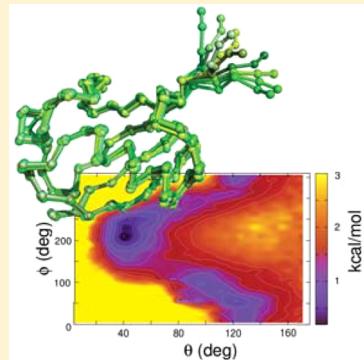


# Coarse-Grained Langevin Equation for Protein Dynamics: Global Anisotropy and a Mode Approach to Local Complexity

J. Copperman and M. G. Guenza\*

Department of Chemistry and Biochemistry and Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403, United States

**ABSTRACT:** We utilize a multiscale approach where molecular dynamic simulations are performed to obtain quantitative structural averages used as input to a coarse-grained Langevin equation for protein dynamics, which can be solved analytically. The approach describes proteins as fundamentally semiflexible objects collapsed into the free energy well representing the folded state. The normal-mode analytical solution to this Langevin equation naturally separates into global modes describing the fully anisotropic tumbling of the macromolecule as a whole and internal modes which describe local fluctuations about the folded structure. Complexity in the configurational free-energy landscape of the macromolecule leads to a renormalization of the internal modes, while the global modes provide a basis set in which the dipolar orientation and global anisotropy can be accounted for when comparing to experiments. This simple approach predicts the dynamics of both global rotational diffusion and internal motion from the picosecond to the nanosecond regime and is quantitative when compared to time correlation functions calculated from molecular dynamic simulations and in good agreement with nuclear magnetic resonance relaxation experiments. Fundamental to this approach is the inclusion of internal dissipation, which is absent in any rigid-body hydrodynamical modeling scheme.



## 1. INTRODUCTION

The dynamics of proteins, and the related biological function, are inherently determined by the complexity of their energy landscape. As proteins mostly populate the energy states at the minimum of a free-energy well, the dynamics in these states mainly consist of local fluctuations around a nontrivial three-dimensional folded structure. This justifies the application of linear Langevin equations, such as the Rouse–Zimm approach, that were originally developed to describe viscoelastic relaxation in synthetic polymer systems to describe the dynamics of the protein.<sup>1–3</sup> In a previous paper, the structure of the Langevin equation was modified to account (i) for the specific effective friction of each amino acid, which all have variable shapes and (ii) the dynamics inside the hydrophobic core of the protein, which is screened from the solvent and as such is not affected by the presence of the solvent-mediated hydrodynamic interaction.<sup>3</sup> In this current work, we develop the theory, which we name Langevin equation for protein dynamics (LE4PD), further, focusing in particular on the treatment of anisotropic rotational decorrelation and a mode approach to the local complexity in the folded free-energy well. The theory uses short-time molecular dynamics (MD) simulations for the input parameters to the analytical solution of the Langevin dynamics, and as a test of the theoretically predicted dynamical properties, (i.e., time correlation functions) of the protein. Theoretical predictions are then directly compared to experimental data of Nuclear Magnetic Resonance (NMR) relaxation, providing a test for the relation among the theoretical model, the MD simulation, and the experiments.

The LE4PD approach represents the protein as a collection of interacting units centered on the  $\alpha$  carbon and relies on

having  $N(N - 1)/2$  pairwise structural bond correlations (where  $N$  is the number of residues of the protein) and  $N$  site-specific friction parameters. Bond correlation and friction parameters are extracted from relatively short simulation trajectories of  $\sim 10$  ns or less, and with them our approach can obtain the equilibrium dynamics of processes such as global tumbling and large-amplitude internal motion, which occur in the same time regime as the simulation or longer.

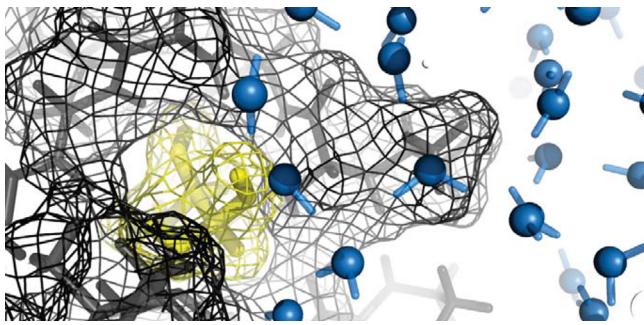
The approach is based on the fundamental picture of proteins as heterogeneous polymers, which are collapsed into a definite tridimensional structure, which nevertheless retains some amount of flexibility. As opposed to a rigid body, where the global modes are the only degrees of freedom in the system, protein dynamics include both rotational and internal fluctuation modes. Our description includes internal dissipation due to fluctuations in the hydrophobic region by accounting for an effective protein internal viscosity and considering the relative exposure of each amino acid to the hydrophobic region (see Figure 1). We show that with the correct dissipation, the linear modes of harmonically coupled objects provide a simple but accurate description of the fluctuations of the molecule.

Because fluctuations occur in the energy well, they can be conveniently approximated by harmonic potentials, and the related intramolecular distribution of sites (e.g.,  $\alpha$ -carbons) can be well-approximated by a Gaussian distribution, to at least the

**Special Issue:** Branka M. Ladanyi Festschrift

**Received:** September 18, 2014

**Revised:** October 28, 2014



**Figure 1.** Surface of a tagged residue (threonine7) in yellow, which is in contact with the internal protein environment (black) and the external solvent environment.

first order. These harmonic interactions are not restricted to bonded pairs along the backbone and construct a network of interactions which are nonlocal along the sequence.

In this way, our approach takes advantage of the same physical principles that motivate the gaussian network models (GNM) and the normal mode analysis (NMA);<sup>4–6</sup> however, it differs from those models in some important aspects. (1) In Gaussian network models,<sup>4–6</sup> the interaction between sites is taken to be a uniform harmonic interaction as long as site separation is within a distance cutoff; this cutoff is typically  $\sim 0.7$  nm, and the interaction strength is adjustable. In the LE4PD theory, all sites are considered to have a pairwise interaction potential whose parameters are defined by the structural correlations obtained from simulations. (2) The GNM and NMA models calculate fluctuations starting from a single equilibrium structure, which is determined experimentally. One important question is how well the experimental starting structure, usually determined from a crystal phase, is representative of the ensemble of protein structures that are present in solution, in physiological conditions.<sup>7</sup> The LE4PD method uses configurations generated from simulations of the protein in aqueous solutions at physiological conditions, which is a more realistic representation of the protein's configurational ensemble.<sup>8</sup> As the LE4PD model aims at building a direct connection between protein structure and their dynamics as measured experimentally for example by NMR relaxation experiments, the representation of the protein from simulations is more realistic. (3) The LE4PD is a diffusive dynamical description with site-specific dissipation, hydrodynamic coupling, and barriers to internal fluctuations, calculated directly from the structural ensemble created by the simulation of the protein in aqueous solvent. The hydrodynamic interaction is key to this description, which is conventionally neglected in the GNM and NMA approaches, which describe collective vibrational fluctuations.

In a nutshell, when comparing the LE4PD to GNM-type approaches, the advantage of the latter is their ability to qualitatively represent the type of fluctuations displayed by biomolecules within a formally simple description which requires only limited computational power, while our approach is more computationally intensive. The limitations of GNM-type approaches are related to their difficulty to obtain accurate and quantitative values of the energetics and dynamical properties from their normal modes of fluctuation, as the motion sampled by these methods is confined to the fluctuations around one equilibrium structure determined experimentally, and the amplitude of the harmonic fluctuations

is fictitious. Furthermore, the GNM-type models have limited information on the free-energy barriers along the paths of important fluctuations,<sup>9</sup> and so the kinetics and the time-dependent phenomena are not accurately determined. In the LE4PD approach, the knowledge of the roughness of the free-energy landscape (i.e., the sampled energy barriers) provides information on the long-time dynamics. With the use of the mode formalism presented in this paper, the LE4PD includes a microscopic description of the energetics and barrier crossing present in the molecular dynamic simulation.

In this paper, the ubiquitin protein is our primary model system, though the approach has been tested on other proteins as well. Ubiquitin is a small well-folded protein whose structure and dynamics have been well-characterized utilizing a number of experimental and computational techniques, including NMR backbone relaxation.<sup>10–12</sup> In the first section of this paper, we will discuss the general properties of the solution of the Langevin equation for biological polymers. Molecular Dynamics Simulations of Ubiquitin presents the methodology we used to perform molecular dynamic simulations of the protein in the canonical ensemble. Treatment of the rotational dynamics to account for the anisotropic shape of the protein is presented in Global Tumbling Modes and Internal Fluctuations. A mode-specific dynamical renormalization of the internal modes based on the free-energy surface sampled in the simulation is described in Dynamical Renormalization of the Internal Modes from Free Energy Surfaces. A method to calculate the dynamics of any bond in the protein in terms of the LE4PD solution, necessary to obtain the N–H bond dynamics probed in experimental NMR backbone relaxation, is presented in Dynamics of the N–H Dipole Vector. A discussion of the newly presented method to calculate long-time protein dynamics from shorter-time simulations concludes the paper.

## 2. THEORETICAL APPROACH: LANGEVIN EQUATION FOR PROTEIN DYNAMICS

The Langevin equation formalism for the coarse-grained representation of biological macromolecules is obtained by formally projecting all atomic degrees of freedom of both solute and solvent onto the coordinates of the coarse-grained sites.<sup>13</sup> For the dynamics of proteins, studied here, the chosen coarse-grained sites are the  $\alpha$ -carbon ( $C_\alpha$ ) of each amino acid in the primary sequence. These coarse-grained variables contain the relevant slow degrees of freedom involved in protein bond orientational relaxation. If all the needed information is contained in those variables, the system memory is small and can be neglected without affecting the precision of the calculation. A covariance matrix analysis has shown that the  $C_\alpha$  positions span the essential fluctuations of proteins.<sup>14</sup> The inertial term in the Langevin equation is also negligible as the protein dynamics in aqueous solvent are generally overdamped. All beads interact with a pair-specific harmonic potential. Detailed derivation of this potential can be found in our previous work.<sup>3,15</sup> The formalism presented in this paper can be applied with little modification to any choice of coarse-grained sites, whether they are atomistic in nature or fictitious (e.g., center of mass sites).

Applying the projection operator technique, the resulting simplified Langevin equation represents the balance of viscous dissipation with the intramolecular force and the random Brownian force due to the surrounding solvent. The time evolution of the coordinate of the site  $i$  is well-described by the following equation

$$\bar{\zeta} \frac{\partial \vec{R}_i(t)}{\partial t} = -\frac{3k_B T}{l^2} \sum_{j,k} H_{ij} A_{jk} \vec{R}_k(t) + \vec{F}_i^{\text{random}}(t) \quad (1)$$

where  $k_B$  is the Boltzmann constant,  $T$  is the temperature,  $l^2$  is the squared bond distance, bond distance, and  $\bar{\zeta}$  is the average monomer friction coefficient, defined as  $\bar{\zeta} = N^{-1} \sum_{i=1}^N \zeta_i$ , with  $\zeta_i$  the friction of the monomer  $i$ .  $\vec{F}_i^{\text{random}}(t)$  is a delta-correlated random force due to projecting the system dynamics onto the coarse-grained sites, where fluctuation-dissipation yields  $\langle \vec{F}_i^{\text{random}}(t) \cdot \vec{F}_j^{\text{random}}(t') \rangle = 2k_B T \zeta_i \delta(t - t') \delta_{ij}$ . The matrix  $H$  is the preaveraged hydrodynamic interaction matrix, which describes the interaction between protein sites occurring through the liquid, described as a continuum medium. Equation 1 is the well-known Rouse-Zimm equation for the dynamics of polymers in solution.<sup>16,17</sup>

We assume a well-folded state where site-site correlations are Gaussian in nature, and the potential of mean force is derived from the standard structural  $A$  matrix built from the set of pairwise bond correlations. Though this assumption seems radical, this is no more than the claim that the fluctuations of the macromolecule are spanned by their pairwise covariances; we will account for the complexity of the internal free-energy landscape and non-Gaussian effects *a posteriori* by rescaling of the local effective friction coefficients. The structural force matrix is calculated as

$$\mathbf{A} = \mathbf{M}^T \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{U} \end{pmatrix} \mathbf{M} \quad (2)$$

where  $\mathbf{M}$  is the matrix that defines the center of gyration and the connectivity between sites,  $\sum_j M_{ij} \vec{R}_j = \vec{l}_i$ . In a protein the  $\alpha$ -carbons are connected linearly, so that for  $i > 1$  the matrix is defined as  $M_{i,i-1} = -1$  and  $M_{i,i} = 1$ , with  $i = 2, \dots, N$ , while  $M_{1,1} = 1/N$  for the first row, and  $M_{i,j} = 0$  otherwise. Also, given that  $\vec{l}_i$  is the bond vector between site  $i - 1$  and site  $i$ , the matrix of bond-bond correlation is defined as

$$(\mathbf{U}^{-1})_{ij} = \frac{\langle \vec{l}_i \cdot \vec{l}_j \rangle}{\langle |\vec{l}_i|^2 \rangle \langle |\vec{l}_j|^2 \rangle} \quad (3)$$

The components of the  $\mathbf{U}^{-1}$  matrix are calculated directly from the simulation trajectory.

A pairwise nonlocal hydrodynamic interaction has been derived with use of the Oseen tensor.<sup>3</sup> The elements in the matrix of the hydrodynamic interaction are defined as

$$H_{ij} = \frac{\bar{\zeta}}{\zeta_i} \delta_{ij} + (1 - \delta_{ij}) \bar{r}^w \left\langle \frac{1}{r_{ij}} \right\rangle \quad (4)$$

where  $\bar{r}^w = N^{-1} \sum_{i=1}^N r_i^w$  is the average hydrodynamic radius which is defined below. This is a perturbative hydrodynamic interaction accounting for the nature of the amino acid primary structure as a heteropolymer made up of building blocks of different chemical types. Figure 1 illustrates how each amino acid is exposed in varying amounts to the external solvent environment and to the internal hydrophobic core of the protein. The site-specific friction parameters,  $\zeta_i$ , are obtained by calculating the solvent-exposed surface area during the course of the simulation, and calculating the total friction of the  $i$ th site via a simple extension of Stoke's law as

$$\zeta_i = 6\pi(\eta_w r_i^w + \eta_p r_i^p) \quad (5)$$

Here  $\eta_w$  and  $r^w$  denote, respectively, the viscosity of water and the radius of a spherical bead of identical surface area as the solvent-exposed surface area of the residue, the hydrodynamic radius.<sup>3</sup> The internal viscosity is  $\eta_p$ , and  $r^p$  denotes the hydrodynamic radius related to the surface not exposed to the solvent. To approximate the protein internal viscosity, we renormalize the solvent viscosity by a numerical factor that is calculated from the energy barriers to the internal motions determined from the simulations, as discussed by Sagnella, Straub, and Thirumalai<sup>18</sup> and described more extensively in Dynamical Renormalization of the Internal Modes from Free Energy Surfaces below.

In the hydrodynamic interaction,  $\langle 1/r_{ij} \rangle$  is the preaveraged distance between two sites in the chain, which has been shown to be a good representation of the hydrodynamic interaction for polymers in solution in standard thermodynamic conditions.<sup>17,19</sup> The preaveraging of the Oseen tensor is necessary to maintain the linear form of the Langevin equation; however, it can lead to unphysical, negative eigenvalues for physically allowed conformations.<sup>20</sup> For ubiquitin, this problem does not occur, and we adopt the average radius,  $\bar{r}^w = 0.23$  nm, calculated from the solvent-exposed surface area of each residue in the protein as measured from simulations. In those cases, for which the use of the preaveraging approximation leads to negative eigenvalues, we adopt instead the largest possible value of  $\bar{r}^w$  that maintains a positive definite solution. Typically, the two values are very close numerically. Finally, we want to point out that this issue has no consequences on the determination of the site-specific friction contributions.

Because we focus only on the bond orientational dynamics and not translation, in the interest of a simpler notation, we separate out the zeroth order translational mode from the internal dynamics. Following the same notation introduced for the orientational dynamics of star polymers,<sup>21</sup> we define  $\mathbf{a}$  as the  $\mathbf{M}$  matrix after suppressing the first row used to define the center of mass, and define  $\mathbf{L} = \mathbf{a} \mathbf{H} \mathbf{a}^T$ . The orientational Langevin equation governing the bond dynamics is

$$\frac{\partial \vec{l}_i(t)}{\partial t} = -\sigma \sum_{j,k} L_{ij} U_{jk} \vec{l}_k(t) + \vec{v}_i^{\text{random}}(t) \quad (6)$$

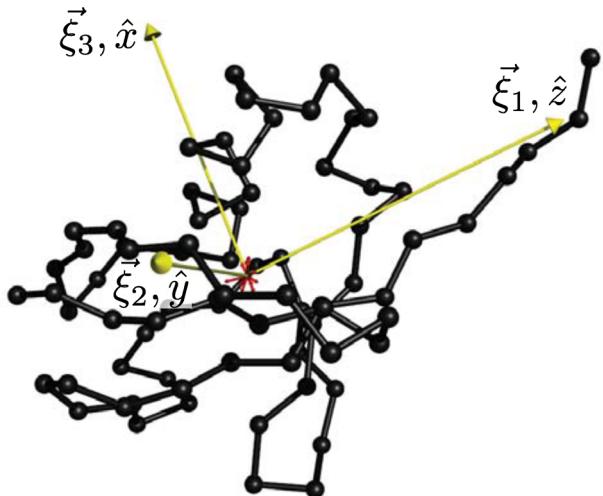
with  $i, j = 1, \dots, N - 1$ , and where  $\sigma = 3k_B T / (l^2 \bar{\zeta})$  and  $\vec{v}_i^{\text{random}}(t)$  is the random delta-correlated bond velocity.

Equation 6 represents a set of  $N - 1$  first-order coupled differential equations, which are solved by finding the matrix of eigenvectors  $\mathbf{Q}$  which diagonalizes the product of matrices  $\mathbf{L}\mathbf{U}$ . In these normal modes, we have  $N - 1$  uncoupled linear diffusion equations

$$\frac{\partial \vec{\xi}_a(t)}{\partial t} = -\sigma \lambda_a \vec{\xi}_a(t) + \vec{v}_a^{\text{random}}(t) \quad (7)$$

with  $a = 1, \dots, N - 1$ , where each mode is just a linear sum of the original bond vector basis  $\vec{\xi}_a(t) = \sum_i Q_{ai} \vec{l}_i(t)$ . We define  $\lambda_a$  to be the eigenvalues of  $\sum_{i,j,k} Q_{ai}^{-1} L_{ij} U_{jk} Q_{kb} = \delta_{ab} \lambda_a$ , and  $\mu_a$  as the eigenvalues of the bond correlation matrix alone  $\sum_{i,j} Q_{ai}^{-1} U_{ij} Q_{jb} = \delta_{ab} \mu_a$ . The  $\mathbf{Q}$  matrix is a set of eigenvectors which spans the  $N - 1$ -dimensional space of bond vectors.

Like the set of bond vectors  $\vec{l}_i(t)$ , the set of normal coordinates  $\vec{\xi}_a(t)$  define the instantaneous conformation of the macromolecule. The normal mode basis spans the same space as the bond vector basis with the added benefit of linearity:  $\langle \vec{\xi}_a(t) \cdot \vec{\xi}_b(t) \rangle = \delta_{ab} l^2 / \mu_a$ . Figure 2 shows a snapshot of the



**Figure 2.** A snapshot of the Ubiquitin protein, represented in the bond basis (black) and in the mode basis (global modes in yellow, internal modes in red).

protein ubiquitin in the bond vector and in the mode vector representation, which is obtained by applying the linear transformation of the inverse eigenvector matrix to the coordinates of the snapshot of ubiquitin. It shows how the first three normal modes, which are the slowest ones, are much larger in magnitude than the internal modes, indicating that the dynamics of this protein, at least in the simulation runs, largely conserves the shape of the molecule, while fluctuations do not involve large conformational transitions or slow cooperative domain motion.

**2.1. Local Dynamics.** The physical quantities of interest in this paper are the bond autocorrelation function and the second-order Legendre polynomial of the time-dependent bond orientation.

For each bond  $i$  along the backbone of the protein, the bond autocorrelation function is defined as

$$M_{1,i}(t) = \frac{\langle \vec{l}_i(t) \cdot \vec{l}_i(0) \rangle}{\langle l_i^2 \rangle} \quad (8)$$

and in the formalism of the Langevin equation

$$\begin{aligned} M_{1,i}(t) &= \sum_{a=1}^{N-1} \frac{Q_{ia}^2}{\langle l_i^2 \rangle} \langle \vec{\xi}_a(t) \cdot \vec{\xi}_a(0) \rangle = \sum_{a=1}^{N-1} A_a^i \exp[-\sigma \lambda_a t], \\ &= \sum_{a=1}^{N-1} A_{ia} \exp[-t/\tau_a], \end{aligned} \quad (9)$$

with  $\tau_a$  the correlation time for the  $a$ th mode.

Another quantity of interest is the second-order Legendre polynomial of the time-dependent bond orientation function  $P_2(t) = (3/2)\langle \cos^2[\theta(t)] \rangle - (1/2)$  which can be related to the first-order bond autocorrelation by

$$P_{2,i}(t) = 1 - 3 \left[ x^2 - \frac{2}{\pi} x^3 \left( 1 - \frac{2}{\pi} \arctan x \right) \right] \quad (10)$$

which is a function of  $M_{1,i}(t)$  as

$$x = \frac{[1 - M_{1,i}(t)^2]^{1/2}}{M_{1,i}(t)} \quad (11)$$

This expression relies on assuming a Gaussian form for the joint probabilities in normal mode coordinates, and it was derived in the paper by Perico and Guenza.<sup>1</sup> The theory accounts for the distribution of effective bond lengths between two sites, here alpha carbons, but does not explicitly account for anisotropy in the global modes. For dipolar relaxation, the Fourier transform of  $P_{2,i}(t)$  defines the spectral density

$$J(\omega) = \frac{2}{5} \int_0^\infty P_2(t) \cos[\omega t] dt \quad (12)$$

from which spin-lattice ( $T_1$ ) and spin-spin ( $T_2$ ) relaxation times, and nuclear Overhauser effect (NOE) as measured in NMR can be calculated as

$$\begin{aligned} \frac{1}{T_1} &= \frac{d^2}{4} [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] \\ &+ c^2 J(\omega_N) \end{aligned} \quad (13)$$

$$\begin{aligned} \frac{1}{T_2} &= \frac{d^2}{8} [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) \\ &+ 6J(\omega_H + \omega_N)] + \frac{c^2}{6} [3J(\omega_N) + 4J(0)] \end{aligned} \quad (14)$$

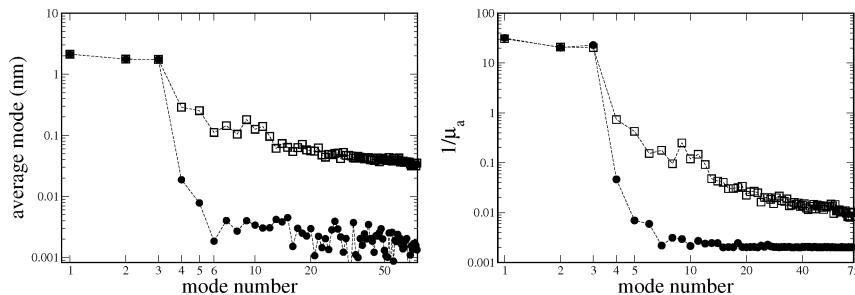
$$\text{NOE} = 1 + \frac{d^2}{4} \frac{\gamma_H}{\gamma_N} [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] T_1 \quad (15)$$

where  $c = [(\omega_N \delta_N)/\sqrt{3}]$  and  $d = [(\mu_0 h \gamma_H \gamma_N)/(8\pi^2 \langle r_{NH}^3 \rangle)]$ . Here  $\mu_0$  is the vacuum permeability,  $h$  is Planck's constant,  $\omega_H$  and  $\omega_N$  are the <sup>1</sup>H and <sup>15</sup>N Larmor frequencies of the experimental field,  $\gamma_H$  and  $\gamma_N$  are their respective gyromagnetic ratios,  $\delta_N$  is the chemical shift anisotropy of the <sup>15</sup>N nucleus, and  $r_{NH}$  is the N–H bond length.

### 3. MOLECULAR DYNAMICS SIMULATIONS OF UBIQUITIN

Molecular dynamic simulations have a double purpose in our studies as they provide the statistical parameters that define the conformational structure of the protein, which enters the Langevin equation, and second because they provide a test of the time correlation functions predicted with the proposed approach in the time regime covered by the simulations. In the short-time regime sampled by the simulations, the time correlation functions predicted by the theory, with input statistics from the simulations, have to be quantitatively consistent with the same functions directly sampled in the simulations.

Molecular dynamics simulations were performed for the protein ubiquitin in explicit solvent using the spc/e water model with the addition of the appropriate number of sodium and chloride ions to obtain a neutral system with 45 mM NaCl, identical to the NMR experiments of Lienin et al.<sup>22</sup> Simulations were performed in the canonical ensemble with the temperature set at 300 K. We utilized the AMBER99SB-ILDN<sup>23</sup> atomic force field for proteins starting from structures obtained from the RCSB protein databank (1UBQ<sup>24</sup> crystal structure of ubiquitin). The GROMACS<sup>25–28</sup> molecular dynamics engine was utilized running multiple nodes (12–64 cores) on the local ACISS cluster at the University of Oregon. The systems were solvated and energy minimized then tempered for 50 ps with all bonds constrained, utilizing a 1 fs time step. The system was then equilibrated for an additional 5 ns with a velocity rescaling



**Figure 3.** Left panel: The modulus of the average mode vector  $|\langle \vec{\xi}_a \rangle|$  (black circles) and the average of the mode modulus  $\langle |\vec{\xi}_a| \rangle$  (squares), calculated from the body-fixed simulation coordinates. The two quantities are nearly identical for the global modes, indicating the stability of the global structure, while  $|\langle \vec{\xi}_a \rangle|$  is almost zero for all internal modes. Right panel:  $\mu_a^{-1}$ , inverse eigenvalues of the bond correlation matrix  $\mathbf{U}$  calculated from the PDB crystal structure assuming the only flexibility to be uniform backbone bond length fluctuations of typical size ( $U_{ii} = 1.002$ ) in (black circles) and from the simulations (squares). The global modes are nearly identical while internal modes are drastically smaller in amplitude.

thermostat. After this equilibration, a 10 ns run was performed switching to a Nose-Hoover thermostat. Following this, ten configurations separated by 1 ns were randomly chosen, and these were used as initial conditions for ten production runs of 10 ns each. This was done to efficiently generate production trajectory; however, these ten simulations are clearly not independent since they were not equilibrated independently. In the calculations of the free-energy surfaces and the bond correlation matrix  $\mathbf{U}^{-1}$ , statistics from these ten production runs were used.

Ubiquitin was selected as a test system to analyze the proposed method because it is a protein which has been extensively studied both experimentally<sup>10–12</sup> and by computational methods.<sup>22,29</sup> While ubiquitin is a well-folded protein, it also contains domains of both rigid ( $\beta$ -sheets,  $\alpha$ -helices) and highly flexible (active loops, C-terminus) secondary structures and, as such, is an excellent test of the theoretical model.

#### 4. GLOBAL TUMBLING MODES AND INTERNAL FLUCTUATIONS

The dynamics of macromolecules develops in  $3N$ -dimensional configurational space, which for well-folded proteins largely reduces to fluctuations about an arbitrary three-dimensional folded structure. The LU matrix, which spans the configurational space of the protein, has three eigenvectors which span the  $\mathbb{R}^3$  subspace that contains the average folded structure; we will call these the global modes. For a well-folded globular protein, these global eigenvectors are just the ones with the three smallest eigenvalues,  $\lambda_a$  so they are the slowest modes to relax. However, this is not a strict criteria because sometimes internal modes can be as slow as the global ones. Furthermore, this rule will not hold for a highly anisotropic protein for which the values of the three eigenvalues can be very different in magnitude, or for very loosely ordered proteins, for which the scaling of the modes follows more closely that of unstructured synthetic polymers.

The global rotational modes can be unambiguously identified by the fact that they are the only modes with a nonzero time average in the body-fixed frame; that is  $\langle \vec{\xi}_a \rangle = \vec{0}$  for all internal modes, while for the global modes  $\langle \vec{\xi}_a \rangle$  is a vector which defines the principal diffusion axes of the protein, as shown in Figure 2. For clarity, Figure 3 shows the average mode length  $\langle |\vec{\xi}_a| \rangle$  and the length of the average mode vector  $|\langle \vec{\xi}_a \rangle|$  calculated directly from the simulation coordinates of the protein ubiquitin. The two quantities are practically identical for the global modes, indicating the stability of the global fold,

while the average internal mode vectors all tend to  $\vec{0}$  by 1 or 2 orders of magnitude. The fact that the average is different from zero indicates that in ubiquitin there is some, very small, amount of anisotropy in the internal fluctuations, mostly in the first few internal modes. Having identified them by this procedure, the protein's global orientational dynamics is spanned by these three modes of motion,<sup>2</sup> which describe rotations along the three main directions of the global structure.

The infinite-time limit of the time correlation function in the body-fixed reference frame gives

$$\lim_{t \rightarrow \infty} M_{1,i}(t) = \frac{\langle \vec{l}_i \rangle^2}{\langle \vec{l}_i^2 \rangle} \quad (16)$$

Expressing  $\vec{l}_i$  in its normal mode expansion leads to

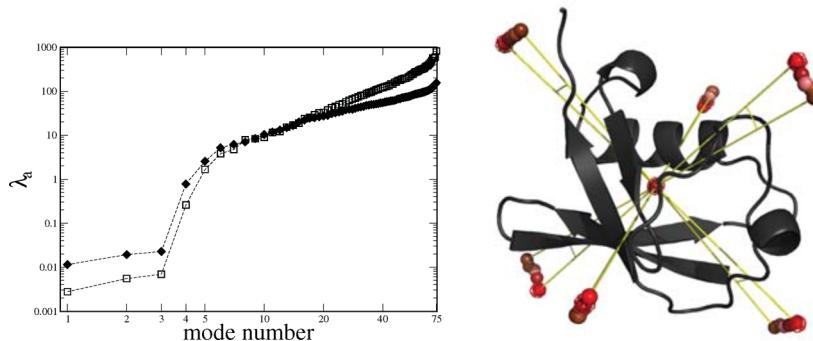
$$\frac{\langle \vec{l}_i \rangle^2}{\langle \vec{l}_i^2 \rangle} = \frac{\sum_{a,b=1}^{N-1} Q_{ia} \langle \vec{\xi}_a \rangle \cdot Q_{ib} \langle \vec{\xi}_b \rangle}{\langle \vec{l}_i^2 \rangle} \quad (17)$$

Due to the stationary nature of the global modes as a set of body-fixed axes as shown in Figure 3,  $\langle \vec{\xi}_{1,2,3} \rangle^2 \approx \langle (\vec{\xi}_{1,2,3})^2 \rangle$ , while due to the isotropic nature of the internal modes  $\langle \vec{\xi}_{a=4} \rangle = \vec{0}$ . With this property along with mode orthogonality, in the long time limit, eq 17 reduces to

$$\frac{\langle \vec{l}_i \rangle^2}{\langle \vec{l}_i^2 \rangle} \approx \sum_{a=1}^3 \frac{Q_{ia}^2}{\mu_a} = \sum_{a=1}^3 A_{ia} \quad (18)$$

indicating that the sum of the global mode amplitudes is a local bond-order parameter.

As another check that the LE4PD solution has separated into three global modes and  $N - 4$  internal modes, Figure 3 plots  $1/\mu_a$ , a dimensionless measure of the amplitude of mode fluctuations. The LE4PD is solved using as input the full bond correlation matrix  $\mathbf{U}$  calculated from the simulation, and that calculated from the PDB crystal structure assuming the only flexibility to be uniform backbone bond length fluctuations of typical size ( $U_{ii} = 1.002$ ). We have ordered the modes by the size of the eigenvalues of the LU matrix,  $\lambda_a$ . Modes  $a = 1, 2$ , and 3 are nearly identical in the two descriptions, reflecting only the small difference in structure between the solvated protein in the simulation and that in the crystal (Figure 5). Modes  $a = 4, N - 1$  are drastically smaller in amplitude in the LE4PD with only bond length fluctuations, indicating that these are the internal modes and the small perturbation of the bond fluctuations in the model we constructed are not nearly as collective as the larger-amplitude internal motion sampled in the simulation.



**Figure 4.** Left panel: Eigenvalues of the LU matrix, which are proportional to the rates of the diffusive processes governing the dynamics of the Ubiquitin protein, without hydrodynamic interaction ( $\square$ ) and with hydrodynamic interaction ( $\blacklozenge$ ). Right panel: The red mesh and the red ball correspond to the orientation of the principal axis of the average structure, and the orientation of the first three modes in eq 6 solution with  $H_{\text{int}} = 0$ . The first three eigenvectors of the U matrix alone correspond to the principal inertial axes of the protein. For simplicity, this model calculation is performed with identical friction coefficients for each amino acid. With the use of this simplification, the hydrodynamic interaction is ramped up by tuning the hydrodynamic radius  $r^w$  from 0 (i.e., no hydrodynamic interaction) to its maximum possible value yielding positive relaxation times. The rotational axes are modified by the presence of the hydrodynamic interaction.

These three global modes define the body-fixed coordinate system. To illustrate the relation of the global modes to the structure of the molecule, we calculate the average gyration tensor of the molecule,

$$\langle S_{\alpha\beta} \rangle = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=i+1}^N \langle (\vec{R}_j - \vec{R}_i) \times \hat{\alpha} (\vec{R}_j - \vec{R}_i) \times \hat{\beta} \rangle \quad (19)$$

where  $\alpha, \beta$  run over the Cartesian indices  $x, y$ , and  $z$ . The eigenvalues of the gyration tensor  $\langle R_g^2 \rangle$  are typically used to denote the shape of a macromolecule.<sup>30</sup> Noting that the bead separation vector can be written as  $(\vec{R}_j - \vec{R}_i) = \sum_{k=i}^{j-1} \vec{l}_k$ , and using the notation  $\vec{l}_k \cdot \hat{\alpha} = l_k^\alpha$  to denote the  $\alpha$ th Cartesian component of the  $k$ th bond vector,

$$\langle S_{\alpha\beta} \rangle = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=i+1}^N \sum_{k,l=i}^{j-1} \langle l_k^\alpha l_l^\beta \rangle \quad (20)$$

and expanding into normal modes

$$\langle S_{\alpha\beta} \rangle = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=i+1}^N \sum_{k,l=i}^{j-1} \sum_{a,b=1}^{N-1} Q_{ka} Q_{lb} \langle \xi_a^\alpha \xi_b^\beta \rangle \quad (21)$$

where we can now exploit the orthogonality of the mode description.

The internal modes  $a = 4, N - 1$  in the harmonic theory are isotropic, reducing to  $\langle \xi_a^\alpha \xi_a^\beta \rangle = (1/3)(l^2/\mu_a)\delta_{\alpha\beta}$  independent of the orientation of the  $x, y$ , and  $z$  coordinate system. However, the three global modes pick out three orthogonal directions in 3D space and define a natural body-fixed coordinate system where by convention we align the  $z$  axis with the long axis of the molecule corresponding to  $\xi_1$  and the  $x$  axis with the shortest axis corresponding to  $\xi_3$ , and the  $y$  axis with the middle axis corresponding to  $\xi_2$ . With this global mode coordinate system,  $\langle \xi_a^\alpha \xi_a^\beta \rangle = (l^2/\mu_a)\delta_{\alpha\beta}\delta_{aa}$ , and the result

$$\begin{aligned} \langle R_g^2 \rangle &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=i+1}^N \sum_{k,l=i}^{j-1} \left( Q_{ka} Q_{la} \frac{l^2}{\mu_a} \right. \\ &\quad \left. + \frac{1}{3} \sum_{a=4}^{N-1} Q_{ka} Q_{la} \frac{l^2}{\mu_a} \right) \end{aligned} \quad (22)$$

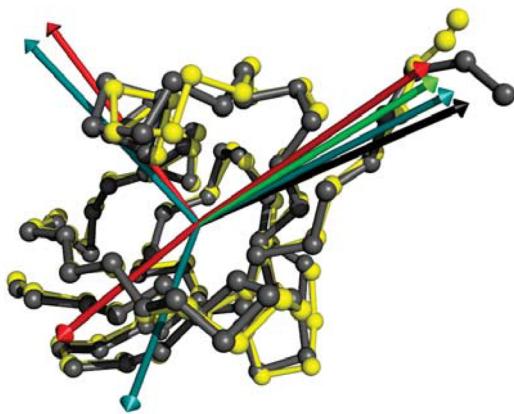
where  $\alpha$  corresponds to Cartesian directions  $x, y$ , and  $z$  or modes  $a = 3, 2, 1$ . The distinction between the three global modes and the  $N - 4$  internal modes is clear; the global modes provide the contribution from the shape of the global folded structure, while the internal modes provide the contributions from the fluctuations. In a random walk model like the freely jointed chain where there is no global structure, the gyration radius contains only contributions from internal degrees of freedom; this expression reduces to the expected  $\langle R_g^2 \rangle = Nl^2/6$ . In a rigid body model, the internal modes would have zero amplitude and the only contributions would be from the global modes. This calculation of the gyration radius of the protein using the LE4PD modes yields a root-mean-square length of  $R_g = 5.509 \text{ \AA}$ ,  $R_{g_x} = 6.126 \text{ \AA}$ ,  $R_{g_y} = 8.575 \text{ \AA}$ , and <1% error from that directly calculated from the simulation trajectory.

To analyze the effect of the hydrodynamic interaction on the dynamics, we report in the right panel of Figure 4 the protein with its global mode axes from the global tumbling modes without hydrodynamic contribution, which corresponded to the inertial axes of the protein. The figure shows that as the hydrodynamic interaction becomes more and more relevant, the principal diffusion axes shift, as expected.<sup>31</sup> In addition, the right panel of Figure 4 shows how the eigenvalues of LU change quite dramatically upon inclusion of the hydrodynamic interaction; global and low-order internal modes are faster, while higher-order internal modes are slowed with softer scaling with mode number. This is similar to the effect of including the hydrodynamic interaction in flexible polymer descriptions (the Rouse–Zimm theory<sup>17</sup>), where the scaling of relaxation rate with mode number goes from  $k_a \sim a^2$  to  $k_a \sim a^{3/2}$ . The hydrodynamic interaction component strongly affects both the global mode relaxation rates, and the orientation of the rotational diffusion axes and has a relevant contribution to the protein dynamics in general and cannot be discarded.

Equation 6 deals with the flexibility about the folded structure in an ensemble averaged fashion. For well-folded proteins, the LE4PD, with simulations as an input, is a good approximation to determine the diffusion tensor, while including the contribution to the dynamics due to the fact that proteins are semiflexible objects. Most computational approaches to determining the rates of rotational diffusion and the diffusion tensor orientation are designed around the idea of

the protein as a rigid body. While in these models the hydrodynamical treatment is often more sophisticated than in our model,<sup>32</sup> this sophistication comes at the cost of neglecting the contributions from internal degrees of freedom and conformational flexibility to protein dynamics.

Figure 5 shows that the average orientation of the principal diffusion axes calculated from the simulation trajectory of ubiquitin with the LE4PD theory is very close to that calculated using other methods; the orientation of the most relevant long axis is 10.9 degrees from that found by fitting the NMR relaxation data to an axially symmetric anisotropic model, using the N–H bond orientations from the crystal structure, 6.5 degrees from that found using the HYDRO<sup>33</sup> program with the crystal structure, and 2.8 degrees from the orientation found using the HYDRO program and a set of snapshots taken from a Langevin simulation with the tail unhindered by crystal contacts.<sup>10</sup> Figure 5 shows that the difference in the orientation



**Figure 5.** A comparison of the Ubiquitin crystal structure (gray), the ensemble averaged structure from the simulations (yellow), the orientation of the long axis of the anisotropic diffusion tensor fit to NMR experiment (black),<sup>10</sup> HYDRO calculation from the crystal structure (blue), and the average global tumbling mode orientation from the theory (red). The long-axis orientation is very close to other hydrodynamical modeling treatments when tail flexibility is taken into account, although the short axes orientation is not.

of the diffusion axes between the LE4PD and other methods appears to be related to the different possible configurations of the tail. Given that the tail is quite mobile, the observed difference in the diffusion axes orientation could be just related to a difference in the sampling of the configurations. This suggests that eq 6 has captured the orientation of the principal diffusion axes of the protein.

We also compare the rates of rotational diffusion calculated using the LE4PD constructed from the crystal structure alone, to that calculated using the HYDRONMR program.<sup>34</sup> The overall rotational diffusion rate in the LE4PD  $D_{av} = (1/3)TrD$  is identical to that calculated using the HYDRONMR program when the AER, the adjustable atomic element radius, is set to  $a = 1.6 \text{ \AA}$ . Fit to NMR experiment was found using the HYDRONMR program at  $a = 2.2 \text{ \AA}$ ,<sup>35</sup> typical values of the AER are much larger,  $\sim 3.1 \text{ \AA}$ . Overall anisotropy  $2D_z/(D_x + D_y)$  in the HYDRONMR calculation and the LE4PD was comparable at 1.45 and 1.48, respectively. These results suggest that the simplified hydrodynamical treatment of the LE4PD with the inclusion of internal friction sources can lead to similar

results for rotational diffusion as other approaches that have a more sophisticated description of hydrodynamics and exclusively consider external solvent friction sources and overlook structural fluctuations.

The differences in the rotational diffusion tensor between that found when fitting to NMR relaxation data, computational rigid-body hydrodynamical modeling, and the LE4PD, can be attributed to three main sources: namely, (i) the hydrodynamical treatment and its effect on the large-scale dynamics, (ii) the presence of flexibility (i.e., differences between the solvated structural ensemble in solution and the crystal structure), and (iii) the sources of dissipation considered, both solvent and internal friction in our model. The qualitative agreement between different methods observed for ubiquitin implies that these effects in this protein are not large. Though ubiquitin has a highly flexible C-terminal tail, there is no large-scale domain reorientation, so that the static structure of the protein is close enough to the equilibrated structure in solution and is a good representation of the equilibrium structural ensemble. In general, however, for other proteins this is not true, and we argue that starting from a static picture of a protein to calculate the diffusion tensor can have some degree of limitation in providing an accurate representation of the rotational dynamics for the reasons just discussed.

**4.1. Rotational Dynamics in an Inertial Frame.** The mode solution of the LE4PD eq 7 correctly describes the dynamics of the molecule in a body-fixed coordinate system attached to the molecule. However, when calculating time correlation functions in an inertial (lab) frame of reference, care must be taken to deal with the inherent anisotropy of the global tumbling modes. The internal modes, approximated here to be completely isotropic, are not affected by the transformation to an inertial coordinate system, so the first step is to isolate the global tumbling modes in eq 7. The global modes are an orthogonal set of vectors defining a set of body-fixed coordinates, and aligning the body-fixed coordinate system,

$$\begin{pmatrix} \hat{x}' \\ \hat{y}' \\ \hat{z}' \end{pmatrix} = \begin{pmatrix} \hat{\xi}'_3 \\ \hat{\xi}'_2 \\ \hat{\xi}'_1 \end{pmatrix} \quad (23)$$

we obtain an orientational diffusion equation in the body-fixed system

$$\frac{\partial}{\partial t} \begin{pmatrix} \vec{\xi}'_3(t) \\ \vec{\xi}'_2(t) \\ \vec{\xi}'_1(t) \end{pmatrix} = -\sigma \begin{pmatrix} \lambda_3 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix} \begin{pmatrix} \vec{\xi}'_3(t) \\ \vec{\xi}'_2(t) \\ \vec{\xi}'_1(t) \end{pmatrix} + \begin{pmatrix} \vec{v}'_3(t) \\ \vec{v}'_2(t) \\ \vec{v}'_1(t) \end{pmatrix}$$

The global modes define the orientation of the protein which in the lab-fixed frame of reference is rotating; denoting vectors in this inertial frame as unprimed  $\vec{\xi}_3, \vec{\xi}_2, \vec{\xi}_1$ , at  $t = 0$ , we align the  $x, y$ , and  $z$  lab system with the  $\vec{\xi}'_3, \vec{\xi}'_2, \vec{\xi}'_1$  body axis of the protein. For  $t \neq 0$ , the body-fixed and lab axes are related by

$$\begin{pmatrix} \vec{\xi}'_3(t) \\ \vec{\xi}'_2(t) \\ \vec{\xi}'_1(t) \end{pmatrix} = \hat{\mathcal{R}}(t) \begin{pmatrix} \vec{\xi}_3(t) \\ \vec{\xi}_2(t) \\ \vec{\xi}_1(t) \end{pmatrix} \quad (24)$$

where  $\hat{\mathfrak{R}}(t)$  is the time-dependent rotation which takes the protein into its actual orientation, which is in general not aligned with the  $xyz$  lab system.

In the inertial frame, the LE4PD is now

$$\frac{\partial}{\partial t} \hat{\mathfrak{R}}(t) \begin{pmatrix} \vec{\xi}_3(t) \\ \vec{\xi}_2(t) \\ \vec{\xi}_1(t) \end{pmatrix} = -\sigma \begin{pmatrix} \lambda_3 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix} \hat{\mathfrak{R}}(t) \begin{pmatrix} \vec{\xi}_3(t) \\ \vec{\xi}_2(t) \\ \vec{\xi}_1(t) \end{pmatrix} + \hat{\mathfrak{R}}(t) \begin{pmatrix} \vec{v}_3(t) \\ \vec{v}_2(t) \\ \vec{v}_1(t) \end{pmatrix}$$

and left multiplying by the inverse rotation which takes the lab coordinate system into the body-fixed coordinates,

$$\frac{\partial}{\partial t} \begin{pmatrix} \vec{\xi}_3(t) \\ \vec{\xi}_2(t) \\ \vec{\xi}_1(t) \end{pmatrix} = -\sigma \hat{\mathfrak{R}}^{-1}(t) \begin{pmatrix} \lambda_3 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix} \hat{\mathfrak{R}}(t) \begin{pmatrix} \vec{\xi}_3(t) \\ \vec{\xi}_2(t) \\ \vec{\xi}_1(t) \end{pmatrix} + \begin{pmatrix} \vec{v}_3(t) \\ \vec{v}_2(t) \\ \vec{v}_1(t) \end{pmatrix}$$

where  $\hat{\mathfrak{R}}(0) = \hat{1}$  and identifying the operator

$$\hat{D}(0) = \begin{pmatrix} \lambda_3 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix} \quad (25)$$

and

$$\hat{D}(t) = \hat{\mathfrak{R}}^{-1}(t) \begin{pmatrix} \lambda_3 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix} \hat{\mathfrak{R}}(t) \quad (26)$$

The time-dependent  $\hat{D}(t)$  can be redefined by assuming that for short time intervals the rotation of the protein takes place in a series of small angular displacements. We note here that  $(\sigma/2) \hat{D}(0)$  is the diagonalized rotational diffusion tensor. This procedure yields a time-independent operator  $\hat{D}_{\mathfrak{R}}$  valid for rotational diffusion. While rigid-body rotational diffusion is well-known,<sup>36</sup> we present a formal derivation for the rotational diffusion of a semiflexible, fluctuating macromolecule.

The rotation can be expressed as three consecutive rotations around the lab-fixed axes  $\hat{\mathfrak{R}}(t) = \hat{\mathfrak{R}}_z[\gamma(t)]\hat{\mathfrak{R}}_y[\beta(t)]\hat{\mathfrak{R}}_x[\alpha(t)]$ . Considering small timesteps ( $\Delta t$ ), we assume the angles  $\alpha(t)$ ,  $\beta(t)$ , and  $\gamma(t)$  to be proportional to  $\Delta t$ , which gives to lowest order in  $\Delta t$ ,

$$\hat{D}(\Delta t) = \hat{\mathfrak{R}}_x^{-1}(\Delta t)\hat{H}(0)\hat{\mathfrak{R}}_x(\Delta t) + \hat{\mathfrak{R}}_y^{-1}(\Delta t)\hat{H}(0)\hat{\mathfrak{R}}_y(\Delta t) + \hat{\mathfrak{R}}_z^{-1}(\Delta t)\hat{H}(0)\hat{\mathfrak{R}}_z(\Delta t) \quad (27)$$

The rotations, to linear order in  $\Delta t$ , are  $\hat{\mathfrak{R}}_{x,y,z}(\Delta t) = \hat{1} + (\Delta t) \hat{L}_{x,y,z}$  where

$$\hat{L}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \hat{L}_y = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \hat{L}_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (28)$$

are angular momentum operators. Since we are dealing with a diffusive process, we can take the limit of infinitely small time steps

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} [\hat{\mathfrak{R}}_{x,y,z}(\Delta t) - \hat{\mathfrak{R}}_{x,y,z}(0)] = \hat{L}_{x,y,z} \quad (29)$$

In this limit, we obtain for the time-independent operator

$$\hat{D}_{\mathfrak{R}} = \frac{1}{2} [\hat{L}_x^T \hat{D}(0) \hat{L}_x + \hat{L}_y^T \hat{D}(0) \hat{L}_y + \hat{L}_z^T \hat{D}(0) \hat{L}_z] \quad (30)$$

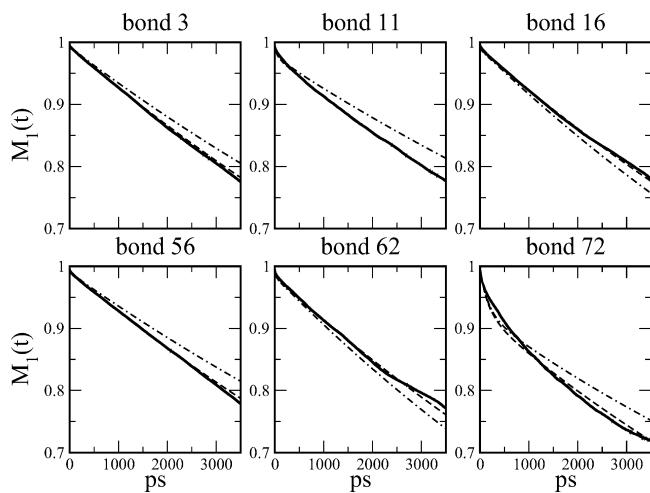
where the factor of  $(1/2)$  comes from requiring that  $\text{Tr}[\hat{D}(0)] = \text{Tr}(\hat{D}_{\mathfrak{R}})$ . Performing the matrix multiplication, the final LE4PD for the global modes in the lab-fixed inertial coordinate system is

$$\frac{\partial}{\partial t} \begin{pmatrix} \vec{\xi}_3(t) \\ \vec{\xi}_2(t) \\ \vec{\xi}_1(t) \end{pmatrix} = -\sigma \begin{pmatrix} \frac{1}{2}(\lambda_3 + \lambda_2) & 0 & 0 \\ 0 & \frac{1}{2}(\lambda_3 + \lambda_1) & 0 \\ 0 & 0 & \frac{1}{2}(\lambda_2 + \lambda_1) \end{pmatrix} \times \begin{pmatrix} \vec{\xi}_3(t) \\ \vec{\xi}_2(t) \\ \vec{\xi}_1(t) \end{pmatrix} + \begin{pmatrix} \vec{v}_3(t) \\ \vec{v}_2(t) \\ \vec{v}_1(t) \end{pmatrix} \quad (31)$$

In the LE4PD, the  $a = 4, N - 1$  internal modes are treated to be isotropic and are unaffected by this change from a body-fixed to an inertial frame of reference, which is translating with the center of mass of the protein. This is an approximation, and for a more flexible protein with large-amplitude highly anisotropic fluctuations, the internal modes may need to be treated in a similar fashion.

The Rouse-Zimm model, eq 6, is typically applied to systems that are either statistically spherically symmetric (a flexible freely jointed chain, or a semiflexible freely rotating chain) or possess a single long-axis (a rod). In these cases, there is only one global tumbling mode.  $D_{\text{LE}}$  is just a number, not a tensor, and the rotational diffusion equation leaves the single rotational decorrelation rate unaltered. However, for proteins which have an arbitrary anisotropic folded structure, the full rotational diffusion equation of an arbitrary three-dimensional body must be solved.

**4.2. Rotational Diffusion and Solvent Viscosity in Simulations.** Extracting rotational dynamics directly from a simulation is problematic due to the computational cost of long simulation runs exceeding the relaxation time of the global rotation by one or 2 orders of magnitude. Furthermore, current water models have been seen to lead to inaccurate viscosity and surface hydration.<sup>37</sup> The viscosity of the spc/e water model has been estimated to be the same as real water<sup>38</sup> or about 27% lower.<sup>39</sup> After utilizing the proposed LE4PD formalism, the rates of rotational decorrelation are found to be consistent with both the simulation and experiment. Figure 6 displays a comparison between  $M_1(t)$  calculated directly from the simulation and from theory where internal barrier crossings



**Figure 6.**  $M_1(t)$  calculated from the simulation (black), LE4PD with  $l = 1$  rotational diffusion eigenvalues (dashed line), and the bare LE4PD solution in the body-fixed frame (dashed-dotted line). The LE4PD approach includes the calculation of the internal mode energy barriers as described in Dynamical Renormalization of the Internal Modes from Free Energy Surfaces

are included as discussed in Dynamical Renormalization of the Internal Modes from Free Energy Surfaces. The figure shows data for six selected bonds where the rotational diffusion eigenvalues so calculated improve the comparison to the simulation, which is true for 71% of the 75 bonds in ubiquitin.

Quantitative results are obtained for the rotational dynamics in the simulation when the viscosity in the LE4PD is set to 1.33 times that of real water. For internal processes, the relaxation is less sensitive to the exact viscosity used in the LE4PD because the internal timescales are dominated by the eigenvalue spectra of the LU matrix and by the energy barriers to relaxation. However, at longer times the rotational decorrelation dominates the dynamics, making the long-time slope of the bond autocorrelation highly sensitive to the viscosity used as input to the theory. The statistical error in the longer-time regime of the time-correlation functions calculated from the 10 ns simulations makes the analysis of the reasons for discrepancy between spc/e water viscosity estimates and the viscosity used in the LE4PD not particularly conclusive. In particular, Wong and Case estimated the statistical error in an exponential rotational relaxation process at  $t \sim \tau$  with  $\tau$  a rotational correlation time to be about 23% when calculated from a 200 ns MD trajectory.<sup>37</sup> The relationship between the diffusive dynamics of proteins, solvent viscosity, internal viscosity, and hydration layer properties of the in silico protein in water model certainly warrants further study and comparison with much longer simulation runs where the simulation time exceeds the timescale of global diffusive processes by 1–2 orders of magnitude. Studies of diffusive motions of proteins in these time regimes and the tuning of water models for better viscosity have recently been undertaken,<sup>40</sup> but the intent of this work is to demonstrate the utility of the LE4PD to obtain the protein dynamics from relatively short simulation runs. For comparison to experiment, the correct viscosity of 10% D<sub>2</sub>O and 90% H<sub>2</sub>O at the experimental temperature is used, while for comparison to simulation results, the viscosity of 1.33 times that of real water at 300 K is used in the LE4PD.

## 5. DYNAMICAL RENORMALIZATION OF THE INTERNAL MODES FROM FREE-ENERGY SURFACES

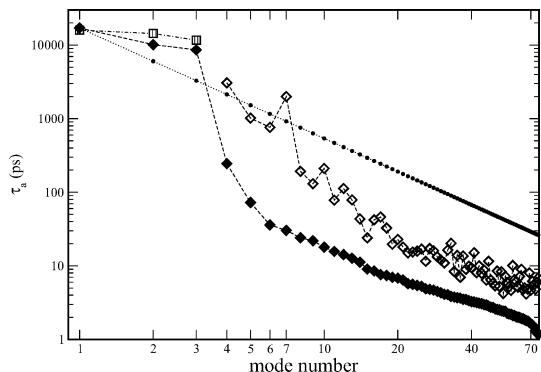
In this section, we describe how we account for local-mode barrier crossing in the Langevin equation for protein dynamics. In our previous paper, this correction was not included in the theory,<sup>3</sup> but we noticed how the disagreement between theoretical predictions and the simulations, from which the theoretical parameters were calculated, for the function  $M_{1,i}(t)$  was largely a consequence of neglecting the local energy barriers.

Here, we explicitly account for the complexity of the free-energy landscape, working with the internal mode description of the linearized Langevin equation and correcting the modes  $a$  *a posteriori* by introducing the local energy barriers, as measured in simulations. We are interested in calculating the bond autocorrelation function, where we separate the first three nontranslational modes from the internal ones as

$$M_{1,i}(t) = \sum_{a=1}^3 A_{ia} \exp[-\sigma\lambda_a t] + \sum_{a=4}^{N-1} A_{ia} \exp[-\sigma\lambda_a t] \quad (32)$$

under the assumption that the first three modes represent the rotational dynamics of the molecule, while the higher modes represent the internal dynamics. This separation of local and global motion in  $M_1(t)$  is due entirely to the mode structure of the solution of eq 6 and does not rely in any way upon the separation of the dynamics into slow and fast processes. In fact, we will show that upon including internal energy barriers in the mode coordinates some internal modes can become slow enough to occur on the same timescale than the rotational decorrelation modes. It is important to note that the lack of time-scale separation between the first three modes and the other ones does not affect the validity of the treatment that is presented here. Furthermore, this separation of local and global motion occurs in the context of a solution for a first-order correlation, while in higher-order correlation functions, such as  $P_{2,i}(t)$ , global and local modes are necessarily mixed.

While the first nonzero three modes of relaxation in a largely stiff object mainly describe the rotational dynamics of the macromolecule, the modes with higher index,  $a = 4, \dots, N - 1$ , describes the internal dynamics of the protein and the collective breathing modes of an overdamped, harmonically connected object. The more cooperative the motion, the slower the characteristic timescale of the normal mode. When applied to linear synthetic polymers in solvent, which are uniform in their composition and are rotationally symmetric in their structure, the time of relaxation scales in the Rouse–Zimm theory with mode number  $a$  as  $\tau_a \sim a^{-3/2}$ , with  $\tau_a = (\sigma\lambda_a)^{-1}$ .<sup>17</sup> Representing the same function for a protein, for example ubiquitin, shows a different and more interesting behavior. Figure 7 compares the Rouse–Zimm scaling law with the data from the simulations of ubiquitin before and after accounting for the local energy barriers in the internal normal modes. Without energy barriers, the global modes are clearly slower than the internal ones. The inclusion of the local energy barriers in the local modes leads to a more complex mode-dependence of the relaxation times. The first handful of internal modes display relaxation times comparable to the rotational modes: those are the modes which span the most collective and large-amplitude fluctuations of the molecule. In general, the relaxation time still decreases with mode number and, interestingly, the slowest more-cooperative internal modes also have the highest free-energy barriers. This result is conceptually consistent with the behavior



**Figure 7.** Mode-dependent relaxation time for the protein ubiquitin. In the bare LE solution (◆), the rotational modes are well-separated from the internal modes, while once the dynamics are modified to account for the local free-energy barriers, the internal modes (◇) become closer to the global modes. The three global modes after accounting for anisotropic rotational diffusion become more uniform in timescale (striped squares). Also reported is the standard Rouse-Zimm model scaling  $a^{-3/2}$  of the normal mode dynamics (●). Even for ubiquitin, a well-folded protein, there is no separation in the timescale between global and internal modes after dynamical renormalization.

emerging in the first few diffusion coordinates in the LSDmap method,<sup>41</sup> where eigenfunctions of the Fokker–Planck operator are estimated.

Each normal mode obtained from the diagonalization of eq 6 is a vector defined by the linear combination of the bond vectors weighted by the eigenvectors of the product of matrices LU, as  $\vec{\xi}_a(t) = \sum_i Q_{ai}^{-1} \vec{l}_i(t)$ . In polar coordinates, the vector is represented as  $\vec{\xi}_a(t) = \{|\vec{\xi}_a(t)|, \theta_a(t), \phi_a(t)\}$ . In eq 6, the orientation of each Langevin internal mode diffuses isotropically about all solid angles; however, the probability distribution of the internal mode vector orientation per solid angle calculated from the simulation trajectory is unique for each mode.

As a first approximation, we assume that the fluctuations in the modulus of the normal mode vector are random and independent of mode orientation, so that the relevant changes in the normal mode free energy occur as the angles, expressed in the spherical coordinates, span the configurational space. For a generic normal mode,  $a$ , the free-energy surface is defined as a function of the spherical coordinate angles  $\theta_a$  and  $\phi_a$  as

$$F(\theta_a, \phi_a) = -k_B T \log\{P(\theta_a, \phi_a)\} \quad (33)$$

with  $P(\theta_a, \phi_a)$  the probability of finding the normal mode vector having the given value of the solid angle.

The internal normal modes provide information about the complexity of the configurational free-energy landscape around the folded state. By calculating the mode-dependent free-energy surface from the simulation probability distribution, we observe that even close to the global folded minima the free-energy landscape is rough and contains multiple local minima, metastable states, and local barriers with related multiple possible pathways for the local dynamics (see, for example, Figure 8).

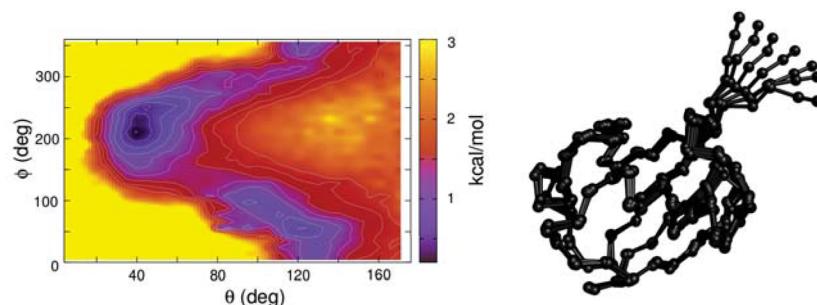
For an approximate rescaling of the dynamics, we correct the rate of reorientational diffusion of each mode by assuming thermal activation over the mode-dependent energy barrier  $\langle E^\dagger \rangle$  where

$$\langle k_a \rangle = k_a^0 \exp[-\langle E_a^\dagger \rangle / (k_B T)] \quad (34)$$

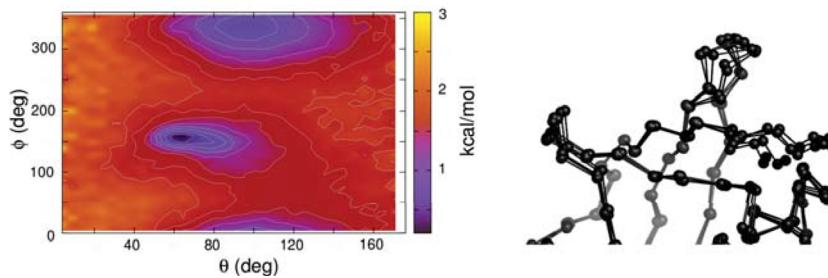
This simple dynamical renormalization provides an average correction to the dynamics of the LE, which approximately accounts for the local barrier crossing and is in agreement with free-energy landscape theories, suggesting an underlying dynamical glass transition at low enough temperature.<sup>42,43</sup> Other more detailed methods can be used to model barrier crossing, such as Markov network models<sup>44</sup> and approaches which explicitly deal with the complexity of the free-energy landscape.<sup>41,45–47</sup> As a first approximation, the depth of the minimum free-energy well in the mode serves as the relevant barrier to transport, and by requiring thermal activation over this barrier, the timescale of all mode-dependent dynamical quantities are simply renormalized by these factors. This simple model gives a realistic first-order correction to the barrier-free Langevin dynamics.

The presence of internal free-energy barriers is not limited to the first internal modes. Figure 11 shows that the complexity in the free-energy surface is present in any normal mode, extending out to fluctuations on mode 17th, which is an intermediate mode, to mode 75th, which is the most local and the last of the internal modes. In this way, the renormalization of the dynamics has to be included for each internal normal mode of motion.

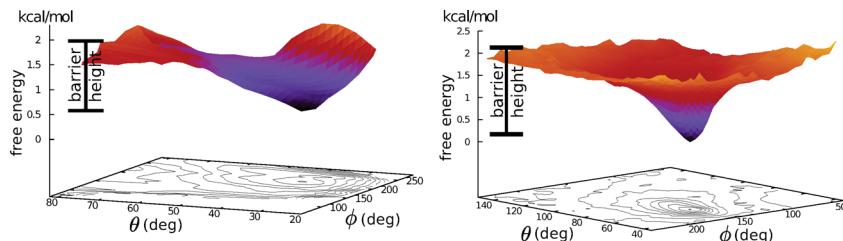
**5.1. Protein Real-Space Configurations from the Normal Mode Pathways.** In simulations, each time frame corresponds to a specific molecular configuration. Each molecular configuration corresponding to a given simulation frame at time  $t$  can be described by the set of bond vectors  $\vec{l}_1(t), \vec{l}_2(t), \dots, \vec{l}_{N-1}(t)$  or equivalently by the set of mode vectors  $\vec{\xi}_1(t), \vec{\xi}_2(t), \dots, \vec{\xi}_{N-1}(t)$ . However, a single mode vector orientation is associated with an ensemble of molecular configurations. Given that we are interested in the explicit representation of the structure along a pathway of interest for each normal mode, all



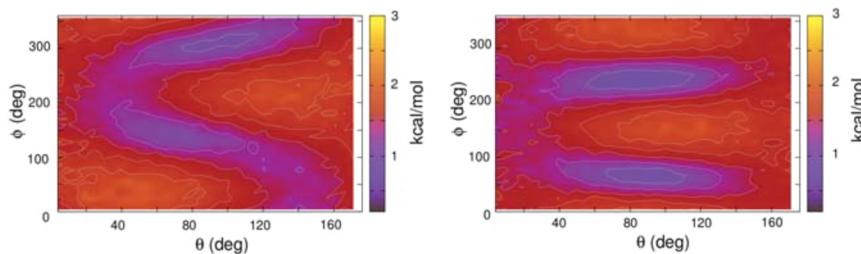
**Figure 8.** Left panel: free-energy surface of the first internal mode. Right panel: structural fluctuations in the first internal mode along the path of minimum energy.



**Figure 9.** Left panel: free-energy surface of the fourth internal mode. Right panel: structural fluctuations in the fourth internal mode, involving the loop containing Lys11, a relatively abundant linkage site involved in cell-cycle regulation<sup>49</sup>



**Figure 10.** Depth of the free-energy well is used as  $E^\dagger$ , the free-energy barrier to mode fluctuations. Left panel: first internal mode of ubiquitin. Right panel: fourth internal mode of ubiquitin.



**Figure 11.** Free-energy surface of the 17th internal mode (left panel) and of the 75th internal mode (right panel). Both at intermediate and local length scales, there is still complexity in the internal mode FES.

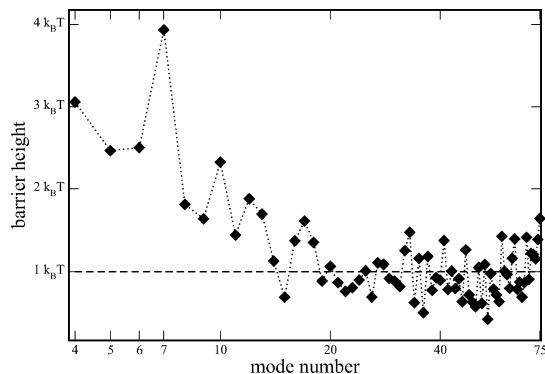
structures from the simulation ensemble which pertain to a particular  $\theta, \phi$  orientation of a mode vector of interest are extracted and averaged. By calculating this average structure along a particular pathway in the free-energy surface of a particular mode, for example along the path of lowest free energy connecting two minima, the structural pathway of relaxation in that mode can be obtained. This is a very different process than projecting the protein structures along particular directions, often used for example to visualize eigenvector directions in essential dynamics analysis.<sup>48</sup> Figure 8 displays the free-energy surface and structural pathway of the first of the internal modes of ubiquitin. Because the modes are in general ordered from the most collective to the most local, this is the most collective internal motion for ubiquitin. Using this process of transforming the protein coordinates from the simulation into the mode representation, and then extracting back the average structure for each specific mode orientation, is possible to reconstruct the ensemble of protein configurations along a structural pathway for each mode.

The transform shows that the first internal mode mainly involves large-amplitude fluctuations of the C-terminal tail, which are biologically important as the C-tail is the linkage point for polyubiquitination.<sup>49</sup> Figure 9 shows the free-energy surface of the fourth internal mode which captures concerted fluctuations of the tail and the loop containing Lys11, a relatively abundant linkage site involved in cell-cycle

regulation.<sup>49</sup> The most collective internal modes are those which span the important functional motion of the protein, illustrating the fundamental relationship between protein dynamics and biological function.<sup>50</sup>

**5.2. Coarse-Gained Representation, Dynamics, and Internal Energy Rescaling.** In the independent normal modes representation, the LE4PD theory renormalizes the rough free-energy surface measured in simulations and replaces it with a smooth isotropic surface upon which dynamical processes become faster than in the real system.<sup>51</sup> From the solution of eq 6, reorientational diffusion takes place at a rate  $k_a^0 = \sigma\lambda_\omega$  while the real free-energy surface of the normal mode, as it emerges from the molecular dynamic simulations, presents complex roughness. Dynamical processes need to include activation dynamics to overcome internal energy barriers.<sup>52</sup> To find this average activation barrier  $\langle E_a^\dagger \rangle$ , we calculate the depth of the well in which the minimum free-energy state lies, for each normal mode. Figure 10 shows, as an example, the average free-energy barriers in the first and fourth internal mode of ubiquitin. The difference in energy is taken between the deepest value of the energy in the energy well and the barrier that the system needs to overcome to escape the same.

The height of the energy barrier is mode-dependent (see Figure 12). Complex energy landscape is observed for any mode and at all length scales, but large energy barriers, present in the first internal modes, converge to smaller barriers for



**Figure 12.** Height of the local-mode free-energy barrier to conformational diffusion as a function of mode number. The first handful of internal modes display large energy barriers, while the average energy barrier for the higher order internal modes goes to  $\sim k_B T$ .

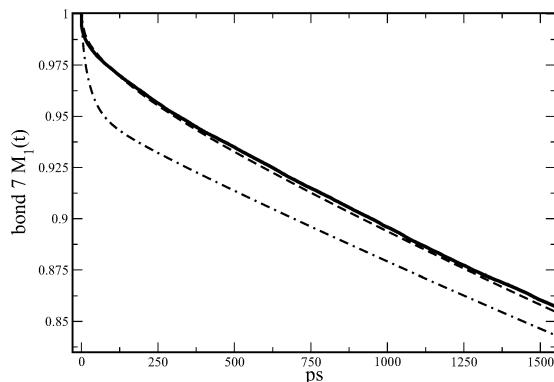
more local modes, and finally to a minimum value given by  $\langle E_{\text{int}}^{\dagger} \rangle \sim k_B T$ . Assuming that the slowing down of the dynamics in the hydrophobic region is largely dominated by local barrier-crossing, we evaluated the protein internal viscosity as

$$\eta_p = \eta_s e^{\langle E_{\text{int}}^{\dagger} \rangle / k_B T} \quad (35)$$

where the external solvent viscosity is taken as the reference point in the calculation.<sup>18</sup> The solvent accessible surface area calculation shows that all protein residues are at least partially exposed to the solvent, which justifies the use of the solvent viscosity as the valid reference point, as all protein sites are at least partially solvated and all dissipation must ultimately lead to dissipation into the solvent to maintain equilibrium. Because the protein's free-energy landscape is in general rough and funnel-shaped, containing barriers of all heights,<sup>42</sup> it seems typical that at any finite temperature, the most highly sampled and relevant local internal barriers should be on the order of  $\sim k_B T$ . This simple estimate of the protein internal viscosity in the hydrophobic core is used as an input in eq 5. At this point, all the information needed to solve the equation of motion for protein dynamics is defined.

**5.3. Testing the Free-Energy Surface Approximation against Simulations.** As a first test of the proposed approach, we compare the decay of the autocorrelation function of a  $C_{\alpha}$ – $C_{\alpha}$  bond,  $M_{1,i}(t)$  defined in eq 32, with the data from simulations. The theoretical function is calculated from the solution of the Langevin equation for protein dynamics, with and without taking into account the barrier crossing of the internal normal modes. Figure 13 reports, as an example, the bond autocorrelation for bond seven: bond seven lies in a particularly active loop with large barriers to fluctuations; the inclusion of the renormalization of the local modes dynamics, due to the internal energy barrier, drastically improves the agreement with simulations. Further comparison of the LE4PD theory predictions for  $M_i(t)$ , with the renormalization from the energy barriers in the internal modes, can be seen in Figures 6 and 16.

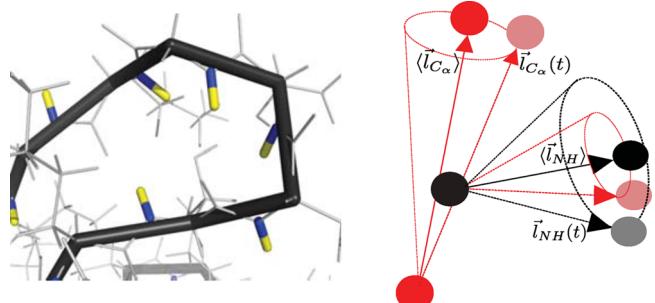
Figure 15 displays similar agreement between theory and simulations for bonds belonging to  $\alpha$ -helices,  $\beta$ -sheets, other loops, and the flexible C-terminal tail, where rotational contributions are removed from both the simulation and the theoretical time correlation functions to highlight the internal dynamics.



**Figure 13.** Bond autocorrelation function  $M_1(t)$ , with both rotational and internal motion, for  $C_{\alpha}$ – $C_{\alpha}$  bond 7 located in a highly active loop. Data calculated directly from the simulations (solid line), from eq 6 with dynamical renormalization of internal modes (dotted line), and without renormalization of the internal modes (dashed line).

## 6. DYNAMICS OF THE N–H DIPOLE VECTOR

When comparing to experiments probing dipolar relaxation, it is important to account for the fact that the experimental probe either refers to the dipole along the N–H or to the one along the C–H bond. The orientation and dynamics of both these probes is not necessarily equivalent to the dynamics of the  $C_{\alpha}$ – $C_{\alpha}$  bond vector, described by eq 6. The left panel of Figure 14



**Figure 14.** Left panel: the orientation of the bond basis ( $C_{\alpha}$ – $C_{\alpha}$ ) is in black and the N–H vector measured in  $^{15}\text{N}$  NMR is blue-yellow. Right panel: model representation of the relative position and orientation of the instantaneous and averaged bond vectors for the N–H and the  $C_{\alpha}$ – $C_{\alpha}$  bonds.

shows, as an example, the relative orientation of the N–H dipole and the  $C_{\alpha}$ – $C_{\alpha}$  bond vector in one sample configuration of the protein ubiquitin from simulations. The N–H dipole, relevant to  $^{15}\text{N}$  NMR backbone relaxation experiments of ubiquitin in this paper, has a different orientation than the related  $C_{\alpha}$ – $C_{\alpha}$  bond vector. The right panel of the same figure represents a model for the relative orientation of the instantaneous vectors,  $\vec{l}_{\text{NH},i}(t)$  and  $\vec{l}_{\text{C}_\alpha,i}(t)$ , and the averaged vectors,  $\langle \vec{l}_{\text{NH},i} \rangle$  and  $\langle \vec{l}_{\text{C}_\alpha,i} \rangle$ . In addition to a difference in average orientation, the  $C_{\alpha}$ – $C_{\alpha}$  vector and the N–H dipole vector can have different contributions to their internal dynamics, both additional short-time librational processes, and additional slower processes where the N–H bond can rotate with the peptide plane without affecting the overall backbone orientation.

The orientational bond autocorrelation for N–H bond relaxation

$$M_{1,\text{NH}}(t) = \frac{\langle \vec{l}_{\text{NH}}(t) \cdot \vec{l}_{\text{NH}}(0) \rangle}{\langle (l_{\text{NH}})^2 \rangle} \quad (36)$$

is calculated from the LE4PD theory by assuming that the modes form a complete set, which is equivalent to saying we can express the N–H bond vector as an expansion in the LE4PD normal modes as

$$\vec{l}_{\text{NH},i}(t) = \sum_{a=1}^{N-1} Q_{\text{NH},ia} \vec{\xi}_a(t) \quad (37)$$

where the transformation matrix  $\mathbf{Q}_{\text{NH}}$  is defined as

$$Q_{\text{NH},ia} = \langle \vec{l}_{\text{NH},i} \cdot \vec{\xi}_a \rangle / \langle \xi_a^2 \rangle \quad (38)$$

The bond autocorrelation function is then

$$M_{1,\text{NH},i}(t) = \sum_{a=1}^{N-1} A_{\text{NH},ia} \exp[-\sigma \lambda_a t] \quad (39)$$

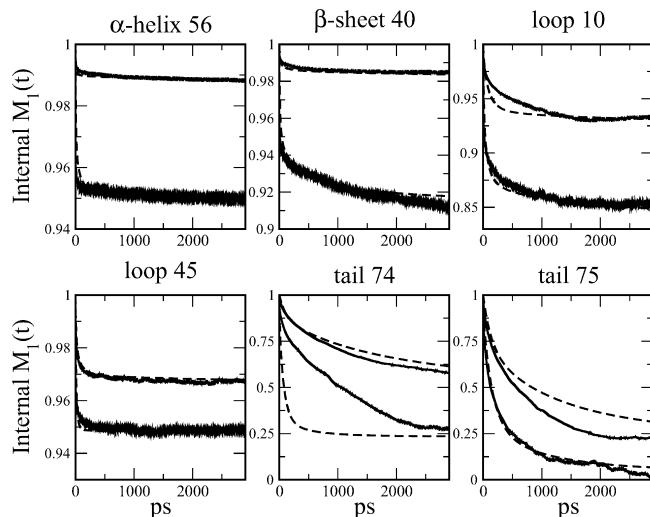
with

$$A_{\text{NH},ia} = \frac{Q_{\text{NH},ia}^2 \langle \xi_a^2 \rangle}{l_{\text{NH}}^2} \quad (40)$$

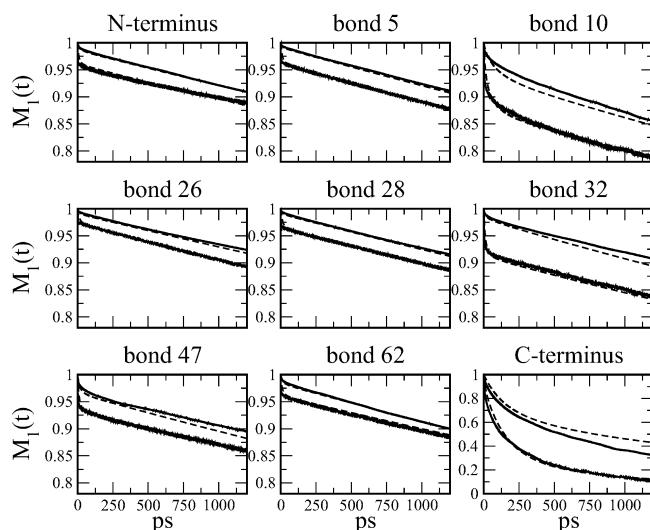
The expansion coefficient  $Q_{\text{NH},ia}$  can be calculated directly from the simulation using eq 38; in this fashion, it is possible to obtain the dynamics of any protein-based vector from the LE4PD modes.

For accuracy, when calculating the normal mode expansion of the N–H bond vector from the simulation coordinates, the local bond-order parameter  $\sum_{a=1}^3 A_{\text{NH},ia}$  is enforced to be the same as  $\langle \vec{l}_{\text{NH},i} \rangle^2 / \langle l_{\text{NH},i}^2 \rangle$  calculated directly from the simulation. This normalization is performed prior to enforcing the overall normalization  $\sum_{a=1}^{N-1} A_{\text{NH},ia} = 1$ . This is only a small correction of a few percent for all bonds except for those located in the C-terminal tail bonds 72–75. Here, the mode expansion fails and  $\sum_{a=1}^{N-1} A_{\text{NH},ia}$  calculated from the simulation before normalization is much less than 1. For these highly flexible C-terminal tail bonds, the  $C_\alpha$ – $C_\alpha$  bond dynamics are obtained well by the LE4PD theory, but the predictions for the N–H bond dynamics are much too fast, showing that there are large amplitude slow processes which are uncorrelated with protein backbone fluctuations in these bonds.

**6.1. Comparing the N–H Bond and the  $C_\alpha$ – $C_\alpha$  Dynamics to Simulation.** As a test of the normal mode solution of the Langevin equation for protein dynamics to describe the dynamics of both the N–H bond vector and the  $C_\alpha$ – $C_\alpha$  we compare the decay of the first-order autocorrelation function with data from computer simulations. Calculations were performed for each bond, and as an example, Figure 15 displays the comparison between theoretical predictions and simulations for the autocorrelation functions of both the N–H bond and the  $C_\alpha$ – $C_\alpha$  bond related to a representative bond in all secondary structure types. To emphasize the difference in the internal dynamics, in this figure, overall rotation is removed from simulation before calculating the time correlation functions, and the global mode contributions are removed from the theoretical calculation. Figure 16 compares the results of the full theory with rotation included for all of the polyubiquitination linkage sites, showing the high variability in the local dynamics at these biologically important residues. The comparison shows that the agreement is quantitative; similar quality of agreement is observed for all bonds along the primary



**Figure 15.** Temporal decay of the bond autocorrelation functions due to internal processes, where rotational relaxations have been subtracted, for the  $C_\alpha$ – $C_\alpha$  vector (top) and for the N–H vector (bottom) for bonds in many secondary structure types along the protein primary sequence. The dotted lines are calculated from the LE4PD normal mode expansion, solid lines are calculated directly from the simulation.



**Figure 16.** Temporal decay of the bond autocorrelation functions due to internal and rotational processes, for the  $C_\alpha$ – $C_\alpha$  vector (top) and for the N–H vector (bottom) for all poly ubiquitination linkage sites. The dotted lines are calculated from the LE4PD normal mode expansion, solid lines are calculated directly from the simulation.

sequence, up to the anomalous relaxation of the N–H bond vectors in some of the tail residues, supporting the validity of the model proposed in this paper. The theoretical model just presented proves to be fully consistent with the simulations, which provide the input quantities to the theory, in the short time regime where the simulations efficiently sample the local configurational space.

**6.2. Comparing the LE4PD Predictions to NMR Experiment.** As a second step in our study, we use the theoretical predictions for  $P_{2,i}(t) = (1/2)[3 \cos^2 \theta_i(t) - 1]$ , obtained from  $M_{1,i}(t)$  using eq 10, to calculate  $T_1$  and  $T_2$  relaxation times, and NOE, which are measured experimentally.  $^{15}\text{N}$  NMR backbone relaxation experiments are very sensitive

to the site-specific dynamics in the picosecond to the nanosecond regimes.<sup>53</sup> The target function is the bond relaxation for the N–H vector, which is calculated using information from the simulations. It is known that the forcefields used to calculate the N–H bond interaction in the atomistic simulation are approximate, because they are not designed for high accuracy at the single-hydrogen level. In addition, the LE4PD modes span the configurational space of the protein backbone but are not necessarily complete for the local N–H bond dynamics. To overcome these issues, we present three different methods to estimate the internal N–H bond dynamics. In all three methods, the global mode contributions  $A_{\text{NH},ia}$  for  $a = 1, 2$ , and 3 are calculated by evaluating eq 40 from the simulation, while the internal dynamics and the overall normalization between global and internal modes differ.

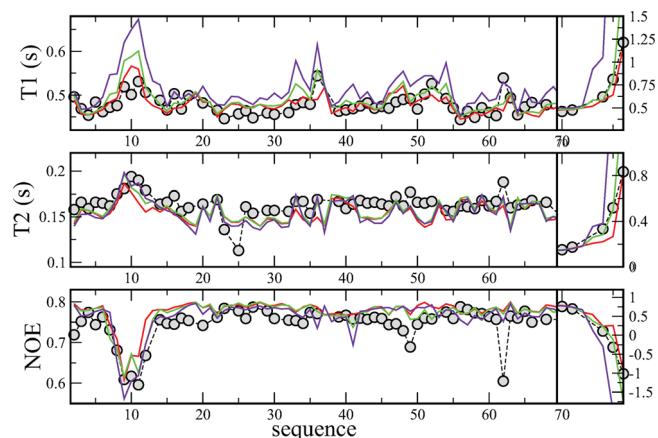
Each method is tested against three different sets of NMR experiments, those of Lee and Wand<sup>11</sup> at 298 K, and those of Tjandra et al.<sup>10</sup> and Lienin et al.<sup>22</sup> at 300 K. The solvent viscosity used in the theoretical expression is adjusted to account for the mixture of 90% H<sub>2</sub>O and 10% D<sub>2</sub>O used in all experiments.<sup>37</sup> The simulation conditions (temperature and salt concentration) were set to match the experiments of Lienin et al., and only the temperature and the temperature dependence of the viscosity in the theoretical expression was set to match the different experiments. The experimental data are for the most part self-consistent, and the theory agrees quite well with all experiments, with correlation coefficients between 0.792 and 0.983, depending on the method used to describe the internal dynamics of the N–H bond, as can be seen in Table 1. Figure

**Table 1. Correlation Coefficient of the Theoretical Models with Experimental Data of NMR Relaxation, Using the Three Different Methods to Estimate the Independent N–H Bond Fluctuations, (1) Assuming Identical Internal Dynamics to the A-Carbon Bond Basis up to an Independent Librational Process, (2) Fitting the Independent Dynamics of the N–H Bond to the Simulation, and (3) Using the Full LE4PD Mode Expansion of the N–H Bond Vector**

NMR reference and theoretical model	T1 Corr.	T2 Corr.	NOE Corr.	RMS Rel. Error
Lee <sup>11</sup> et al. 1	0.792	0.885	0.937	7.7%
Lee <sup>11</sup> et al. 2	0.875	0.974	0.967	5.9%
Lee <sup>11</sup> et al. 3	0.880	0.944	0.940	14.8%
Tjandra <sup>10</sup> et al. 1	0.806	0.879	0.948	13.7%
Tjandra <sup>10</sup> et al. 2	0.890	0.971	0.978	12.1%
Tjandra <sup>10</sup> et al. 3	0.914	0.938	0.955	20.6%
Lienin <sup>22</sup> et al. 1	0.959	0.936	0.967	7.0%
Lienin <sup>22</sup> et al. 2	0.960	0.960	0.970	6.6%
Lienin <sup>22</sup> et al. 3	0.880	0.983	0.982	16.6%

17 shows the comparison between the theoretical results and the experimental data of  $T_1$ ,  $T_2$ , and NOE from Lienin et al.<sup>22</sup> The different theoretical models proposed here provide comparable agreement with the experiments.

In the first method, the internal dynamics are taken to be identical to that of the  $C_\alpha$ – $C_\alpha$  vector, with the addition of a trivial contribution from the fast independent librational motion of the N–H bond. This librational motion is taken to be a subpicosecond process with identical magnitude for all bonds. From NMR experiments in oriented media, Ottiger and



**Figure 17.**  $T_1$ ,  $T_2$ , and NOE 15N NMR backbone relaxation, comparison between experiment (black),<sup>22</sup> the first theoretical method (red) correcting for N–H orientation and only adding an identical librational process onto the  $C_\alpha$ – $C_\alpha$  internal motion, the second theoretical method (green) accounting for differences in internal N–H fluctuations by fitting indendent N–H relaxation to simulation, and the third theoretical method with the full LE4PD expansion of the N–H bond vector (purple).

Bax estimated an average order parameter  $S$  for independent librational motion of the N–H bond from the backbone basis to be  $S = 0.94$ ,<sup>54</sup> corresponding to an approximate amplitude of the independent librational process in  $M_{1,\text{NH}}(t)$  of  $A_{\text{librational}} = 0.02$ . This gives the expression for the relaxation of the  $i$ th N–H bond

$$M_{1,\text{NH},i}(t) = C \sum_{a=1}^3 A_{\text{NH},ia} \exp[-\sigma\lambda_a t] + \sum_{a=4}^{N-1} A_{C\alpha,ia} \exp[-\sigma\lambda_a t] + A_{\text{librational}} \exp\left[-\frac{t}{\tau_{\text{lib}}}\right] \quad (41)$$

where  $\tau_{\text{lib}} = 0.2$  ps. The results are practically independent of  $\tau_{\text{lib}}$ , as long as it is taken to be a fast process. To enforce the correct normalization, the constant  $C = (\sum_{a=1}^3 A_{C\alpha,ia} - A_{\text{librational}})/\sum_{a=1}^3 A_{\text{NH},ia}$  so that  $M_{1,\text{NH},i}(0) = 1$ . As can be seen in Table 1, the  $\alpha$ -carbon-based internal dynamics of the backbone (model 1) correlates highly with the N–H bond dynamics measured experimentally for all three different sets of data.

The second method allows for an arbitrary independent relaxational mode for each N–H bond that is directly fit to the simulation; that is,

$$M_{1,\text{NH},i}(t) = C \sum_{a=1}^3 A_{\text{NH},ia} \exp[-\sigma\lambda_a t] + \sum_{a=4}^{N-1} A_{C\alpha,ia} \exp[-\sigma\lambda_a t] + A_{\text{indNH}}(t) \quad (42)$$

where  $A_{\text{indNH}}(t)$  is a three-exponential fit to simulations of the difference in the internal dynamics of the N–H bond and the  $C_\alpha$ – $C_\alpha$  vector,

$$A_{\text{indNH}}(t) = \frac{\langle \vec{l}_{\text{NH}}(t) \cdot \vec{l}_{\text{NH}}(0) \rangle}{\langle (l_{\text{NH}})^2 \rangle} - \frac{\langle \vec{l}_{C\alpha}(t) \cdot \vec{l}_{C\alpha}(0) \rangle}{\langle (l_{C\alpha})^2 \rangle} \quad (43)$$

evaluated in the body-fixed frame. Normalization is set by  $C = (\sum_{a=1}^3 A_{Ca,ia} - A_{indNH}(0)) / \sum_{a=1}^3 A_{NH,ia}$ . Table 1 shows that accounting for the difference between the local N–H bond dynamics and the backbone leads to lower error and higher correlation to experiments than the first method. The dynamics of the N–H bonds calculated with the second method for the highly flexible C-terminal tail agrees well with simulations; however, it disagrees with the experimental data for the tail-bond 75. This suggests the simulations predict dynamics that are too fast when compared with the experimental findings, indicating that some slow-energy barrier is not accounted for in the simulation.

In the third method, the full LE4PD mode expansion is used,  $M_{1,NH,i}(t) = C \sum_{a=1}^3 A_{NH,ia} \exp[-\sigma\lambda_a t]$ , with no additional independent processes added. Despite the errors in the C-terminal tail bonds, the overall agreement and correlation to experiment is nearly identical to the second method in which the difference in the internal dynamics between the  $\alpha$ -carbon basis and N–H bond is fitted (see Table 1 and Figure 17). Only the second method requires any fitting to a time-dependent quantity, and as can be seen in the Table, this improves the agreement with the experiment only modestly, if at all.

It is important to point out that the good agreement between theory and experiment has been obtained without the need of fitting the theory to experimental data, and in this way, our approach differs from other approaches commonly used to interpret NMR relaxation.<sup>55</sup> When a model is parametrized by fitting to the experimental data, it carries their uncertainty and errors, and it cannot directly relate the measured data to actual structural relaxation processes. In this respect, our approach has a clear advantage. The level of agreement between theoretical prediction, simulation, and experiment suggests that the LE4PD approach models the protein backbone dynamics with accuracy, while the disagreement observed for specific bonds can be related to insufficient sampling of the free-energy landscape that enters the relaxation dynamics of NMR, or possible experimental errors. The LE4PD approach has the advantage of being firmly grounded in the underlying physical processes which relax the dipole orientation.

## 7. CONCLUSIONS

We have presented a coarse-grained description for the dynamics of a folded protein in aqueous solution. The theory is closely related to the Rouse–Zimm model of the dynamics of synthetic, unfolded, polymers in dilute solutions but carefully incorporates the complexity of the protein free-energy landscape into a linear Langevin description of the protein dynamics. The theory is conveniently expressed in normal modes of motion. Because the modes are linearly independent, this facilitates the inclusion of correction terms that are specific of the dynamics spanned by each single mode.

The theory, which we call the Langevin equation for protein dynamics (LE4PD), modifies the Rouse–Zimm equation for both the global and internal modes. Since proteins have a specific tridimensional structure which is anisotropic, global modes must describe the fully anisotropic rotational diffusion of the molecule. The three global modes, properly modified, are able to capture the fully anisotropic rotational dynamics of the folded structure. The mode framework also allows for the calculation of the dynamics of bonds in the protein structure other than the  $\alpha$ -carbon bond basis chosen for the coarse-grained description. This is especially important when

comparing to experiments which probe the dynamics of specific dipoles measured in experimental techniques, here the N–H dipole whose relaxation is measured in <sup>15</sup>N NMR backbone relaxation.

To obtain a formalism that is solvable analytically, our treatment of the hydrodynamics involves preaveraging; in principle, other bead or shell hydrodynamical models may be more accurate in this regard.<sup>32</sup> However, while the treatment of the hydrodynamics can be more exact for a completely rigid protein, our approach captures correctly the essential physics of a collapsed polymer system with internal fluctuations. Typical synthetic polymers have an isotropic shape and all the monomers are statistically equally exposed to the solvent. This is not true for proteins; residues that are inside the protein, in the hydrophobic “core”, have very little contact with the solvent; however, they experience friction due to interactions with other protein residues. The physics of the tumbling of a rigid body, no matter how detailed the parametrization and hydrodynamical treatment, incorrectly attributes all frictional sources to solvent contacts. This would imply that the sites belonging to the hydrophobic core of the protein have little or no friction, which is unphysical. Furthermore, for the system of a protein in solvent there is no conservation law mandating that the friction associated with global diffusive processes be due only to direct solvent contacts. And it has been observed that the fit to experimental data in bead and shell hydrodynamical models for rotational diffusion always requires extra friction. We argue here that while the conventional inclusion of an adjustable bound layer of water and ions leads to quantitative comparison between rigid body hydrodynamical calculations and experiment,<sup>32</sup> the neglect of internal friction may be an additional factor not accounted for in the hydrodynamic modeling of biological polymers as rigid objects.

The dynamics of the internal modes predicted by the Rouse–Zimm equation for synthetic polymers is unrealistically fast due to neglecting the complex nature of the internal free-energy landscape. We have presented here a new model of a Langevin equation for protein dynamics, which includes a first-order correction where the timescale of relaxation in each internal mode is rescaled by the mode-specific mean free-energy barrier to orientational diffusion. After rescaling, we observe that there is no longer a separation in timescale between internal and global processes, even in the well-folded ubiquitin protein. Accounting for global anisotropy and the complexity of the internal free-energy landscape leads to simultaneous quantitative agreement with simulation and NMR backbone relaxation rates.

The Langevin equation for protein dynamics is based upon the inherent nature of proteins as polymers with both flexibility and global structure; it is a modified Langevin approach for polymer dynamics. The approach seamlessly describes both internal and rotational fluctuations, as well as dissipation in the internal hydrophobic core of the protein and the external solvent environment. With the ease of use of current simulation packages and the availability of computational power, when starting structures are available, the LE4PD approach can help bridging the gap between the often short timescales of simulations and longer timescales probed by experiments, providing a direct formal connection between the protein’s primary sequence, three-dimensional structure, free-energy landscape, and the dynamics.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: mguenza@uoregon.edu.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors thank James McCarty for useful discussions. This material is based upon work partially supported by the National Science Foundation under Grant CHE-1362500. This material is based upon work supported by the National Science Foundation under the Graduate STEM Fellows in K-12 Education (GK-12) program, Grant DGE-0742540. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant ACI-1053575.

## REFERENCES

- (1) Perico, A.; Guenza, M. Viscoelastic relaxation of segment orientation in dilute polymer solutions. *J. Chem. Phys.* **1985**, *83*, 3103–3109.
- (2) Perico, A.; Guenza, M.; Mormino, M.; Fioravanti, R. Protein dynamics: Rotational diffusion of rigid and fluctuating three dimensional structures. *Biopolymers* **1995**, *35*, 47–54.
- (3) Caballero-Manrique, E.; Bray, J. K.; Deutschman, W. A.; Dahlquist, F. W.; Guenza, M. G. A theory of protein dynamics to predict NMR relaxation. *Biophys. J.* **2007**, *93*, 4128–4140.
- (4) Bahar, I.; Attilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Design* **1997**, *2*, 173–181.
- (5) Attilgan, A.; Durell, S.; Jernigan, R.; Demirel, M.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505–515.
- (6) Go, N.; Noguti, T.; Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3696–3700.
- (7) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F.; Becker, S.; Meiler, J.; Grubmuller, H.; Griesinger, C.; De Groot, B. L. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **2008**, *320*, 1471–1475.
- (8) Long, D.; Brüschweiler, R. In silico elucidation of the recognition dynamics of ubiquitin. *PLoS Comput. Biol.* **2011**, *7*, e1002035.
- (9) Chu, J.-W.; Voth, G. A. Coarse-grained free energy functions for studying protein conformational changes: A double-well network model. *Biophys. J.* **2007**, *93*, 3860–3871.
- (10) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. Rotational diffusion anisotropy of human ubiquitin from <sup>15</sup>N NMR relaxation. *J. Am. Chem. Soc.* **1995**, *117*, 12562–12566.
- (11) Lee, A. L.; Wand, A. J. Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation. *J. Biomol. NMR* **1999**, *13*, 101–112.
- (12) Lakomek, N.-A.; Lange, O.; Walter, K.; Fares, C.; Egger, D.; Lunkenheimer, P.; Meiler, J.; Grubmuller, H.; Becker, S.; de Groot, B. L.; et al. Residual dipolar couplings as a tool to study molecular recognition of ubiquitin. *Biochem. Soc. Trans.* **2008**, *36*, 1433–1437.
- (13) Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press: Oxford, 2001.
- (14) Amadei, A.; Linssen, A.; Berendsen, H. J. Essential dynamics of proteins. *Proteins: Struct., Funct., Bioinf.* **1993**, *17*, 412–425.
- (15) Guenza, M. G. Theoretical models for bridging timescales in polymer dynamics. *J. Phys.: Condens. Matter* **2008**, *20*, 033101.
- (16) Bixon, M.; Zwanzig, R. Optimized Rouse–Zimm theory for stiff polymers. *J. Chem. Phys.* **1978**, *68*, 1896–1902.
- (17) Doi, M.; Edwards, S. *The Theory of Polymer Dynamics*; Clarendon Press: Oxford, 1986.
- (18) Sagnella, D. E.; Straub, J. E.; Thirumalai, D. Time scales and pathways for kinetic energy relaxation in solvated proteins: Application to carbonmonoxy myoglobin. *J. Chem. Phys.* **2000**, *113*, 7702–7711.
- (19) Bird, R. B.; Armstrong, R. C.; Hassager, O. *Dynamics of Polymeric Liquids*; Wiley-Interscience: Hoboken, 1987; Vol. 2: Kinetics.
- (20) Zwanzig, R.; Kiefer, J.; Weiss, G. H. On the validity of the Kirkwood-Riseman theory. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *60*, 381–386.
- (21) Guenza, M.; Perico, A. A reduced description of the local dynamics of star polymers. *Macromolecules* **1992**, *25*, 5942–5949.
- (22) Lienin, S.; Bremi, T.; Brüscheiler, R.; Ernst, R. Anisotropic intramolecular backbone dynamics of ubiquitin characterized by NMR relaxation and MD computer simulation. *J. Am. Chem. Soc.* **1998**, *120*, 9870–9879.
- (23) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950–1958.
- (24) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **1987**, *194*, 531–544.
- (25) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (26) Lindahl, E.; Hess, B.; Van Der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317.
- (27) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (28) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (29) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5915–5920.
- (30) Rawat, N.; Biswas, P. Size, shape, and flexibility of proteins and DNA. *J. Chem. Phys.* **2009**, *131*, 165104.
- (31) Harvey, S.; Garcia de La Torre, J. Coordinate systems for modeling the hydrodynamic resistance and diffusion coefficients of irregularly shaped rigid macromolecules. *Macromolecules* **1980**, *13*, 960–964.
- (32) García de la Torre, J.; Huertas, M. L.; Carrasco, B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **2000**, *78*, 719–730.
- (33) García De La Torre, J.; Navarro, S.; Lopez Martinez, M.; Diaz, F.; Lopez Cascales, J. HYDRO: A computer program for the prediction of hydrodynamic properties of macromolecules. *Biophys. J.* **1994**, *67*, 530–531.
- (34) García de la Torre, J.; Huertas, M.; Carrasco, B. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **2000**, *147*, 138–146.
- (35) Bernadó, P.; de la Torre, J. G.; Pons, M. Interpretation of <sup>15</sup>N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *J. Biomol. NMR* **2002**, *23*, 139–150.
- (36) Favro, L. D. Theory of the Rotational Brownian Motion of a Free Rigid Body. *Phys. Rev.* **1960**, *119*, 53–62.
- (37) Wong, V.; Case, D. A. Evaluating rotational diffusion from protein MD simulations. *J. Phys. Chem. B* **2008**, *112*, 6013–6024.
- (38) Smith, P. E.; van Gunsteren, W. F. The viscosity of SPC and SPC/E water at 277 and 300 K. *Chem. Phys. Lett.* **1993**, *215*, 315–318.
- (39) Guo, G.-J.; Zhang, Y.-G. Equilibrium molecular dynamics calculation of the bulk viscosity of liquid water. *Mol. Phys.* **2001**, *99*, 283–289.
- (40) Takemura, K.; Kitao, A. Water model tuning for improved reproduction of rotational diffusion and NMR spectral density. *J. Phys. Chem. B* **2012**, *116*, 6279–6287.

- (41) Rohrdanz, M. A.; Zheng, W.; Maggioni, M.; Clementi, C. Determination of reaction coordinates via locally scaled diffusion map. *J. Chem. Phys.* **2011**, *134*, 124116.
- (42) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Struct., Funct., Bioinf.* **1995**, *21*, 167–195.
- (43) Onuchic, J. N.; Luthey-Schulzen, Z.; Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
- (44) Noé, F.; Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–162.
- (45) Weinan, E.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*, 052301.
- (46) Best, R. B.; Hummer, G. Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732–6737.
- (47) Yang, S.; Onuchic, J. N.; Levine, H. Effective stochastic dynamics on a protein folding energy landscape. *J. Chem. Phys.* **2006**, *125*, 054910.
- (48) Berendsen, H. J.; Hayward, S. Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* **2000**, *10*, 165–169.
- (49) Komander, D. The emerging complexity of protein ubiquitination. *Biochem. Soc. Trans.* **2009**, *37*, 937–953.
- (50) Pettitt, B. M. *Advances in Chemical Physics, Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; John Wiley & Sons: Hoboken, 1990.
- (51) Lyubimov, I.; Guenza, M. First-principle approach to rescale the dynamics of simulated coarse-grained macromolecular liquids. *Phys. Rev. E* **2011**, *84*, 031801.
- (52) Zwanzig, R. Rate processes with dynamical disorder. *Acc. Chem. Res.* **1990**, *23*, 148–152.
- (53) Palmer, A. G., III NMR probes of molecular dynamics: Overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 129–155.
- (54) Ottiger, M.; Bax, A. Determination of Relative N-HN, N-C', Cα-C', and Cα-Hα Effective Bond Lengths in a Protein by NMR in a Dilute Liquid Crystalline Phase. *J. Am. Chem. Soc.* **1998**, *120*, 12334–12341.
- (55) Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.