g., due to interaction of nitroxide radical with cysteine thiols[106] catalyzed by microquantities of metals). The experimental data also show some (tentative) evidence of slow exchange between the paramagnetic and diamagnetic species. It is possible that the exchange reflects the following reduction/reoxidation equilibrium, protein-NO• + protein-NOH ⇄ protein-NOH + protein-NO•, as described previously.[107] In principle, the equilibrium can be driven toward the

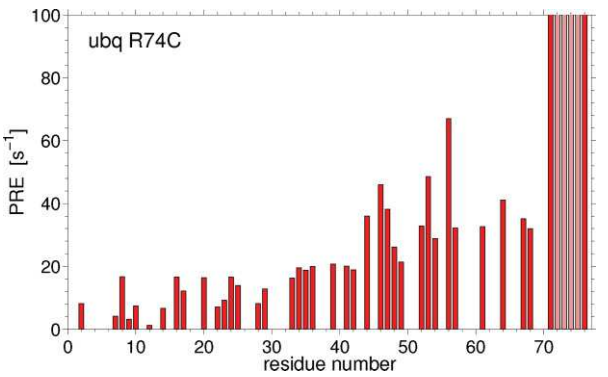

Figure 15. PRE rates as experimentally measured in the sample of R74C-MTSL ubiquitin in 8*M* urea (pH 2, protein concentration 0.35 m*M*, reference compound *N*-acetylglycine, temperature 5°C, proton frequency 600 MHz; see Supporting Information for additional information).

radical form by bubbling oxygen through the solution or by adding another, more potent, oxidant.[108] Note that similar issues arise when addition of ascorbate fails to completely reduce the sample.[109,110]

From a practical perspective, we conclude that the results of conventional relaxation experiments are severely compromised by the presence of diamagnetic species in the MTSL-tagged samples. On the contrary, the method based on peak volume determination provides reasonably accurate results. To address this problem quantitatively, we have undertaken a series of numeric simulations (see Supporting Information). As it turns out, for the PRE rates under 100 s^{-1} the approach based on Eq. (20) produces an estimated error of less than 3 s^{-1}. By comparison, in the spin-echo relaxation experiments, the error is by an order of magnitude larger. In what follows, we rely on the peak volume ratios, Eq. (20), to measure the PRE rates.

The experimental PRE profiles, Figure 14, can be readily compared with the theoretical predictions, Figures 5–10. We find it most convenient to narrow the comparison to the MD simulation results which illustrate different dynamic behaviors—from collapsed chain, Figure 8, to fully extended chain, Figure 10. (Note that, in principle, analytical models can also be used to sample different dynamic regimes—for instance, in the harmonic potential model this can be achieved by varying the parameter b).

Surveying Figure 14, we note that the experimental data from N-terminal mutant A3C are fairly similar to the theoretical profile from 1000 K MD simulation (top panel in Fig. 9). The same can be said of the C-terminal mutant, D59C (bottom panel in Fig. 9). Although the simulated profiles are "smoother" than the experimental ones, this can be perhaps attributed to an exaggerated amount of conformational averaging in the high-temperature MD simulations. The situation with the middle mutant, D32C, is somewhat different. The experimental PRE data bear more resemblance to the 700 K simulation, especially toward the C-terminal region (middle panel in Fig. 8).

Let us summarize. The experimental PRE data for the middle mutant of drkN SH3 are consistent with the compact protein structure. The data from N- and C-terminal mutants point toward the moderately expanded structure. Finally, the experimental data on radius of gyration (see Table I and related discussion) suggest that the structure is highly expanded.

How all of this evidence can be reconciled? The obvious explanation is that none of the current MD models—compact, moderately expanded, or highly expanded—provides a fully satisfactory description of the actual protein. It can be speculated, for instance, that the protein spends most of the time in the extended form, yet occasionally undergoes partial collapse involving, in particular, residues 30–55. Such a hypothetical scenario would account for the sum of experimental evidence: the dominant extended form of drkN SH3 is responsible for the large experimental R_g, whereas the appearance of the collapsed species explains the experimentally measured high PRE rates. Given that PRE rates are extremely sensitive to distance, already subtle structural preferences may be sufficient to cause the observed behavior.*

To test the generality of these observations, we also collected a limited amount of data from a sample of denatured ubiquitin. Recently, Wirmer *et al.* described ubiquitin in 8*M* urea at pH 2 as a perfect example of random-coil protein.[113] In Figure 15, we present the data acquired from the sample of ubiquitin R74C-MTSL under these experimental conditions. The data set is more sparse than in the case of drkN SH3 because of the spectral overlaps; furthermore, some of the assignments[114] could not be transferred to 5°C. Nevertheless, certain preliminary observations can be made. The PRE rates at the residues located far away from the labeling site remain substantial—on average, 10 s^{-1} for residues from 2 to 29. This is comparable to the PRE rates measured in the N-terminal region of drkN SH3 D59C-MTSL. Further similarities can be noted if one turns to the R_g data—the value of b for the two proteins is virtually identical.[10] It remains to be seen whether the same trends, especially with respect to the PRE rates, hold for other chemically denatured proteins.

Conclusions

The PRE effect is one of the most effective tools for studying unfolded proteins. Following the pioneering work of Gillespie and Shortle, the interpretation of the PRE data from unfolded proteins has relied on the conventional formulas for rotational spin relaxation. It is clear, however, that the relative motion of paramagnetic label attached to the unfolded protein and the reporter spin from the protein backbone should be generally viewed as translation rather than rotation. In this work, we developed a number of models—analytical, numerical, and MD-based—where the PRE effect is treated as translational relaxation, subject to constraints imposed by the peptide chain.

The new models lead to the appreciable difference in the predicted dependence of the PRE rates on the residue number and the static magnetic field. It can be suggested, however, that the Gillespie–Shortle approach is generally adequate when the PRE data are used to model the distance distributions $P(r_{IS})$ (provided that the characteristic correlation time of translational motion is known with reasonable accuracy). Of note, the developed theoretical tools can be applied to a number of related problems such as long-range NOEs in unfolded proteins. Another class of problems that can be treated in a similar fashion involves paramagnetic relaxation or fluorescence lifetime measurements (time-resolved FRET) in proteins consisting of multiple domains connected via flexible linkers.

*Somewhat similar arguments were put forth before in comparative analysis of R_g and NMR data.[10,11]

Experimental measurements of the PRE effect in MTSL-tagged proteins are faced with certain difficulties. Conventional relaxation experiments tend to fail because of the presence of diamagnetic impurities in the samples (apparently, a small fraction of reduced species, possibly involved in reduction/reoxidation equilibrium). The traditional method based on comparison of the spectra from oxidized and reduced samples is much more stable in this regard. In applying this method, we have devised two improvements: (i) using a reference compound to calibrate the signals in the two spectra and (ii) using peak volumes instead of intensities to calculate the PRE rates. These two modifications lead to a substantially more straightforward and reliable measurement procedure.

We also note interesting experimental opportunities arising from the use of alternative paramagnetic labels. For instance, in the case of the thiol-reactive MTS-EDTA tag loaded with a paramagnetic ion, such as Ni^{2+}, the PRE effect is controlled by the relatively short (on the order of tens of picoseconds[115]) electron-spin relaxation time $T_{1,S}$. The segmental diffusion of the peptide chain, which is much slower, essentially has no influence on paramagnetic relaxation. Under these circumstances, the interpretation of the PRE data should be straighforward.[33]

To test our theoretical constructs, we acquired PRE data for three MTSL-tagged mutants of drkN SH3 domain in 2*M* guanidinium chloride. As it turns out, some of the PRE data are consistent with the model of moderately expanded random-coil protein, whereas other data point toward the existence of more compact structure (hydrophobic cluster). At the same time, the experimentally determined radius of gyration[94] suggests that the protein is highly expanded. Eventually, it should be possible to build a dynamic model of denatured drkN SH3 that would satisfy all pieces of experimental data. Specifically, we envisage the structural ensemble that is dominated by extended conformations, yet also contains the compact (collapsed) species. This anticipated behavior is apparently more complex than predicted by the model of a random-coil protein in good solvent/poor solvent.

Acknowledgments

We are grateful to Dr. Irina Bezsonova and Prof. Julie Forman-Kay (Hospital for Sick Children, Toronto) for the generous gift of drkN SH3 plasmid and advice on protein expression and purification, Prof. Valery V. Khramtsov (Ohio State University) for consultation on free radicals, and Prof. Anatoly N. Podkorytov (St. Petersburg State University) for mathematical insights. Ubiquitin plasmid was supplied by Rachel Klevit laboratory through Addgene repository. The scripts for calculating PRE rates are available from the authors upon request.

References

1. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CR, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang CH, Kissinger CR, Bailey RW, Griswold MD, Chiu M, Garner EC, Obradovic Z (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59.
2. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6: 197–208.
3. Neri D, Billeter M, Wider G, Wüthrich K (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor. Science 257:1559–1563.
4. Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: Resolving the reconciliation problem. Proc Natl Acad Sci USA 102: 13099–13104.
5. McCarney ER, Kohn JE, Plaxco KW (2005) Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. Crit Rev Biochem Mol Biol 40:181–189.
6. Daggett V, Fersht A (2003) The present view of the mechanism of protein folding. Nat Rev Mol Cell Biol 4: 497–502.
7. Dyson HJ, Wright PE, Scheraga HA (2006) The role of hydrophobic interactions in initiation and propagation of protein folding. Proc Natl Acad Sci USA 103: 13057–13061.
8. Ptitsyn OB (1995) Molten globule and protein folding. Adv Protein Chem 47:83–229.
9. Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry 38:16424–16431.
10. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci USA 101:12491–12496.
11. Tran HT, Pappu RV (2006) Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions. Biophys J 91: 1868–1886.
12. Jennings PA, Wright PE (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science 262:892–896.
13. Schwarzinger S, Kroon GJA, Foss TR, Chung J, Wright PE, Dyson HJJ (2001) Sequence-dependent correction of random coil NMR chemical shifts. Am Chem Soc 123: 2970–2978.
14. Farrow NA, Zhang OW, Forman-Kay JD, Kay LE (1995) Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. Biochemistry 34:868–878.
15. Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM (1996) Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. J Mol Biol 255:494–506.
16. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci USA 102:17002–17007.
17. Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM (1997) Structural and dynamical properties of a denatured

protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. Biochemistry 36:8977–8991.
18. Yi Q, Scalley-Kim ML, Alm EJ, Baker D (2000) NMR characterization of residual structure in the denatured state of protein L. J Mol Biol 299:1341–1351.
19. Crowhurst KA, Choy WY, Mok YK, Forman-Kay JD (2003) Corrigendum to the paper by Mok et al. (1999). NOE data demonstrating a compact unfolded state for an SH3 domain under nondenaturing conditions. J Mol Biol 329:185–187.
20. Crowhurst KA, Forman-Kay JD (2003) Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain. Biochemistry 42: 8687–8695.
21. Chang I-J, Lee JC, Winkler JR, Gray HB (2003) The protein-folding speed limit: Intrachain diffusion times set by electron-transfer rates in denatured $Ru(NH_3)_5$ (His-33)-Zn-cytochrome c. Proc Natl Acad Sci USA 100: 3838–3840.
22. Segal MS, Bye JM, Sambrook JF, Gething MJHJ (1992) Disulfide bond formation during the folding of influenza-virus hemagglutinin. Cell Biol 118:227–244.
23. Lyubovitsky JG, Gray HB, Winkler JR (2002) Structural features of the cytochrome c molten globule revealed by fluorescence energy transfer kinetics. J Am Chem Soc 124:14840–14841.
24. Steinhoff HJ (2004) Inter- and intra-molecular distances determined by EPR spectroscopy and site-directed spin labeling reveal protein-protein and protein-oligonucleotide interaction. Biol Chem 385:913–920.
25. Gillespie JR, Shortle DJ (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. 1. Paramagnetic relaxation enhancement by nitroxide spin labels. Mol Biol 268:158–169.
26. Berliner LJ, Grunwald J, Hankovszky HO, Hideg K (1982) A novel reversible thiol-specific spin label - papain active site labeling and inhibition. Anal Biochem 119:450–455.
27. Hubbell WL, Mchaourab HS, Altenbach C, Lietzow MA (1996) Watching proteins move using site-directed spin labeling. Struct Fold Des 4:779–783.
28. Gillespie JR, Shortle D (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease .2. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. J Mol Biol 268:170–184.
29. Lietzow MA, Jamin M, Dyson HJ, Wright PE (2002) Mapping long-range contacts in a highly unfolded protein. J Mol Biol 322:655–662.
30. Teilum K, Kragelund BB, Poulsen FM (2002) Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. J Mol Biol 324:349–357.
31. Bertoncini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, Zweckstetter M (2005) Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. Proc Natl Acad Sci USA 102:1430–1435.
32. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM (2005) Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J Am Chem Soc 127:476–477.
33. Freed JH (1978) Dynamic effects of pair correlation functions on spin relaxation by translational diffusion in liquids. 2. Finite jumps and independent T_1 processes. J Chem Phys 68:4034–4037.
34. Pyka J, Ilnicki J, Altenbach C, Hubbell WL, Froncisz W (2005) Accessibility and dynamics of nitroxide side chains in T4 lysozyme measured by saturation recovery EPR. Biophys J 89:2059–2068.
35. Jun S, Becker JS, Yonkunas M, Coalson R, Saxena S (2006) Unfolding of alanine-based peptides using electron spin resonance distance measurements. Biochemistry 45:11666–11673.
36. Hyde JS, Yin JJ, Subczynski WK, Camenisch TG, Ratke JJ, Froncisz W (2004) Spin-label EPR T_1 values using saturation recovery from 2 to 35 GHz. J Phys Chem B 108:9524–9529.
37. Mailer C, Nielsen RD, Robinson BH (2005) Explanation of spin-lattice relaxation rates of spin labels obtained with multifrequency saturation recovery EPR. J Phys Chem A 109:4049–4061.
38. Lindorff-Larsen K, Kristjansdottir S, Teilum K, Fieber W, Dobson CM, Poulsen FM, Vendruscolo M (2004) Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme A binding protein. J Am Chem Soc 126:3291–3299.
39. Kristjansdottir S, Lindorff-Larsen K, Fieber W, Dobson CM, Vendruscolo M, Poulsen FM (2005) Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies. J Mol Biol 347:1053–1062.
40. Felitsky DJ, Lietzow MA, Dyson HJ, Wright PE (2008) Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. Proc Natl Acad Sci USA 105:6278–6283.
41. Hwang LP, Freed JH (1975) Dynamic effects of pair correlation functions on spin relaxation by translational diffusion in liquids. J Chem Phys 63:4017–4025.
42. Albrand JP, Taieb MC, Fries PH, Belorizky E (1983) NMR study of spectral densities over a large frequency range for intermolecular relaxation in liquids: Pair correlation effects. J Chem Phys 78:5809–5815.
43. Abragam A (1961) The principles of nuclear magnetism. Oxford: Clarendon Press.
44. Bertini I, Luchinat C, Parigi G (2001) Solution NMR of paramagnetic molecules. Amsterdam: Elsevier.
45. White GF, Ottignon L, Georgiou T, Kleanthous C, Moore GR, Thomson AJ, Oganesyan VS (2007) Analysis of nitroxide spin label motion in a protein-protein complex using multiple frequency EPR spectroscopy. J Magn Reson 185:191–203.
46. Zwanzig R (1988) Diffusion in a rough potential. Proc Natl Acad Sci USA 85:2029–2030.
47. van Kampen NG (1981) Stochastic processes in physics and chemistry. Amsterdam: North-Holland.
48. Halle B, Jóhannesson H, Venu K (1998) Model-free analysis of stretched relaxation dispersions. J Magn Reson 135:1–13.
49. Halle B (2003) Cross-relaxation between macromolecular and solvent spins: The role of long-range dipole couplings. J Chem Phys 119:12372–12385.
50. Ayant Y, Belorizky E, Alizon J, Gallice J (1975) Calculation of spectral densities for relaxation resulting from random molecular translational modulation of magnetic dipolar coupling in liquids. J Phys (Paris) 36:991–1004.
51. Sholl CA (1981) Nuclear spin relaxation by translational diffusion in liquids and solids: high and low-frequency limits. J Phys C: Solid State Phys 14:447–464.
52. MathWorks (2001) Using MATLAB. MA, Natick.
53. Machonkin TE, Westler WM, Markley JL (2002) $^{13}C\{^{13}C\}$ 2D NMR: a novel strategy for the study of paramagnetic proteins with slow electronic relaxation rates. J Am Chem Soc 124:3204–3205.

54. Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R (2003) ^{13}C direct detection experiments on the paramagnetic oxidized monomeric copper, zinc superoxide dismutase. J Am Chem Soc 125:16423–16429.
55. Ullman R (1965) Nuclear magnetic relaxation of polymer solutions. J Chem Phys 43:3161–3177.
56. Kowalewski J, Nordenskiöld L, Benetis N, Westlund P-O (1985) Theory of nuclear spin relaxation in paramagnetic systems in solution. Prog NMR Spectrosc 17: 141–185.
57. Szabo A, Schulten K, Schulten Z (1980) 1st passage time approach to diffusion controlled reactions. J Chem Phys 72:4350–4357.
58. Doi M, Edwards SF (1986) The theory of polymer dynamics. Oxford: Clarendon Press.
59. Fierz B, Kiefhaber T (2007) End-to-end vs interior loop formation kinetics in unfolded polypeptide chains. J Am Chem Soc 129:672–679.
60. Feldman HJ, Hogue CWV (2000) A fast method to sample real protein conformational space. Protein Struct Funct Genet 39:112–131.
61. Marsh JA, Baker JMR, Tollinger M, Forman-Kay JD (2008) Calculation of residual dipolar couplings from disordered state ensembles using local alignment. J Am Chem Soc 130:7804–7805.
62. Feldman HJ, Hogue CWV (2002) Probabilistic sampling of protein conformations: New hope for brute force? Protein Struct Funct Genet 46:8–23.
63. Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. Biochemistry 39:5355–5365.
64. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:65–73.
65. Murzyn K, Rog T, Blicharski W, Dutka M, Pyka J, Szytula S, Froncisz W (2006) Influence of the disulfide bond configuration on the dynamics of the spin label attached to cytochrome c. Proteins: Struct Funct Bioinf 62:1088–1100.
66. Frisch MJ (2004) Gaussian 03, Revision C. 02. Wallington, CT: Gaussian.
67. Chapra SC (2006) Applied numerical methods with MATLAB for engineers and scientists. New York: McGraw-Hill.
68. Nordio PL, General magnetic resonance theory. In: Berliner LJ, Ed. (1976) Spin labeling: theory and applications. New York: Academic Press, pp 5–52.
69. Conte SD, Deboor C (1972) Elementary numerical analysis. New York: McGraw-Hill.
70. Garcia AE, Onuchic JN (2003) Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. Proc Natl Acad Sci USA 100:13898–13903.
71. Schwieters CD, Clore GM (2001) Internal coordinates for molecular dynamics and minimization in structure determination and refinement. J Magn Reson 152: 288–302.
72. Kuszewski J, Gronenborn AM, Clore GM (1999) Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. J Am Chem Soc 121:2337–2338.
73. Clore GM, Schwieters CD (2003) Docking of protein-protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from $^{1}H^{N}/^{15}N$ chemical shift mapping and backbone ^{15}N-^{1}H residual dipolar couplings using conjoined rigid body/torsion angle dynamics. J Am Chem Soc 125:2902–2912.
74. Brüschweiler R, Roux B, Blackledge M, Griesinger C, Karplus M, Ernst RR (1992) Influence of rapid intramolecular motion on NMR cross-relaxation rates. A Molecular Dynamics study of antamanide in solution. J Am Chem Soc 114:2289–2302.
75. Xue Y, Pavlova MS, Ryabov YE, Reif B, Skrynnikov NR (2007) Methyl rotation barriers in proteins from ^{2}H relaxation data. Implications for protein structure. J Am Chem Soc 129:6827–6838.
76. Day R, Bennion BJ, Ham S, Daggett V (2002) Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. J Mol Biol 322: 189–203.
77. Mayor U, Guydosh NR, Johnson CM, Grossmann JG, Sato S, Jas GS, Freund SMV, Alonso DOV, Daggett V, Fersht AR (2003) The complete folding pathway of a protein from nanoseconds to microseconds. Nature 421: 863–867.
78. Day R, Daggett V (2005) Ensemble versus single-molecule protein unfolding. Proc Natl Acad Sci USA 102: 13445–13450.
79. Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. J Chem Phys 103:4613–4621.
80. Qiu D, Shenkin PS, Hollinger FP, Still WC (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. J Phys Chem A 101:3005–3014.
81. Dominy BN, Brooks CL (1999) Development of a generalized Born model parametrization for proteins and nucleic acids. J Phys Chem B 103:3765–3773.
82. Calimet N, Schaefer M, Simonson T (2001) Protein molecular dynamics with the generalized Born/ACE solvent model. Proteins: Struct Funct Bioinf 45:144–158.
83. Francis CJ, Lindorff-Larsen K, Best RB, Vendruscolo M (2006) Characterization of the residual structure in the unfolded state of the Δ131Δ fragment of staphylococcal nuclease. Proteins: Struct Funct Bioinf 65:145–152.
84. Choy WY, Forman-Kay JD (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. J Mol Biol 308:1011–1032.
85. Gottfried DS, Haas E (1992) Nonlocal interactions stabilize compact folding intermediates in reduced unfolded Bovine Pancreatic Trypsin Inhibitor. Biochemistry 31:12353–12362.
86. Buckler DR, Haas E, Scheraga HA (1995) Analysis of the structure of ribonuclease A in native and partially denatured states by time-resolved nonradiative dynamic excitation energy transfer between site-specific extrinsic probes. Biochemistry 34:15965–15978.
87. Guo ZY, Thirumalai D (1995) Kinetics of protein folding: nucleation mechanism, time scales, and pathways. Biopolymers 36:83–102.
88. Lapidus LJ, Eaton WA, Hofrichter J (2000) Measuring the rate of intramolecular contact formation in polypeptides. Proc Natl Acad Sci USA 97:7220–7225.
89. Lapidus LJ, Steinbach PJ, Eaton WA, Szabo A, Hofrichter J (2002) Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: testing a 1-dimensional diffusion model for peptide dynamics. J Phys Chem B 106:11628–11640.
90. Yeh IC, Hummer G (2002) Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. J Am Chem Soc 124:6563–6568.
91. Möglich A, Joder K, Kiefhaber T (2008) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation (vol 103, pg 12394, 2006). Proc Natl Acad Sci USA 105:6787–6787.

92. Doi M (1996) Introduction to polymer physics. USA: Oxford Univeristy Press.
93. Schäfer L (1999) Excluded volume effects in polymer solutions as explained by the renomralization group. Berlin: Springer.
94. Choy WY, Mulder FAA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. J Mol Biol 316:101–112.
95. Huang XM, Powers R (2001) Validity of using the radius of gyration as a restraint in NMR protein structure determination. J Am Chem Soc 123:3834–3835.
96. Garnier J, Gibrat JF, Robson B (1996) GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol 266:540–553.
97. Redfield AG (2003) Shuttling device for high-resolution measurements of relaxation and related phenomena in solution at low field, using a shared commercial 500 MHz NMR instrument. Magn Reson Chem 41:753–768.
98. Ma LX, Jørgensen AMM, Sørensen GO, Ulstrup J, Led JJ (2000) Elucidation of the paramagnetic R_1 relaxation of heteronuclei and protons in Cu(II) plastocyanin from *Anabaena variabilis*. J Am Chem Soc 122:9473–9485.
99. Crowhurst KA, Tollinger M, Forman-Kay JD (2002) Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain. J Mol Biol 322: 163–178.
100. Mok YK, Elisseeva EL, Davidson AR, Forman-Kay JD (2001) Dramatic stabilization of an SH3 domain by a single substitution: Roles of the folded and unfolded states. J Mol Biol 307:913–928.
101. Zhang OW, Forman-Kay JD (1997) NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. Biochemistry 36:3959–3970.
102. Grzesiek S, Bax A (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. J Magn Reson 99:201–207.
103. Donaldson LW, Skrynnikov NR, Choy WY, Muhandiram DR, Sarkar B, Forman-Kay JD, Kay LE (2001) Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy. J Am Chem Soc 123:9843–9847.
104. Iwahara J, Tang C, Clore GM (2007) Practical aspects of ^{1}H transverse paramagnetic relaxation enhancement measurements on macromolecules. J Magn Reson 184: 185–195.
105. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe – a multidimensional spectral processing system based on unix pipes. J Biomol NMR 6:277–293.
106. Hiramoto K, Ojima N, Kikugawa K (1997) Conversion of nitroxide radicals by phenolic and thiol antioxidants. Free Radical Res 27:45–53.
107. Nettleton DO, Morse PD, Swartz HM (1989) Exchange and shuttling of electrons by nitroxide spin labels. Arch Biochem Biophys 271:414–423.
108. Sentjurc M, Mason RP (1992) Inhibition of radical adduct reduction and reoxidation of the corresponding hydroxylamines in in-vivo spin trapping of carbon tetrachloride-derived radicals. Free Radical Bio Med 13: 151–160.
109. Bobko AA, Kirilyuk IA, Grigor'ev IA, Zweier JL, Khramtsov VV (2007) Reversible reduction of nitroxides to hydroxylamines: Roles for ascorbate and glutathione. Free Radical Bio Med 42:404–412.
110. Liang BY, Bushweller JH, Tamm LK (2006) Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. J Am Chem Soc 128:4389–4397.
111. Saab-Rincón G, Gualfetti PJ, Matthews CR (1996) Mutagenic and thermodynamic analyses of residual structure in the alpha subunit of tryptophan synthase. Biochemistry 35:1988–1994.
112. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. Science 295:1719–1722.
113. Wirmer J, Peti W, Schwalbe H (2006) Motional properties of unfolded ubiquitin: a model for a random coil protein. J Biomol NMR 35:175–186.
114. Peti W, Smith LJ, Redfield C, Schwalbe H (2001) Chemical shifts in denatured proteins: Resonance assignments for denatured ubiquitin and comparisons with other denatured proteins. J Biomol NMR 19:153–165.
115. Miller JC, Lohr LL, Sharp RR (2001) NMR paramagnetic relaxation enhancement: test of the controlling influence of ZFS rhombicity for S=1. J Magn Reson 148:267–276.

Supplemental Materials

Paramagnetic relaxation enhancements in unfolded proteins: theory and application to drkN SH3 domain.

Yi Xue,[*] Ivan S. Podkorytov,[*] D. Krishna Rao, Nathan Benjamin, Honglei Sun, Nikolai R. Skrynnikov[†]

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, USA

[*] These two authors contributed equally to this work

[†] Reprint requests to: Nikolai R. Skrynnikov, Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47907-2084, USA; e-mail: nikolai@purdue.edu; fax: 1-765-494-0239.

Diffusion in a square-well potential: *derivation of the PRE rates*

The correlation function for the dipolar interaction between the two spins is:[1]

$$g_m(\tau) = 4\pi \int d\Omega_0 r_0^2 dr_0 \int d\Omega r^2 dr \frac{Y_{2m}(\Omega_0)}{r_0^3} \frac{Y_{2m}^*(\Omega)}{r^3} P(r_0, \Omega_0, 0) P(r_0, \Omega_0, 0 \mid r, \Omega, \tau) \qquad \text{(S1)}$$

where $P(r_0, \Omega_0, 0)$ is the probability distribution describing the length and the orientation of the dipolar vector at the point in time $t = 0$; $P(r_0, \Omega_0, 0 \mid r, \Omega, \tau)$ is the conditional probability for the relative diffusion of the two spins, i.e. given the vector (r_0, Ω_0) at the point in time $t = 0$ it defines the probability that this vector will evolve into (r, Ω) by the time $t = \tau$. Other notations are as described in the text of the paper.

The conditional probability satisfies the diffusion equation:

$$\frac{\partial P(r_0, \Omega_0, 0 \mid r, \Omega, \tau)}{\partial \tau} = D_{tr} \Delta P(r_0, \Omega_0, 0 \mid r, \Omega, \tau) \qquad \text{(S2)}$$

where the operator Δ expressed in spherical coordinates reads:

$$\Delta = \frac{1}{r^2} \left[\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{\sin\theta} \frac{\partial}{\partial \theta} \sin\theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial \phi^2} \right] \qquad \text{(S3).}$$

To model the square-well potential, Eq. (S2) should be supplemented by the reflecting wall boundary conditions at $r = d_0$ and $r = L$:

$$\left. \frac{\partial P(r_0, \Omega_0, 0 \mid r, \Omega, \tau)}{\partial r} \right|_{r = d_0} = 0 \qquad \text{(S4.1)}$$

$$\left. \frac{\partial P(r_0, \Omega_0, 0 \mid r, \Omega, \tau)}{\partial r} \right|_{r = L} = 0 \qquad \text{(S4.2)}$$

We solve Eqs. (S2, S4) by the method of separation of variables.[2, 3] The following ansatz:

$$P(r_0, \Omega_0, 0 \mid r, \Omega, \tau) = R(r)\Psi(\Omega)T(\tau) \qquad \text{(S5)}$$

is substituted into Eq. (S2). This leads to elementary differential equations for $T(\tau)$ and $\Psi(\Omega)$ that can be readily solved:

$$T(\tau) = \exp(-\alpha^2 D_{tr}\tau), \qquad \text{(S6)}$$

$$\Psi(\Omega) = Y_{lm}(\Omega), \quad l = 0, 1, 2, ..., \quad m = -l, -l+1, ..., l, \qquad \text{(S7)}$$

The differential equation for $R(r)$ can be identified as a spherical Bessel equation:

$$r^2 R'' + 2rR' + (\alpha^2 r^2 - l(l+1))R = 0 \qquad \text{(S8)}$$

where α^2 is a separation constant. The boundary conditions Eq. (S4) translate into:

$$R'(d_0) = 0, \qquad \text{(S9.1)}$$

$$R'(L) = 0. \qquad \text{(S9.2)}$$

Eqs. (S8) and (S9) are an example of the Sturm-Liouville boundary value problem.[2,3] The eigenvalues α^2 of the Sturm-Liouville problem are real nonnegative numbers and the corresponding eigenfunctions R are orthogonal (see Theorem 3.3 of Ref. 3). For the problem at hand, the orthogonality is defined in relation to the following scalar product:

$$\langle f \mid g \rangle = \int_{d_0}^{L} r^2 f(r) g(r) dr \qquad \text{(S10).}$$

A general solution of Eq. (S8) can be represented as:

$$R(r) = A j_l(\alpha r) + B n_l(\alpha r) \qquad \text{(S11)}$$

where $j_l(x)$ and $n_l(x)$ are spherical Bessel functions of the first and the second kind, respectively; A and B are arbitrary constants.

Substituting Eq. (S11) into Eq. (S9) we obtain a discrete set of α that satisfy the boundary conditions. These values, enumerated as α_{lk}, are the solutions of the following equation:

$$j_l'(\alpha_{lk} d_0) n_l'(\alpha_{lk} L) - j_l'(\alpha_{lk} L) n_l'(\alpha_{lk} d_0) = 0 . \qquad \text{(S12)}$$

For every value of l, l = 0, 1, 2, ..., Eq. (S12) yields infinitely many positive α_{lk}, which are indexed in ascending order, $0 < \alpha_{l1} < \alpha_{l2} < \alpha_{l3} < \ldots$. The corresponding radial eigenfunctions are:

$$R_{lk}(r) = n_l'(\alpha_{lk} d_0)\, j_l(\alpha_{lk} r) - j_l'(\alpha_{lk} d_0) n_l(\alpha_{lk} r) . \qquad \text{(S13)}$$

As noted above, for any given value of l the functions $R_{lk}(r)$ are orthogonal in a sense of Eq. (S10). Of interest, relevant results can be found in the treatment of acoustic resonators by Kanellopoulos and Fikioris[4] (their paper misses some of the eigenvalues) and by Kokkorakis and Roumeliotis.[5]

In addition to this series of solutions, α equal to zero also satisfies the boundary conditions. In this case the only non-trivial (non-zero) eigenfunction is obtained when $l = 0$. It is, therefore, convenient to denote this root α_{00}. The radial eigenfunction corresponding to $\alpha_{00} = 0$ is a constant, $R_{00}(r) = const$. Because $\alpha_{00} = 0$ corresponds to a spherical harmonic with $l = 0$, which is also a constant, and because the time-dependent term Eq. (S6) turns into a constant as well, the resulting partial solution is $P(r_0, \Omega_0, 0 \mid r, \Omega, \tau) \equiv P_{00} = const$.

Combining P_{00} with the series of solutions arising from Eq. (S13) we obtain the following representation for $P(r_0, \Omega_0, 0 \mid r, \Omega, \tau)$:

$$P(r_0, \Omega_0, 0 \mid r, \Omega, \tau) = P_{00} + \sum_{l=0}^{\infty} \sum_{k=1}^{\infty} \sum_{m=-l}^{l} c_{lkm} R_{lk}(r) Y_{lm}(\Omega) \exp\left(-\alpha_{lk}^2 D_{tr} \tau\right) \qquad \text{(S14).}$$

Because $P(r_0, \Omega_0, 0 \mid r, \Omega, \tau)$ is a probability density it should integrate to one:

$$\int_V d\Omega r^2 dr P(r_0, \Omega_0, 0 \mid r, \Omega, \tau) = 1 \qquad \text{(S15)}.$$

Here V is the space enclosed between the two concentric spheres of radii d_0 and L. The only term in Eq. (S14) that produces nonzero contribution into the integral Eq. (S15) is P_{00}. This can be readily demonstrated by using the orthogonality of $Y_{lm}(\Omega)$ and the orthogonality of $R_{lk}(r)$ ($l = 0$). Thus the normalization condition Eq. (S15) yields the following result for P_{00}:

$$P_{00} = \frac{1}{V} = \frac{1}{\frac{4}{3}\pi(L^3 - {d_0}^3)} \qquad \text{(S16)}.$$

Note that in the limit of $\tau \to \infty$ the conditional probability density $P(r_0, \Omega_0, 0 \mid r, \Omega, \tau)$ reduces to P_{00}:

$$P(r_0, \Omega_0, 0 \mid r, \Omega, \infty) = P(r_0, \Omega_0, 0) = P_{00} \qquad \text{(S17)}.$$

Thus, in the limit of infinitely long time τ, $P(r_0, \Omega_0, 0 \mid r, \Omega, \tau)$ converges to the uniform probability distribution Eq. (S16).

On the other hand, it is required that at $\tau = 0$ the conditional probability density $P(r_0, \Omega_0, 0 \mid r, \Omega, \tau)$ is given by Dirac δ-function:[6]

$$P(r_0, \Omega_0, 0 \mid r, \Omega, 0) = \frac{1}{r^2}\delta(r - r_0)\delta(\Omega - \Omega_0) \qquad \text{(S18)}.$$

Replacing l.h.s. of Eq. (S18) with the corresponding expression from Eq. (S14), integrating the result with $R_{lk}(r)Y_{lm}^*(\Omega)$, and using the orthogonality properties of spherical harmonics and radial functions, we obtain the following results for the coefficients c_{lkm} in Eq. (S14):

$$c_{lkm} = \frac{1}{\int_{d_0}^{L} r^2 R_{lk}^2(r)dr} R_{lk}(r_0) Y_{lm}^*(\Omega_0) \qquad \text{(S19).}$$

Inserting Eqs. (S16) and (S19) into Eq. (S14) we obtain:

$$P(r_0, \Omega_0, 0 \mid r, \Omega, \tau) = \frac{1}{\frac{4}{3}\pi(L^3 - d_0^3)} + \sum_{l=0}^{\infty}\sum_{k=1}^{\infty}\sum_{m=-l}^{l} \frac{1}{\int_{d_0}^{L} r^2 R_{lk}^2(r)dr} R_{lk}(r_0) Y_{lm}^*(\Omega_0) R_{lk}(r) Y_{lm}(\Omega) \exp(-\alpha_{lk}^2 D_{tr}\tau) \qquad \text{(S20)}$$

The results Eq. (S20) and Eq. (S17) can be now substituted into Eq. (S1), and the integrals can be readily evaluated making use of the orthogonality property for spherical harmonics:

$$g(\tau) = \frac{3}{L^3 - d_0^3} \sum_{k=1}^{\infty} \frac{\left(\int_{d_0}^{L} r^{-1} R_{2k}(r)dr \right)^2}{\int_{d_0}^{L} r^2 R_{2k}^2(r)dr} \exp(-\alpha_{2k}^2 D_{tr}\tau) \qquad \text{(S21)}$$

Note that the same result is obtained for the correlation functions $g_m(\tau)$ independently of the value of *m*. Changing certain notations, $R_{2k}(d_0 x) = \rho_k(x)$ and $\alpha_{2k} d_0 = \beta_k$, and Fourier-transforming Eq. (S21) we arrive at the expression for $J(\omega)$ as listed in the text, Eq. (7.1).

Diffusion in harmonic potential: *derivation of* $J(0)$.

The diffusion in potential $U(r,\Omega)$ is described by Smoluchowski equation:

$$\frac{\partial P(r_0,\Omega_0,0\,|\,r,\Omega,\tau)}{\partial\tau} = D_{tr}\Delta P(r_0,\Omega_0,0\,|\,r,\Omega,\tau) + D_{tr}\beta\nabla[P(r_0,\Omega_0,0\,|\,r,\Omega,\tau)\nabla U(r,\Omega)] \tag{S22}$$

where $\beta = 1/k_BT$, k_B is Boltzmann constant, T is temperature, and the differential operator ∇ is:

$$\nabla = \frac{\partial}{\partial r}\mathbf{e}_r + \frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}\mathbf{e}_\phi + \frac{1}{r}\frac{\partial}{\partial\theta}\mathbf{e}_\theta \tag{S23}.$$

The knowledge of $P(r_0,\Omega_0,0\,|\,r,\Omega,\tau)$ allows for evaluation of the correlation function and further for calculation of the spectral densities. Note that spectral density at zero frequency, $J(0)$, amounts to the integral of the correlation function, cf. Eq. (2) and Eq. (S1), and can be written as follows:

$$J(0) = 4\pi\int d\Omega_0 r_0^2 dr_0 P(r_0,\Omega_0,0)\frac{Y_{2m}^*(\Omega_0)}{r_0^3}\,\tau_A(r_0,\Omega_0) \tag{S24}$$

$$\tau_A(r_0,\Omega_0) = \int d\Omega r^2 dr\int_0^\infty d\tau\frac{Y_{2m}(\Omega)}{r^3}P(r_0,\Omega_0,0\,|\,r,\Omega,\tau) \tag{S25}.$$

Szabo, Schulten, and Schulten demonstrated that Smoluchowski equation can be reduced to a differential equation for $\tau_A(r_0,\Omega_0)$.[7] This latter equation contains no dependence on time and is, therefore, easier to solve:

$$D_{tr}\Delta\tau_A(r,\Omega)-D_{tr}\beta(\nabla\tau_A(r,\Omega))(\nabla U(r,\Omega))=-\frac{Y_{2m}(\Omega)}{r^3} \quad \text{(S26).}$$

For the sake of convenience we have relabeled the variables: instead of r_0,Ω_0 the above equation uses r,Ω.

Consider now the specific problem at hand where the particle diffuses between two concentric spheres subject to a harmonic restraining potential:

$$U(r,\Omega)=\frac{1}{\beta}\frac{r^2}{2\sigma^2} \quad \text{(S27).}$$

The parameter σ is related to the root mean square length of the ideal Gaussian chain, l, as $\sigma=l/\sqrt{3}$.[7] The reflecting boundary conditions at $r=d_0$ and $r=L$, Eq. (S5), can be rewritten for $\tau_A(r,\Omega)$ as follows:

$$\left.\frac{\partial\tau_A(r,\Omega)}{\partial r}\right|_{r=d_0}=\left.\frac{\partial\tau_A(r,\Omega)}{\partial r}\right|_{r=L}=0 \quad \text{(S28)}$$

We seek the solution of Eq. (S26) in a form of the following construct:

$$\tau_A(r,\Omega)=R(r)Y_{2m}(\Omega) \quad \text{(S29).}$$

Substituting Eq. (S29) into Eq. (S26) and invoking the potential Eq. (S27) we arrive at the following differential equation for $R(r)$:

$$R''(r)+\left(\frac{2}{r}-\frac{r}{\sigma^2}\right)R'(r)-\frac{6}{r^2}R(r)=-\frac{1}{D_{tr}r^3} \quad \text{(S30)}$$

where prime denotes the derivative with respect to r as before, and the boundary conditions read:

$$R'(r)\big|_{r=d_0} = R'(r)\big|_{r=L} = 0 \tag{S31}$$

The solution of Eq. (S30) can be constructed from two linearly independent solutions of the homogeneous equation corresponding to Eq. (S30), $R_1(r)$ and $R_2(r)$, plus a particular solution of non-homogeneous equation, $A_3R_3(r)$:[8]

$$R(r) = A_1R_1(r) + A_2R_2(r) + A_3R_3(r) \tag{S32.1}$$

$$R_1(r) = \left(\frac{\sigma}{r}\right)^3 \exp\left[\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right] \tag{S32.2}$$

$$R_2(r) = I_4\left(\frac{r}{\sigma}\right)R_1(r) \tag{S32.3}$$

$$R_3(r) = I_1\left(\frac{r}{\sigma}\right)R_1(r) \tag{S32.4}$$

$$A_0 = R_1'(L)R_2'(d_0) - R_1'(d_0)R_2'(L) \tag{S32.5}$$

$$A_1 = \frac{1}{3D_{tr}\sigma}\frac{1}{A_0}(R_2'(L)R_3'(d_0) - R_2'(d_0)R_3'(L)) \tag{S32.6}$$

$$A_2 = \frac{1}{3D_{tr}\sigma}\frac{1}{A_0}(-R_1'(L)R_3'(d_0) + R_1'(d_0)R_3'(L)) \tag{S32.7}$$

$$A_3 = \frac{1}{3D_{tr}\sigma} \tag{S32.8}.$$

Here

$$I_n(z) = \int_0^z x^n \exp\left[-\frac{x^2}{2}\right] dx \tag{S33.1}$$

with explicit expressions:

$$I_1(z) = 1 - \exp\left[-\frac{z^2}{2}\right] \tag{S33.2}$$

$$I_4(z) = 3\sqrt{\frac{\pi}{2}}\,\mathrm{erf}\left[\frac{z}{\sqrt{2}}\right] - \left(z^3 + 3z\right)\exp\left[-\frac{z^2}{2}\right] \tag{S33.3}$$

Finally, before Eq. (S24) can be evaluated, one needs to define the equilibrium probability distribution:

$$P(r, \Omega, 0) = \frac{\exp\left[-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right]}{4\pi \int_{d_0}^{L} r^2 \exp\left[-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right] dr} \tag{S34}$$

Substituting Eqs. (S29, S32.1) and Eq. (S34) into Eq. (S24) and making use of the orthogonality property of spherical harmonics we obtain the result for $J(0)$, as listed in the text, Eqs. (12.1 – 12.9).

Sample preparation

The N-terminal SH3 domain of the *Drosophila* adapter protein drk (drkN SH3), was expressed and purified using the protocol adapted from Forman-Kay group.[9] Three single-cysteine mutants, A3C, D32C, and D59C, were engineered using the QuikChange kit from Stratagene. Special care was taken to obtain a high degree of MTSL labeling. Following the purification, the protein material was incubated at room temperature with 10-fold excess of dithiothreitol (DTT). The DTT was subsequently removed by ultrafiltration (Amicon, 3 kD cutoff), as the protein was transferred into the denaturing buffer (2 M GuHCl, 50 mM sodium phosphate, pH 6.0). The sample was then diluted with denaturing buffer to the concentration of 20 μM and reacted with 20-fold excess of MTSL (Toronto Research Chemicals; stored at –20 ºC as 50 mM stock solution in acetonitrile). The reaction proceeded for 12 hours on a rocking shaker at room temperature, after which the excess MTSL was removed by ultrafiltration. 2-fold excess of N-acetyl-glycine (5 μl of the stock solution of NAG) has been added to the sample as an intensity reference. As customary, the measurements were first performed on the paramagnetic sample and then repeated on the diamagnetic sample. The sample was reduced by adding 5-fold excess of ascorbic acid (3 μl of the stock solution of ascorbic acid injected into the NMR tube using a Hamilton syringe) and kept on the bench for two hours before returning it to the NMR spectrometer.

The expression and purification procedure for human ubiquitin was adapted from the work by Lazar *et al*.[10] The denaturing buffer in this case was 8 M urea, pH 2. The paramagnetic samples were reduced using 30-fold excess of ascorbic acid. Other details of the labeling procedure are the same as for drkN SH3.

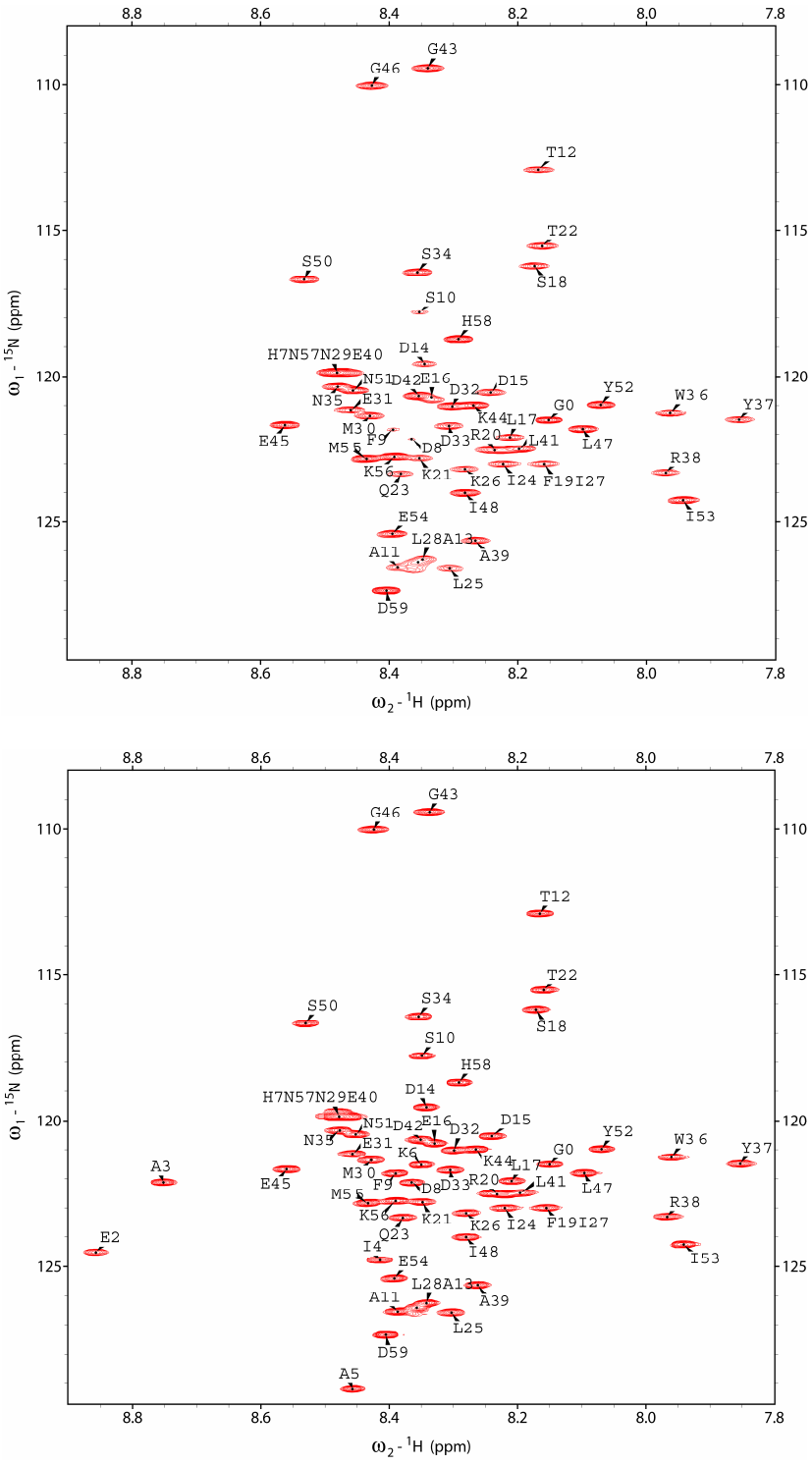

Figure S1. Spectra of drkN SH3 A3C – MTSL: oxidized (top) and reduced (bottom) samples. The resonance from the reference compound, N-acetyl-glycine, is labeled G0.

The effect of partial MTSL reduction on PRE measurements

In this section we model the effect from the presence of diamagnetic species in the presumed fully oxidized protein sample. It is assumed that the evolution of the system is governed by the simple Bloch-McConnell equation:[11]

$$\frac{d}{dt}\begin{bmatrix} M_d \\ M_p \end{bmatrix} = \begin{pmatrix} -R_d - k_{dp} & k_{pd} \\ k_{dp} & -R_p - k_{pd} \end{pmatrix} \begin{bmatrix} M_d \\ M_p \end{bmatrix} \qquad \text{(S35)}$$

where subscripts distinguish between the *d*iamagnetic and *p*aramagnetic components and the exchange rates obey the chemical balance condition, $k_{dp} p_d^0 = k_{pd} p_p^0$.

Here we seek to model a straightforward relaxation experiment, namely HSQC scheme complemented with an additional variable length spin-echo period (i.e. proton R_2 spin-echo experiment).[12, 13] Instead of reproducing the entire sequence, we focus on the portion of the sequence where proton magnetization is transverse – namely, two INEPT periods, the spin-echo element, and the acquisition time. The specific simulation algorithm is as follows. Prior to the beginning of the sequence, the system is assumed to be in equilibrium, $M_d(0) = p_d^0$ and $M_p(0) = p_p^0$. The subsequent propagation through the two INEPT periods and spin-echo period, of total duration $\tau = 2\tau_{INEPT} + \tau_{spin\text{-}echo}$, is computed by means of Eq. (S35).

In doing so the transverse proton relaxation rate for diamagnetic species is set to a certain constant value, $R_{2,d} = 15\ \text{s}^{-1}$, as estimated from our experimental data. The paramagnetic relaxation rate, on the other hand, is assumed to be dependent on the residue number. Specifically, we choose the diffusion in harmonic potential model to simulate residue-dependent paramagnetic rates, $R_{2,p} = R_{2,d} + PRE$, where PRE is the function of the spacing between the two spins, $n_{HN} - n_{MTSL}$. Finally, the exchange rate, $k_{ex} = k_{dp} + k_{pd}$, and the population of diamagnetic species, $p_d^0 = 1 - p_p^0$, are both treated as tunable parameters.

Following the evolution during $\tau = 2\tau_{INEPT} + \tau_{spin\text{-}echo}$, the surviving proton magnetization is passed over for detection (acquisition time t_2). The resulting spectral line, consisting of two Lorentzian contours with relative amplitudes $M_d(\tau)$ and $M_p(\tau)$,

can be readily reconstructed using Eq. (S35). The simulated line is then digitized with the step of 5 Hz and fitted with a *single* Lorentz contour using a procedure that mimics the treatment of the experimental data. In this manner the integral of the spectral line is obtained; in the context of the PRE measurements, this is equivalent to the determination of the peak volume.

The above simulation is repeated multiple times for a series of delays $\tau_{spin\text{-}echo}$ (the duration of the delays is the same as used in our experimental measurements). The result is a simulated decay profile, $V_{ox}(\tau_{spin\text{-}echo})$. This profile is subsequently fitted with a single exponential, yielding an apparent relaxation rate as observed in the oxidized sample, R_2^{ox}. As discussed in the text, this apparent rate can be dramatically different from the target rate, $R_{2,p}$. The fundamental cause of this discrepancy is that the data analysis protocol mistakes the small residual peak associated with diamagnetic 'impurity' for a signal of interest (i.e. a signal from paramagnetic protein).

As an alternative to a full-fledged relaxation experiment,[12, 13] we also modeled a simple measurement scheme which is commonly used for measuring the PRE rates in unfolded proteins.[14, 15] In this scheme only one spectral plane per sample is recorded, $\tau_{spin\text{-}echo} = 0$. The volumes of the peaks from the oxidized and reduced samples are then used to determine the paramagnetic relaxation enhancements, $PRE = (-1/2\tau_{INEPT})\ln(V_{ox}/V_{red})$ (note the difference with the standard procedure that relies on peak intensities). This approach is much more forgiving with respect to the effect of diamagnetic 'impurity'. Indeed, even when V_{ox} represents nothing but diamagnetic impurity, the ratio V_{ox}/V_{red} turns out to be small due to the low content of diamagnetic species. As a result, the PRE rate is correctly predicted to be large.

The relationship between the two above measurement schemes is illustrated in Fig. S2. The top section of the plot shows our experimental data from the sample of drkN SH3 A3C – MTSL in 2 M GuHCl. The left panel displays the PRE rates derived from the ratio V_{ox}/V_{red}, Eq. (20) in the text. The right panel, on the other hand, represents the results of the full-fledged relaxation measurement scheme, $PRE = R_2^{ox} - R_2^{red}$. The results are, clearly, at odds; especially striking is the finding of the slowly relaxing peaks in the vicinity of the paramagnetic label, upper right panel in Fig. S2. A similar pattern has been observed in other mutants of drkN SH3 and in ubiquitin (results not shown).

The lower portion of the plot Fig. S2 demonstrates that the experimentally observed trends can be fairly well reproduced by means of the numeric simulation. The simulation does not aspire to be quantitatively accurate. First, the chosen diffusion in

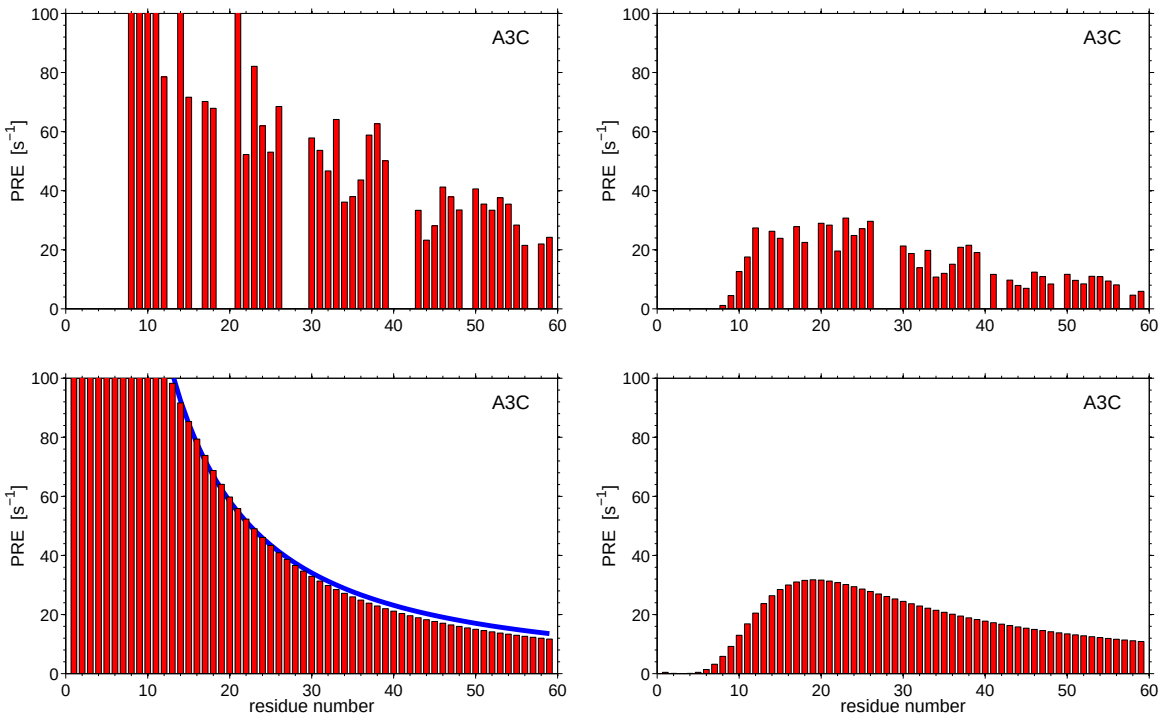

Fig. S2. Comparison of the two methods for PRE determination: based on the ratio of peak volumes V_{ox}/V_{red} (left column) and on the full-fledged proton R_2 relaxation experiment (right column). The top portion of the plot shows the experimental data from the sample of drkN SH3 A3C – MTSL. The data plotted in the upper left corner have been acquired in 32 hrs (two HSQC spectra using recycling delay 5 s). The data plotted in the upper right corner have been acquired in 70 hrs (seven points per decay curve, ranging from 7 to 32 ms in the case of the oxidized sample and from 7 to 60 ms in the case of the reduced sample). The bottom portion of the plot shows the outcome of the simulations as described in the Supporting materials. In the simulations the diamagnetic relaxation rate was set to a generic constant value, $R_{2,d} = 15\ \mathrm{s}^{-1}$. The target PRE rates (blue line in the plot) have been simulated using the diffusion in harmonic potential model with $b = 5$ Å, $d_0 = 4$ Å, $\tau_{tr} = d_0^2 / D_{tr} = 2$ ns. Other simulation parameters were $p_d^0 = 0.15$, $k_{ex} = 0$.

harmonic potential model is approximate and involves some arbitrarily assigned parameters. Second, the diamagnetic relaxation rate was set, for simplicity, to a generic constant, $R_{2,d} = 15\ \mathrm{s}^{-1}$. Third, the experimental data offer some evidence of slow exchange, $k_{ex} \sim 1\ \mathrm{s}^{-1}$ (specifically, modest broadening of the diamagnetic impurity peaks in the vicinity of the paramagnetic label). The simulations using non-zero exchange rates

led to PRE profiles in reasonable agreement with the experimental data (not shown). In the simulations Fig. S2, however, we prefer to use a minimalistic set of parameters and hence set $k_{ex} = 0$.

While exact details pertaining to the diamagnetic 'impurities' and the purported reduction / re-oxidation equilibrium remain unknown, it is clear that these impurities are responsible for the failure of the relaxation measurement scheme illustrated in the right half of Fig. S2. The alternative scheme, on the other hand, proves to be fairly resistant to these effects. In particular, there is only a minimum deviation between the rates extracted on the basis of the V_{ox}/V_{red} ratio (red bars) and the target PRE values (blue curve), as shown in the lower left panel of Fig. S2. This result serves as a justification for the measurement scheme chosen in our study.

References

(1) Hwang, L. P.; Freed, J. H. *J. Chem. Phys.* **1975,** *63*, 4017-4025.
(2) Courant, R.; Hilbert, D., *Methods of Mathematical Physics*. Interscience: New York, 1953.
(3) Cain, G.; Meyer, G. H., *Separation of variables for partial differential equations: an eigenfunction approach.* Chapman & Hall / CRC: Boca Raton, 2006.
(4) Kanellopoulos, J. D.; Fikioris, J. G. *J. Acoust. Soc. Am.* **1978,** *64*, 286-297.
(5) Kokkorakis, G. C.; Roumeliotis, J. A. *J. Sound Vib.* **1997,** *206*, 287-308.
(6) Abragam, A., *The Principles of Nuclear Magnetism.* Clarendon Press: Oxford, 1961.
(7) Szabo, A.; Schulten, K.; Schulten, Z. *J. Chem. Phys.* **1980,** *72*, 4350-4357.
(8) Polyanin, A. D.; Zaitsev, V. F., *Handbook of exact solutions for ordinary differential equations*. 2-nd Edition, Chapman & Hall / CRC: Boca Raton, 2003.
(9) Bezsonova, I.; Singer, A.; Choy, W. Y.; Tollinger, M.; Forman-Kay, J. D. *Biochemistry* **2005,** *44*, 15550-15560.
(10) Lazar, G. A.; Desjarlais, J. R.; Handel, T. M. *Protein Sci.* **1997,** *6*, 1167-1178.
(11) McConnell, H. M. *J. Chem. Phys.* **1958,** *28*, 430-431.
(12) Donaldson, L. W.; Skrynnikov, N. R.; Choy, W. Y.; Muhandiram, D. R.; Sarkar, B.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **2001,** *123*, 9843-9847.
(13) Iwahara, J.; Tang, C.; Clore, G. M. *J. Magn. Reson.* **2007,** *184*, 185-195.
(14) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997,** *268*, 158-169.
(15) Battiste, J. L.; Wagner, G. *Biochemistry* **2000,** *39*, 5355-5365.