

## A Theory of Protein Dynamics to Predict NMR Relaxation

Esther Caballero-Manrique, Jenelle K. Bray, William A. Deutschman, Frederick W. Dahlquist, and Marina G. Guenza

Department of Chemistry, University of Oregon, Eugene, Oregon

**ABSTRACT** We present a theoretical, site-specific, approach to predict protein subunit correlation times, as measured by NMR experiments of  $^1\text{H}$ - $^{15}\text{N}$  nuclear Overhauser effect, spin-lattice relaxation, and spin-spin relaxation. Molecular dynamics simulations are input to our equation of motion for protein dynamics, which is solved analytically to produce the eigenvalues and the eigenvectors that specify the NMR parameters. We directly compare our theoretical predictions to experiments and to simulation data for the signal transduction chemotaxis protein Y (CheY), which regulates the swimming response of motile bacteria. Our theoretical results are in good agreement with both simulations and experiments, without recourse to adjustable parameters. The theory is general, since it allows calculations of any dynamical property of interest. As an example, we present theoretical calculations of NMR order parameters and x-ray Debye-Waller temperature factors; both quantities show good agreement with experimental data.

### INTRODUCTION

Proteins are dynamic systems in which motion is characterized by an extended range of timescales. NMR is a powerful technique to study protein dynamics because it is sensitive to motions on many different timescales. The interpretation of NMR data, however, often involves model-dependent analyses. Indeed, several theoretical models have been successfully developed to interpret NMR relaxation data for protein motion. Nevertheless, such models often contain a finite number (sometimes large) of adjustable parameters that limit the theory's predictive power and generality.

Here we present a theoretical approach that is parameter free, since all of the physical quantities entering the theoretical description of protein dynamics are defined a priori from experiments and molecular dynamics simulations. Protein dynamics spanning a broad range of timescales are predicted as analytical functions of the eigenvalue/eigenvector solution to our equation of motion. We find good quantitative agreement between our theoretical predictions and the experimental and simulation data, without the use of adjustable parameters.

The dynamics of proteins occur over an extended range of time- and length-scales. For example, local bond librational motions can occur on picosecond timescales, whereas bond reorientation can occur on nanoseconds, and the global tumbling and cooperative interdomain motions can occur on tens of nanoseconds and longer timescales (1). Although local

dynamics depends on the chemical structure of the residue and on the local barriers that determine semiflexibility, global dynamics depends on the size of the protein and noticeably slows down with increasing protein molecular weight: timescales of milliseconds or longer are often observed in the global motions of large protein complexes. Due to the many orders of magnitude in timescale, and the many degrees of freedom available to the molecule, understanding the general principles of protein dynamics is a complicated problem.

Nevertheless, the development of a general understanding of protein local dynamics can provide important insights into many physical properties. It is widely recognized that the dynamics of proteins are relevant to function (2–5). For example, kinetic processes involving protein-protein or protein-ligand interactions can depend on conformational fluctuations. In many cases, protein stability is dominated by the thermodynamic entropy, which in turn is a function of the density of conformational states explored by the molecule and increases with increasing flexibility (2,3,6–8). Furthermore, there are subtle ways in which protein flexibility and dynamics can influence function. In allosteric mechanisms of activation, protein dynamics allow for structural fluctuations of coupled domains, which can transmit information between distant sites inside the protein. Such fluctuations can be responsible for promoting reactivity by bringing within close proximity catalytic sites that are otherwise distant. Finally, large cooperative motions are important for protein reactivity, because they can allow substrates to access internal regions of the protein that are sterically hindered.

The combination of experimental NMR methods and capable theoretical tools can provide a complete and quantitative picture of protein dynamics over a wide range of timescales. One might be tempted to directly model NMR relaxation by computer simulations. Unfortunately, this is generally not possible, since simulations cannot reach the longest timescales measured by NMR for most systems of interest

---

*Submitted May 3, 2007, and accepted for publication August 1, 2007.*

Address reprint requests to Marina G. Guenza, Dept. of Chemistry and Institute of Theoretical Science, University of Oregon, Eugene, OR 97403. Tel.: 541-3462877; Fax: 541-3464643; E-mail: mguenza@uoregon.edu.

William A. Deutschman's present address is Dept. of Chemistry, Westminster College, Salt Lake City, UT.

Frederick W. Dahlquist's present address is Dept. of Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, CA.

Editor: Arthur G. Palmer III.

(9–11). The longest timescale sampled by simulations is limited by computer power. For a typical computer, atomistically detailed simulations, explicitly including solvent molecules, are constrained to the nanosecond timescale. Such simulations contain data pertaining to local-scale motions for medium-sized proteins, i.e.,  $\sim 100$  amino acids. Alternatively, simulations can achieve longer than nanosecond timescales when employing multiscale coarse-graining algorithms. In the latter case, local scale information is averaged over, and the effects of solvent interactions enter through effective parameters. Thus, a complete description of the long-time dynamics of proteins, through atomistically detailed simulations, is limited to modest-sized proteins.

Several theoretical approaches have been developed to correlate NMR relaxation data to models of protein mobility. Of the initial attempts to model NMR relaxation by assuming specific mechanisms of diffusive dynamics (12–14), the most widely accepted is the model-free approach by Lipari and Szabo (15,16). In a seminal article they presented a formally simple but physically sound theory that maps the decay of the time autocorrelation function for the single bond vector  $i$ ,  $P_i^2(t)$  (defined in the next section) into a linear combination of two uncorrelated dynamical processes. The theory in its simplest form requires three fitting parameters: the local and global correlation times, and a generalized order parameter determined by the weights assigned to the two dynamical processes. The main contribution of the Lipari-Szabo approach is that it allows for an analysis of NMR relaxation data to provide an approximate, although fairly realistic, picture of the local motions of the protein. Later theoretical approaches correlate the measured order parameters to protein entropy (17,18) and function.

The assumptions in Lipari-Szabo theory are consistent with the dynamics of flexible bonds in a globular protein. For this system, local motions follow two uncorrelated decay processes given by the local bond dynamics and the overall protein tumbling. However, when more complex systems are involved, the original three-parameter theory becomes inadequate. Proteins that do not obey the simple three-parameter scheme are, for example: a), proteins with rigid structure, where local and global dynamics both occur on the slow timescale of the overall protein rotation; b), partially unfolded proteins in which local and global motions are both relatively fast; c), proteins with asymmetric shapes, for which rotational diffusion is anisotropic and the slow motion includes at least three contributions corresponding to the tensorial components of the overall tumbling motion (19); and d), proteins composed of two or more large domains that slowly move relative to one another (e.g., the protein calmodulin). In general, experimental data for such proteins can be reproduced by the theory only at the expense of introducing a high number of fitting parameters (20–23).

The complexity of the underlying dynamics, even for a simple flexible globular protein, has been made apparent in the recent work by Brüschweiler and co-workers. In the

Reorientational Eigenmode Dynamics theory (24,25), local reorientation of backbone bonds is obtained from simulations by initially eliminating the overall tumbling motion from molecular dynamics (MD) trajectories. The global dynamic modes are a posteriori calculated through a fitting of correlation times. This procedure is repeated until good numerical agreement with the data is achieved. It is clear from this model that, even for globular proteins, the dynamics is a combination of many slow dynamical processes.

The theory we present in this article approaches the modeling of NMR relaxation and protein dynamics from a new perspective. In our approach, we derive a first principles equation of motion for protein dynamics that correctly describes a broad range of time regimes, and provides a good prediction of different NMR experiments, (i.e.,  $T_1$ ,  $T_2$ , and nuclear Overhauser effect (NOE)), x-ray crystallography temperature factors, and computer simulations. We are interested in pursuing this alternative-avenue, first-principles approach for the following reasons: it should provide a physically accurate picture of protein dynamics within a general framework, applicable to most proteins and experimental methods. Although parameter-dependent models can achieve good agreement with experiments by including an increasing number of adjustable parameters, this is often accomplished at the expense of physical self-consistency and broad predictive power.

Our approach combines MD simulations and the solution of a diffusion equation in the spirit of our early work (26). However it differs from mode-coupling dynamics approaches (27,28), as developed by La Penna, Perico, and Freed, in many fundamental points. First of all, we are not interested in including higher-order bond-bond correlations. Second, in the proteins we investigate, the solvation dynamics of water molecules are uncorrelated with protein motions so that there is no modification of the protein friction. Finally, we adopt an original form for the hydrodynamic interaction that properly accounts for hydrophobic regions in the protein structure.

Although atomistically detailed simulations provide the needed information for the short-time dynamic regime, the long timescales of protein dynamics is governed mainly by hydrodynamics (29–31). It is known that conventional equations of motion, which are quite accurate in representing the dynamics of synthetic macromolecules (32), do not apply to proteins unless arbitrary numerical prefactors in the hydrodynamic interaction are included. The difficulty with conventional hydrodynamic interaction formalisms to treat biological systems is ascribed to the following: although macromolecules of synthetic origin explore a large conformational space, protein dynamics are characterized by fluctuations around the native conformation, which is stabilized by hydrophobic interactions. Each residue in the folded structure may be partially exposed to the protein hydrophobic core, and partially exposed to the solvent. This effect is essential in determining the dependence on solvent viscosity of the rate of conformational transitions in proteins, as shown

by Ansari et al. in a study of myoglobin dynamics after ligand dissociation (33). The presence of hydrophobic regions modifies local friction coefficients and requires a new formalism for the hydrodynamic interaction, as we discuss further below.

In our procedure, we begin by performing molecular dynamics simulations of a protein in its solvent described at the atomistic level. Simulations provide a trajectory of  $\sim 4$  ns during which the protein is stable and exhibits fluctuations about its equilibrium configuration. From this relatively short “equilibrium” trajectory, we calculate the statistical quantities that characterize the matrices in the equation of motion for the time evolution of the protein spatial coordinates. The equation is solved by diagonalization of these matrices, yielding eigenvalues and eigenvectors that characterize the spectral density. From the latter, we calculate NMR relaxation parameters, which we compare with experimental  $^1\text{H}$ - $^{15}\text{N}$  NOE, spin-lattice relaxation time ( $T_1$ ), and spin-spin relaxation time ( $T_2$ ). We test our procedure against NMR data for the dynamics of the response regulator CheY that controls chemotaxis in *Escherichia coli*. We find that theoretical predictions and experimental NMR data are in good agreement. Moreover, to formally connect our approach to Lipari-Szabo’s model, we calculate the sequence-dependent order parameter,  $S_i^2$ , which is found to agree well with the corresponding values extracted from NMR experiments.

The response regulator protein CheY is part of a two-component regulatory system that controls the chemotactic swimming response of motile bacteria. Understanding the dynamics of CheY is in general relevant, since two-component regulatory systems are conserved through all bacteria, archaea, and low eukaryotes (34). Specifically in bacteria, response regulators control gene expression, chemotaxis, antibiotic resistance, and many other processes (35). CheY is an ideal protein to investigate since it is the best characterized member of the response regulator superfamily (36).

As an additional test of our theory, we use our analytical solution of the equation of motion to calculate the mean-square displacements of  $\alpha$ -carbon atoms in the picosecond regime. We observe that our theoretical results agree well with experimental data of temperature factors for CheY, as measured with high-resolution x-ray crystallography by Volz and Matsumura (36). The good agreement we obtain between our theoretical predictions and both the x-ray and NMR experimental data underlines the power and utility of our general approach.

The remainder of this article is organized in the following manner. In the Theory section, we present the Langevin equation for protein dynamics. The section includes a new treatment of the hydrodynamic interaction that accounts for varying degrees of residue exposure to the solvent, from the completely screened to the fully exposed. In the Methods section, we describe the method to perform molecular dynamics simulations, as well as the methodology we used to analyze simulation data. This section includes a brief description of the NMR relaxation experiments. In the

Results section, we present the theoretical predictions of our approach for a series of experimentally measured quantities. In all of our analyses, we directly compare theory with simulations and experimental data for NMR relaxation and order parameters, and Debye-Waller temperature factors from x-ray crystallography.

## THEORY

### The Langevin equation for protein dynamics

According to Bloch, Wangsness, and Redfield (37–40), spin relaxation parameters measured by NMR are functions of the spectral density,

$$J(\omega) = \frac{2}{5} \int_0^\infty P_2(t) \cos(\omega t) dt, \quad (1.1)$$

defined as the Fourier cosine transform of the orientational autocorrelation function  $P_2^i(t)$ . The latter is the second-order Legendre polynomial of the cosine of the angle,  $\theta_i(t)$ ,

$$P_2^i(t) = \frac{3}{2} \langle \cos^2 \theta_i(t) \rangle - \frac{1}{2}. \quad (1.2)$$

In Eq. 1.2,  $\theta_i(t)$  is the angle between the bond vectors  $\mathbf{l}_i(0)$  and  $\mathbf{l}_i(t)$ , and describes the bond vector reorientation during the time interval  $t$ .

For a semiflexible macromolecule, the function  $P_2^i(t)$  assumes a simple analytical form (41), expressed as a function of eigenvalues and eigenvectors of the matrix product  $\mathbf{H}\mathbf{A}$ , which governs the equation of motion for the time evolution of the spatial coordinates for the protein in solution (see Eqs. 1.6, 1.7, and 1.8). Here  $\mathbf{H}$  is the matrix of hydrodynamic interaction and  $\mathbf{A}$  is the structural matrix, which are defined in the following sections. In this context, the goal is to define an equation of motion that properly accounts for the dynamic properties of the protein.

The equation of motion for the time evolution of the spatial coordinates is a diffusive Langevin equation, where the inertial term is neglected because protein dynamics are overdamped (29,31,41–44). In our approach the protein is represented as a collection of effective units, or residues, centered on the position of the  $\alpha$ -carbon. In this way, the structure of the protein appears to be coarse-grained at the level of its primary sequence of amino acids. The time evolution of the spatial coordinate of the  $i$ th residue inside a protein, i.e., the vector  $\mathbf{R}_i(t)$ , obeys a Langevin equation (LE), given by the balance of forces acting on the residue (29,31,32,41–44)

$$\zeta \frac{\partial \mathbf{R}_i(t)}{\partial t} = -\frac{3k_B T}{l^2} \sum_{j,k=0}^{N-1} H_{ij} A_{jk} \mathbf{R}_k(t) + \mathbf{F}_i(t), \quad (1.3)$$

where  $k_B$  is Boltzmann constant,  $T$  is the temperature, and  $\zeta = N^{-1} \sum_{i=0}^{N-1} \zeta_i$  is the friction coefficient obtained from the average over the residue-dependent local friction coefficient

$\zeta_i$ . It is also useful to define the following quantities,  $\sigma = 3/(\zeta\beta l^2)$ ,  $\beta^{-1} = k_B T$ , that will be used later on in this article. In Eq. 1.3, the viscous force on the left-hand side,  $\zeta(\partial \mathbf{R}_i(t)/\partial t)$ , is balanced on the right-hand side by the intramolecular force,  $-3k_B T/l^2 \sum_{j,k=0}^{N-1} H_{ij} A_{jk} \mathbf{R}_k(t)$ , and by the random force,  $\mathbf{F}_i(t)$ . The random force is assumed to be a Gaussian white noise process with zero mean. The LE is an inhomogeneous, first-order differential equation, linear in  $\mathbf{R}$ , that is solved by diagonalization of the product of matrices  $\mathbf{H}\mathbf{A}$

$$\sum_{i,j,k=0}^{N-1} Q_{a,i}^{-1} H_{ij} A_{jk} Q_{k,b} = \delta_{a,b} \lambda_a. \quad (1.4)$$

The matrix of eigenvectors,  $\mathbf{Q}$ , also diagonalizes  $\mathbf{A}$  through the congruent transformation

$$\sum_{i,j=0}^{N-1} Q_{i,a} A_{ij} Q_{j,b} = \delta_{a,b} \mu_a, \quad (1.5)$$

where  $\mu_a$  defines the mean-square length of the normal mode  $a$  as  $\langle \xi_a^2 \rangle = l^2 \mu_a^{-1}$ .

Perico and Guenza have shown that for a macromolecule, the autocorrelation function  $P_2^i(t)$  can be expressed as a function of  $M_1^i(t)$ , following the relation (41)

$$P_2^i(t) = 1 - 3 \left[ x^2 - \frac{\pi}{2} x^3 \left( 1 - \frac{2}{\pi} \arctan x \right) \right], \quad (1.6)$$

with

$$x = [1 - (M_1^i(t))^2]^{1/2} / M_1^i(t). \quad (1.7)$$

The time correlation function  $M_1^i(t)$  characterizes the disorientation of bond  $\mathbf{l}_i(t)$  and is defined by eigenvalues and eigenvectors of the matrices  $\mathbf{A}$  and  $\mathbf{H}$  according to

$$M_1^i(t) = \frac{\langle \mathbf{l}_i(t) \cdot \mathbf{l}_i(0) \rangle}{\langle |\mathbf{l}_i(t)| \rangle \langle |\mathbf{l}_i(0)| \rangle} = \sum_{a=0}^{N-1} (Q_{i+1,a} - Q_{i,a})^2 \mu_a^{-1} \exp\{-\sigma \lambda_a t\}. \quad (1.8)$$

The condition applies that  $\sum_{a=1}^{N-1} (Q_{i+1,a} - Q_{i,a})^2 \mu_a^{-1} = 1$ , i.e., the normalized average length of  $i$ th the bond. The time correlation function is a linear combination of  $N$  decay processes characterized by the mode index  $a$ . Each decay process has a weight of  $(Q_{i+1,a} - Q_{i,a})^2 \mu_a^{-1}$  and a correlation time  $\tau_a = (\sigma \lambda_a)^{-1}$ , which shows how negative eigenvalues correspond to unphysical correlation times. Two points emerge from this description: i), the time correlation functions characterizing protein dynamics contain contributions from  $N$  correlation times, where  $N$  in our model corresponds to the number of residues in the protein; ii), the function  $P_2^i(t)$  for a macromolecule is, in general, not a simple linear combination of exponential functions (41).

Eqs. 1.6–1.8 define the autocorrelation function  $P_2^i(t)$  as a function of the eigenvalues,  $\lambda_a$ , and eigenvectors,  $Q_{i,a}$ , of the matrix product  $\mathbf{H}\mathbf{A}$ , and of the eigenvalues,  $\mu_a$ , of the connectivity matrix  $\mathbf{A}$ . Therefore the quality of the predicted autocorrelation functions, depends on the accuracy of the definition of matrices  $\mathbf{H}$  and  $\mathbf{A}$ .

## Definition of the structural matrix $\mathbf{A}$

If the matrices  $\mathbf{H}$  and  $\mathbf{A}$  are known, it is straightforward to calculate  $P_2^i(t)$  and the spectral density. Following the approach developed by Bixon and Zwanzig to treat the dynamics of semiflexible macromolecules (43), the matrix  $\mathbf{A}$  is defined as

$$\mathbf{A} = \mathbf{M}^T \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{U} \end{pmatrix} \mathbf{M}, \quad (2.1)$$

and is a function of the equilibrium average bond correlation matrix (41,44)

$$U_{ij}^{-1} = \langle (\mathbf{l}_i \cdot \mathbf{l}_j) \rangle / \langle |\mathbf{l}_i| |\mathbf{l}_j| \rangle, \quad (2.2)$$

and of the connectivity matrix  $\mathbf{M}$ , where all elements equal zero except  $M_{0,i} = 1/N$  with  $i = 0, \dots, N-1$ ,  $M_{i,i} = 1$ , and  $M_{i,i-1} = -1$  with  $i = 1, \dots, N-1$ . The matrix  $\mathbf{A}$  defines the effective mean-force potential,

$$V(\{\mathbf{R}\}) = \frac{3}{2\beta l^2} \sum_{i,j=0}^{N-1} A_{ij} \mathbf{R}_i \cdot \mathbf{R}_j, \quad (2.3)$$

which has been successfully adopted in theories of protein folding to describe the final state of the folding process (45). As a consequence of Eq. 2.3, the distribution of any vector connecting two residues is Gaussian. We tested this property against simulations and we see that for all the quantities investigated (i.e., effective bond length, end-to-end distance, radius-of-gyration, and dihedral angles, among others) the Gaussian shape of the distribution function is fulfilled to a good degree of approximation (29).

With a slightly different form of the matrix  $\mathbf{A}$  our approach recovers known models of protein dynamics. For example, if the residue potential is assumed to be harmonic, and uniform for each residue, while local intra- and intermolecular connectivity are included through the matrix  $\mathbf{M}$ , our approach recovers a Gaussian network model (46–48), implemented to include hydrodynamic interaction. The traditional Gaussian network model is recovered when hydrodynamic interaction is neglected, and  $\mathbf{H} = \mathbf{1}$ .

## Definition of the hydrodynamic interaction matrix $\mathbf{H}$

The generic element of the hydrodynamic interaction matrix,  $H_{ij}$ , describes how the motion of the  $i^{\text{th}}$  residue produces instantaneous waves in the solvent, which perturb the velocity of the fluid surrounding a generic residue  $j \neq i$ . The perturbation of the velocity decreases as the inverse of the interresidue distance  $R_{ij}$ ; in this way the hydrodynamic interaction is a long-ranged perturbation, which affects large-distance and long-time scale dynamics of the protein.

Because the perturbation propagates through the solvent, in hydrophobic regions, which are not in contact with the solvent, the hydrodynamic interaction becomes screened and its perturbation on the dynamics is negligible (29). This effect is not accounted for in the conventional expression of the hydrodynamic interaction matrix (29),

$$H_{ij} = \delta_{ij} + (1 - \delta_{ij}) \frac{\bar{\zeta}}{6\pi\eta_w} \left\langle \frac{1}{R_{ij}} \right\rangle, \quad (2.4)$$

where  $\bar{\zeta} = 6\pi\eta_w r_H$  is the average residue friction,  $\eta_w$  is the solvent viscosity, and  $r_H$  is the average hydrodynamic radius of the residue. In Eq. 2.4, each residue is assumed averaged over the solvent degrees of freedom, which is a good approximation for flexible macromolecules since they explore an extended conformational space; i.e., each residue is statistically exposed to the solvent. Nevertheless, for folded proteins with residues buried in the hydrophobic core, self- and cross-contributions in Eq. 2.4 are not properly balanced. This often leads to negative eigenvalues, and unphysical correlation times, which implies time correlation functions that diverge at long time, instead of decaying to zero (see Eq. 1.8 with  $\lambda_a < 0$  and Fig. 1).

To take into account the presence of residues partially or totally buried in the protein, we derive a new form for the hydrodynamic matrix. We couple to the traditional contribution for solvent exposed regions (49), a solvent-screened equation of motion driving the dynamics of the regions excluded from the solvent. The perturbation of the velocity of residue  $j$ , due to the motion in the solvent of residue  $i$ , is given by  $\mathbf{v}'_j = \sum_{i \neq j=0}^{N-1} \mathbf{T}_{ji} \cdot \mathbf{F}_i^\zeta$ , where  $\mathbf{T}_{ji} \approx (1/6\pi\eta_w) \langle 1/R_{ji} \rangle \mathbf{1}$  is the preaveraged Oseen tensor.

In this new formalism, each residue has a friction coefficient,  $\zeta_i$ , that contains a contribution from the area exposed to solvent,  $\zeta_{i,w}$ , and a contribution from the area screened from the solvent (buried),  $\zeta_{i,p}$ . The effective friction coefficient for each residue,  $\zeta_i$ , is given by a simple extension of Stokes' law as

$$\zeta_i = \zeta_{i,w} + \zeta_{i,p} = 6\pi(\eta_w r_{i,w} + \eta_p r_{i,p}), \quad (2.5)$$

where  $r_{i,w}$  is the radius of a spherical bead of surface area equal to the surface exposed to the solvent (50), whereas  $r_{i,p}$

is the radius of a spherical bead of surface area equal to the surface shielded from the solvent. The internal viscosity of the hydrophobic core of the protein,  $\eta_p$ , is assumed to be twice the magnitude of the viscosity of the water solvent,  $\eta_w$ . This value is chosen on the physical grounds that residues in the hydrophobic core move in a "liquid" of hydrophobic particles, similar to the environment of a monomer in a melt of polymer chains (51). The exact calculation from the decay of the velocity autocorrelation function provides a consistent estimate (52).

The hydrodynamic force due to residue  $i$ ,  $\mathbf{F}_i^\zeta = \zeta_{i,w} \mathbf{v}_i$ , contains the friction related to the exposure to solvent of residue  $i$ ,  $\zeta_{i,w}$ . To allow the solution of Eq. 1.3 through a normal mode representation, we approximate  $\zeta_{i,w}$  as the total friction on residue  $i$ ,  $\zeta_i$ , weighted by the ratio of the average contribution due to solvent exposure,  $\zeta_w$ , to the average total friction coefficient,  $\zeta$ :

$$\zeta_{i,w} \approx (\zeta_w / \zeta) \zeta_i, \quad (2.6)$$

where the average friction coefficients are defined as  $\zeta_w = N^{-1} \sum_{i=0}^{N-1} \zeta_{i,w}$  and  $\zeta = N^{-1} \sum_{i=0}^{N-1} \zeta_i$ .

Once Eqs. 2.5 and 2.6 are included in our derivation, by balancing the hydrodynamic drag force on the  $j$ th residue,  $\zeta_j \mathbf{v}_j$ , with the intramolecular force and the solvent Brownian force, we obtain the equation of motion given by Eq. 1.3, with a novel form of the hydrodynamic interaction matrix, namely

$$H_{ij} = \frac{\zeta}{\zeta_i} \delta_{ij} + (1 - \delta_{ij}) \frac{\zeta_w}{6\pi\eta_w} \left\langle \frac{1}{R_{ij}} \right\rangle. \quad (2.7)$$

The cross-terms in Eq. 2.7 only account for the contribution of residues exposed to the solvent, whereas the self-terms include the contributions of both buried and solvent exposed residues. Equation 2.7 therefore describes buried residues as immersed in a "liquid" of polypeptides, i.e., following a Rouse-like approach, whereas residues exposed to the solvent follow the dynamics of macromolecules in solutions, i.e., a Rouse-Zimm-like approach (29). Partially exposed residues are represented by a weighted combination of the two contributions.

Equation 2.7 is general and holds also for nonbiological macromolecules since it recovers the conventional definition for the dynamics of unperturbed macromolecules of synthetic origin (29), Eq. 2.4, if all the residues are statistically exposed to solvent, i.e., no regions of hydrophobicity are present in the molecule. As mentioned before, if the presence of hydrophobic regions in the dynamics of proteins is overlooked, the diagonalization of the matrix product  $\mathbf{H}\mathbf{A}$  leads, in most cases, to negative eigenvalues, which correspond to unphysical negative correlation times. An arbitrary parameter, the so-called reduced friction coefficient, was previously included as a weight of the cross-contributions to overcome this problem (44). With the new formulation for the hydrodynamic interaction proposed here, no arbitrary parameters need be included.

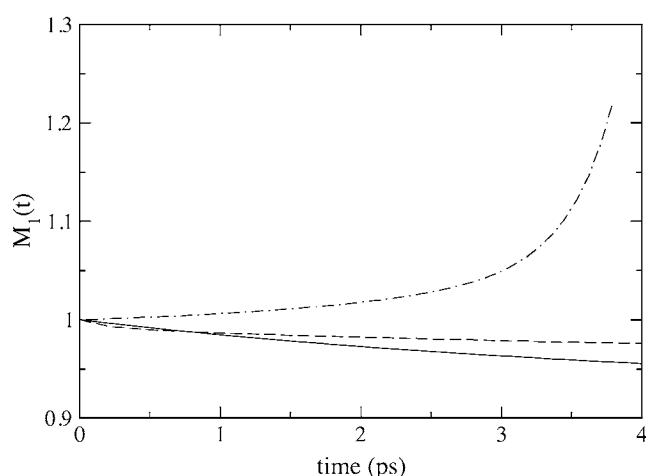


FIGURE 1 Time correlation function  $M_1(t)$ , for the first  $\text{C}^\alpha\text{-C}^\alpha$  bond, calculated using different forms of the hydrodynamic matrix. Hydrodynamic interaction from Eq. 2.4 produces an unphysical, diverging correlation function (dotted-dashed line), whereas Eq. 2.7 leads to a function (solid line) in reasonable agreement with simulation data (dashed line).

## METHODS

### Solving the Langevin equation

The averaged statistical quantities that enter matrices  $\mathbf{H}$  and  $\mathbf{A}$  are calculated from trajectories of MD simulations of the protein in its solvent, performed using the experimental values of the thermodynamic parameters. These include the average normalized statistical correlation between effective bonds defined in Eq. 2.2, as well as the average distance between each pair of  $\alpha$ -carbons,  $\langle R_{ij} \rangle$ , which enters the hydrodynamic matrix Eq. 2.7. We also calculate, for each residue, the average surface area available to solvent as well as the average surface area exposed to the hydrophobic region, which define the local friction coefficients entering Eq. 2.7, through the definition in Eq. 2.5. In this way, all the input parameters to our theory are known, as well as the numerical values of the matrices  $\mathbf{H}$  and  $\mathbf{A}$ .

The procedure just outlined extracts the relevant physical information from short-time all-atom simulations and, through the solution of the equation of motion, allows us to predict experimentally measurable dynamical properties of proteins in a broad range of timescales. The key idea is that if initial position and forces acting on each residue are known, it should be possible to make predictions about its motion for any timescale. This is analogous to solving Newton's equation of motion to predict the trajectory of a point-mass, when its initial position and forces acting on it are known. The idea of using computer simulations as input to diffusive equations is well established (26,32,53–55). In the following sections, we test our approach by investigating the dynamics of the signal transduction protein CheY, as measured in simulations and experiments of NMR relaxation.

### Simulations and NMR experiments of CheY

In the specific case of *E. coli* CheY, we start with the x-ray structure from the Protein Data Bank (entry 3CHY (36)) and perform molecular dynamics simulations using the GROMACS package (56) (GROMOS96 force field) for the protein in a box containing  $\sim 3000$  water molecules, with periodic boundary conditions. Density and temperature are set to those from experiments. The temperature was controlled using a Berendsen algorithm at 300 K with a coupling constant of 100 fs. The cutoff for the Van der Waals and Coulomb interactions was 1.4 nm, with neighbor lists updated every fifth step. The x-ray structure presents seven amino acids in two different rotameric configurations: Asn-23, Asn-32, Glu-37, Ser-56, Met-85, Tyr-106, and Leu-127. All the rotamers interconvert during the simulation with the exception of Tyr-106: we perform two distinct sets of runs starting with CheY having the side chain in Tyr-106 either exposed to solvent (the CheYa conformation) or embedded in a hydrophobic groove (the CheYb conformation). After energy minimization and system equilibration, the longest simulation run for CheYa extended up to 19 ns whereas the longest run for CheYb included 21 ns. Long time “stretches” in the simulation, during which the structure was fluctuating around a transient stable configuration, were identified. A stable folded configuration was defined by having the root mean-square deviation of the  $\alpha$ -carbon positions from the initial equilibrated structure, smaller than  $\sim 4$  Å. We identified one long stable “stretch” of  $\sim 4$  ns for CheYa and of  $\sim 3$  ns for CheYb. From each stretch, we calculated statistical averages and residue-specific friction coefficients, which are input to the protein equation of motion. Available surface areas (ASA) are calculated for each picosecond of simulation (57,58). The contribution to the friction coefficient due to exposure to the hydrophobic region is in turn obtained by subtracting the ASA from the total surface area of the residue (59), and it is also calculated for each picosecond of MD trajectory.

For each NMR experiment a single sample was used, containing  $\sim 1$  mM concentration of  $^{15}\text{N}$ -labeled CheY protein in 50 mM phosphate buffer. Experiments were carried out on a 600-MHz spectrometer, equipped with a triple resonance probe and pulse field gradient, at a  $^1\text{H}$  frequency of 599.98 MHz, and a  $^{15}\text{N}$  frequency of 60.8 MHz, at 25°C. Spectra were collected for the measurement of longitudinal relaxation rates ( $T_1$ ),  $^{15}\text{N}$  transverse relaxation rates ( $T_2$ ), and the  $^1\text{H}$ - $^{15}\text{N}$  steady-state NOE of the  $^{15}\text{N}$  nuclei in CheY.

## RESULTS

### Comparison of theoretical predictions against simulations: test of the theory in the short-time regime

As a test of consistency for the theory, we compare time correlation functions calculated from eigenvalues and eigenvectors of the equation of motion, with the ones obtained directly from the simulation trajectories that provided the input parameters to the theory. If the equation of motion is correct and contains all the needed information the two should be identical. We calculate the decay of the time correlation function for bond disorientation,  $M_1^i(t)$ . The comparison is limited to intervals of 1.5 ns, since the longest simulation “stretch” during which the protein fluctuates around a stable configuration has a total length of  $\sim 4$  ns for CheYa, and  $\sim 3$  ns for CheYb. We find that the theoretical predictions of bond disorientation, as a function of time, are in quantitative agreement with simulations for all the bonds along the primary structure of the rotamer CheYb, whereas a more complex picture emerges for the CheYa rotamer.

To summarize our results in a visual picture and rapidly identify bonds presenting this “anomalous” decay of the correlation function, we plot in Fig. 2 the mean-square difference between theoretical and simulated  $M_1^i(t)$  at fixed time intervals (0.5 ns, 1.0 ns, and 1.5 ns), for each bond along the primary sequence of CheYa.

Fig. 2 shows that in the CheYa rotamer, most of the bonds display quantitative agreement with simulated bond disorientation with the exception of two main peaks clearly above the “noise” level, which correspond to regions including residues 13–15, and bond 117. Uniform “noise” appears in the baseline, more pronounced with increasing time, which is a consequence of the numerical error accumulating in the simulations (60,61). The disagreement between theory and simulations for bonds 13–15, and 117 is due to the presence of local conformational barriers (62), which are not accounted for in the theory. Fig. 3 shows as an example of this effect, the decorrelation in time for bonds 70–72, which are well behaved, and the decorrelation in time of bonds 13–15, which are part of the region that displays anomalous relaxation. For bonds 70–72, theoretical predictions and simulations of the time correlation function superimpose, whereas for bonds 13–15 the theory predicts faster motion than observed in simulations. Each sequence of three bonds defines the comprised dihedral angle. For sequence 70–72, the distribution of dihedral angles in the simulation follows Gaussian statistics, which is consistent with our theory. However, in the region of anomalous decorrelation the distribution of the comprised dihedral angle shows two stable conformational states, which cannot be described by our approach in its present form. Further development of the theory will require the inclusion of activated barrier crossing.

It is interesting to consider the possible implications of the different local conformational behaviors that we observe for the two CheY rotamers. The comparison between theory and

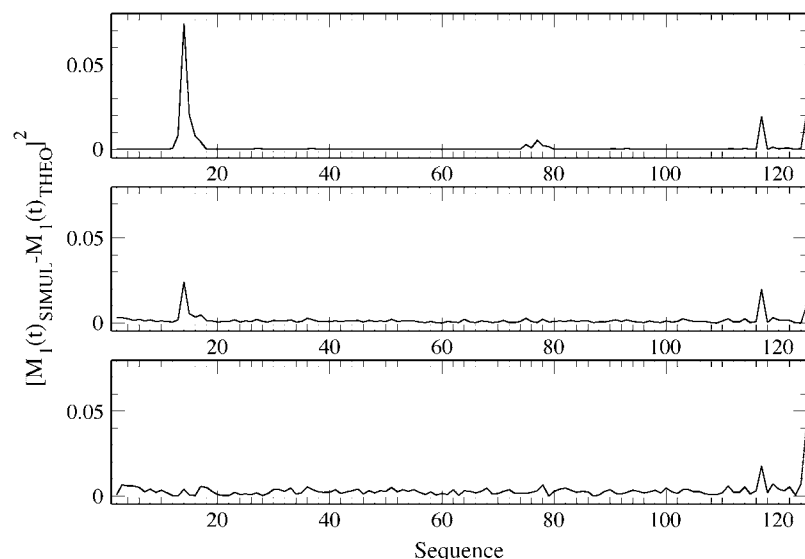


FIGURE 2 Square difference of the bond time correlation function calculated from theory and from simulations as a function of protein sequence at a fixed time interval. From top to bottom panels, data are reported at increasing time intervals (500, 1000, and 1500 ps) for the rotamer CheYa of the signal transduction protein CheY.

simulation identifies regions, along the protein primary structure, characterized by small conformational energy barriers, easily overcome during simulations, but still relevant in affecting the local bond reorientation. These barriers are present only in one of the two main rotamers of CheY, specifically CheYa. The two rotamers differ in the position of the side chain of residue Tyr-106, which is exposed to solvent in CheYa and buried in a hydrophobic pocket in CheYb. Residue 106 is a highly conserved aromatic amino acid, either tyrosine or phenylalanine, in the extended family

of response regulators, and is relevant in the mechanism of signal transduction. The external position of the aromatic ring in residue 106 facilitates binding of the response regulator, here CheY, to its autokinase protein during the first step of the phosphorylation process. Its external position also hinders access of the flagellar motor to the binding surface of CheY (63). Consistently, inactive mutants of CheY have the aromatic ring in 106 exclusively solvent exposed.

Fig. 3 shows that the transition of the Tyr-106 side chain from a local configuration buried in a hydrophobic pocket to

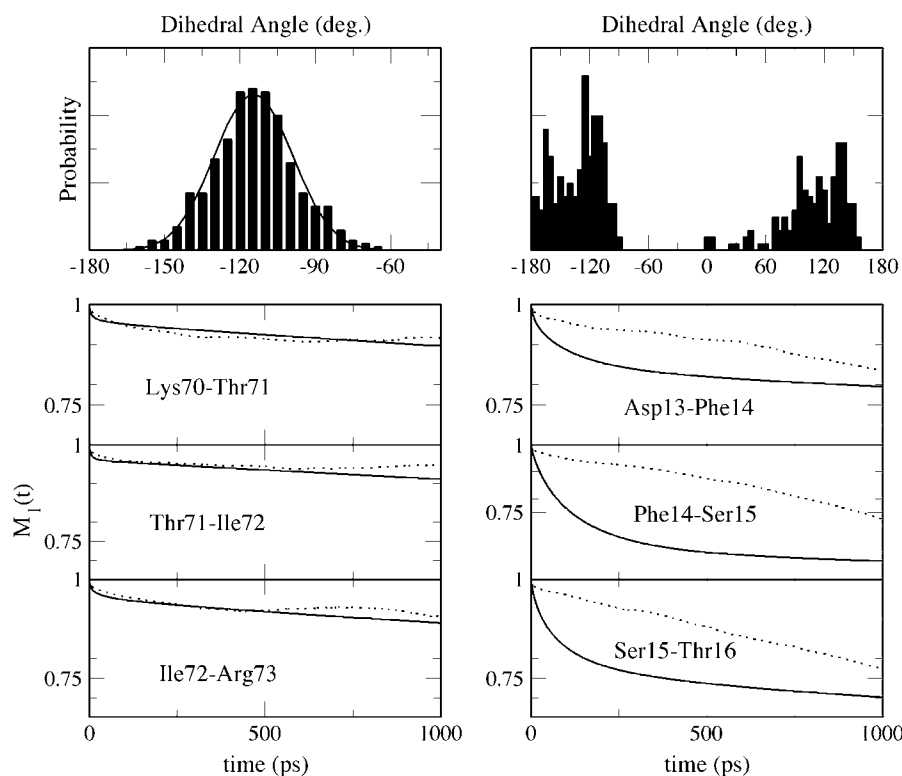


FIGURE 3 Dihedral angle distribution from simulations (*top*) and bond time correlation functions for each bond participating to the dihedral angle (*bottom*). In the bottom plots theoretical predictions (*solid line*) are compared with simulations (*dashed line*).

one exposed to solvent, is coupled to a change in the local energy landscape for the sequence Asp-13, Phe-14, Ser-15, and Thr-16, and the appearance of two stable energetic states. It is possible that the region defined by residues 13–16 in CheY functions as an “on-off switch” during CheY phosphorylation, since it is characterized by two possible states when the ring in Tyr-106 is exposed to solvent, and by only one state when the ring is buried. In support of this hypothesis are the facts that residues 13–16 appear to play some role in the activation of CheY. For example: a), residue Asp-13 is part of the active site of CheY; b), specific amino acids in this region appear to be conserved in the large family of signal transduction proteins; c), mutations that lead to active CheY usually involve residues 13 and 106. In conclusion, this simple procedure of comparing our theoretical predictions for bond decorrelation times with simulation data, allows for the identification of regions along the primary protein sequence where relatively slow conformational transition take place. Such slow regions could potentially play a key role in the protein’s biological function.

Since the theory predicts faster dynamics for bonds 13–15, and bond 117, than should be expected, theoretical predictions of NMR relaxation for these bonds will not be reported. The results presented in Figs. 2 and 3 show that for most of the bonds, the agreement between theory and simulations is good. We emphasize that no adjustable parameters are included in the analytical theory, which characterizes the decays of  $\sim 250$  time correlation functions.

### Comparison of theoretical predictions against NMR relaxation Experiments

From the eigenvalues and eigenvectors obtained from the solution of the equation-of-motion, we calculate the spectral density. The spectral density determines the relaxation times measured in NMR, i.e., the  $^1\text{H}$ - $^{15}\text{N}$  nuclear Overhauser effect (NOE), the spin-lattice relaxation time ( $T_1$ ), and the spin-spin relaxation time ( $T_2$ ), according to

$$\begin{aligned} \frac{1}{T_1} &= \frac{d^2}{4} [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \\ \frac{1}{T_2} &= \frac{d^2}{8} [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) \\ &\quad + 6J(\omega_H + \omega_N)] + \frac{c^2}{6} [3J(\omega_N) + 4J(0)] \\ \text{NOE} &= 1 + \left( \frac{d^2}{4} \right) \left( \frac{\gamma_H}{\gamma_N} \right) [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] T_1, \end{aligned} \quad (3.1)$$

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

bond length. For each effective bond, we compare the theory with these three independent NMR measurements. In the theory, the effective bond is defined as the vector connecting two adjacent  $\alpha$ -carbons along the protein primary sequence, whereas experimental data concern the relaxation of the bond connecting nitrogen and hydrogen atoms. Although the segments considered are different, the relatively good agreement between the analytical theory and experimental NMR data implies that the most relevant contribution to the dynamics of N-H bond relaxation is given by the backbone dynamics, as represented by the reorientation of the coarse-grained statistical segment connecting two alpha carbons.

The theory predicts NMR relaxations for CheYa and CheYb separately because the interconversion between the two rotameric states of residue Tyr-106 exceeds our longest simulation time. We expect both forms to be present in solution, since the transition between the two rotamers of Tyr-106 has been observed experimentally, and in simulations (64), although their relative concentration is not known. The reduced affinity of CheY to bind the flagellar motor could suggest that only a low percentage of CheYb is present in solution. However, since the x-ray structure of the protein (which is input to our simulations) shows 50% occupancy of the two rotamers (36,65), we adopt this concentration for the two species. Because the differences in the theoretically predicted bond relaxation times between CheYa and CheYb are small, and the experimental data are affected by experimental error, it is not possible for us to accurately assess by comparison the most probable concentration of the two allosteric forms.

In Fig. 4, we show a comparison between theoretical predictions and experiments for  $T_1$ ,  $T_2$ , and NOE of *E. coli* CheY, for the 50:50 concentration ratio of CheYa and CheYb. Predictions for bonds 13–16 and 117 are not reported in the plots, as discussed earlier on. Bonds 13–15 show a peak with the maximum corresponding to residue 14, where  $T_1 = 0.94$  s,  $T_2 = 0.19$  s, and a reverse peak with  $\text{NOE} = 0.01$ . Moreover, residue 117 shows a peak with  $T_1 = 0.74$  s,  $T_2 = 0.14$  s, and a reverse peak with  $\text{NOE} = 0.45$ . These peaks are unrealistic, since the theory incorrectly predicts faster dynamics than should be physically expected for these bonds. Experimental data for the NMR relaxation of bonds 13–15 are not available.

In the comparison between the experimental data and our theoretical predictions, we interpolate each data set with a spline fit that serves as a guide to the eye. Each interpolation line contains  $\sim 50\%$  of the available points in the data set. As an estimate of the agreement between theory and experiments, we calculate the mean square predictive error

$$\sigma = \frac{1}{N} \sum_{i=1}^N \frac{(X_i^{\text{exp}} - X_i^{\text{theo}})^2}{(X_i^{\text{theo}})^2}, \quad (3.2)$$

which gives  $1.0E - 2$ ,  $2.5E - 2$ , and  $9.2E - 3$  values for  $T_1$ ,  $T_2$ , and NOE, respectively. The correlation coefficients between theoretical and experimental data are 0.56, 0.36, and



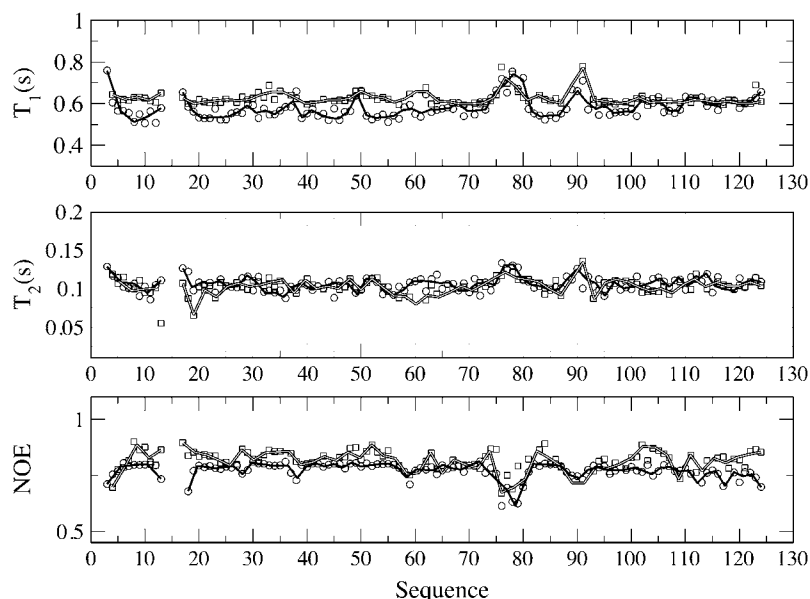


FIGURE 4 Comparison between theoretical predictions (circles) and experiments (squares) of spin-lattice relaxation time ( $T_1$ ), spin-spin relaxation time ( $T_2$ ), and NOE of *E. coli* signal regulator protein CheY, for a 50% mixture of CheYa and CheYb rotamers. Splines interpolate theory (black lines) and experiments (white lines).

0.52 for  $T_1$ ,  $T_2$ , and NOE, respectively. We thus observe slightly better agreement between theory and experiments for  $T_1$  and NOE data, in comparison to that observed for  $T_2$  data. Unfortunately, a full statistical analysis of the quality of the agreement between theory and experiment cannot be performed, since no other CheY relaxation experimental data sets are currently available for comparison with theory. Nevertheless, the relatively good agreement between theory and three independent experimental data sets is encouraging, especially in light of the fact that over  $\sim 350$  data points are described using a first principles theory.

For purposes of visual clarity, we have omitted from Fig. 4 the “trivial” relaxation of the bonds at the two ends of the protein. For those segments, both theory and experiments show enhanced flexibility and fast dynamics due to the lack of connectivity, so that NMR relaxation times largely exceed the range depicted in Fig. 4.

For all of the residues, we find that the theory reproduces well the experimentally observed trend of fast and slow relaxation. Notice that the slow relaxation processes mentioned here do not include slow exchange processes, which cannot be predicted by our theory in its present stage of development. The baseline in the data represents the overall protein rotation, which relates to the protein long-time dynamics and depends on the model for hydrodynamic interaction. Peaks in the spin-lattice relaxation indicate local flexibility and fast relaxation. Both theory and experiments show an enhanced flexibility in regions of the protein corresponding to the  $\alpha_2$ - $\beta_3$  loop (47–52), the  $\beta_4$ - $\alpha_4$  loop (87–92), and the turn including bonds 76–80. In conclusion, the quality of agreement obtained in this study shows that simulations, theory, and experiments are largely compatible in providing consistent information on the physics of the system.

### Theoretical predictions of NMR order parameters and comparison to experiments

Through the direct modeling of NMR relaxation data using the Lipari-Szabo theory, a commonly extracted quantity is the order parameter  $S_i^2 = P_i^2(\Delta t)$ , with  $0 < \Delta t < \tau_R$ , and  $\tau_R$  the protein orientational correlation time, i.e., the time of global molecular rotation. For times short in comparison to the relaxation of the orientational time correlation function,  $P_i^2(t)$ , the bond reorientation can appear restricted to fluctuations around a specific angle. This angle could be similar to that in the protein’s native conformation, suggesting that the bond is part of a rigid local structure and  $S_i^2 = 1$ . Alternatively, the angle could strongly differ from that of the native structure if the bond is flexible, i.e.,  $S_i^2 \leq 0.7$ .

In Fig. 5 (*top panel*), we directly compare our theoretically predicted order parameters with the corresponding values extracted from experimental NMR data using the Lipari-Szabo theory. Several bonds in CheY cannot be modeled and “experimental” values are available only for 90 of the 129 residues. We chose the sampling time interval for the theoretical data  $\Delta t = 13.59$  by optimizing the agreement between the orientational correlation function for the first bond along the protein sequence and its corresponding experimental value. Overall, the agreement between our theory and the experimental data is good for the entire primary sequence.

Both theory and experiments show several regions with significant loss of orientation, including the two ends of the protein, the  $\alpha_2$ - $\beta_3$  loop (47–52), the  $\beta_4$ - $\alpha_4$  loop (87–92), and the turn (76–80), consistent with the relaxation data discussed previously. An unphysical enhanced mobility appears for residues 13–16, where activated barrier crossing cannot be accounted for by the theory, as discussed previously. In general, order parameters extracted from NMR data have

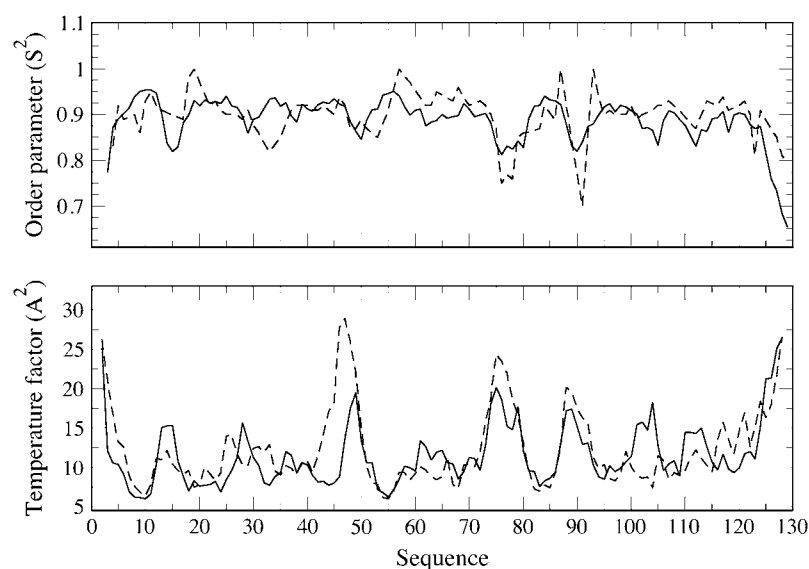


FIGURE 5 Comparison between theoretical predictions (solid line) and experiments (dashed line) of NMR order parameter (top panel) and x-ray temperature factors (bottom panel) of *E. coli* signal regulator protein CheY, for a 50% mixture of CheY $\alpha$  and CheY $\beta$  rotamers.

sharp peaks and fast transitions from completely immobile, to strongly mobile residues along the primary sequence. Our theory appears to exhibit somewhat smoother behavior. In conclusion, the agreement between theory and experiments in identifying regions of high and low mobility is good, with a correlation coefficient  $cc = 0.52$ . Choosing slightly different values of the time interval, from  $\Delta t \approx 5ps$  ( $cc = 0.53$ ) to  $\Delta t \approx 11ps$  ( $cc = 0.56$ ), has a moderate effect on the quality of this agreement.

### Theoretical predictions of short-time dynamics: x-ray temperature factor and comparison to experiments

One of the advantages of our general first-principles approach is that it is independent of adjustable parameters. In contrast, phenomenological models optimize their agreement with a specific set of experimental data by adjusting multiple parameters. Such optimizations seldom achieve equally good agreement with data taken from other experiments that probe similar fundamental quantities.

In this section, we present a comparison between our theoretical predictions of temperature factors and their corresponding experimental values obtained from high resolution x-ray crystallography of *E. coli* CheY by Volz and Matsumura (36). Theoretical values of the temperature factors are function of the mean-square fluctuation about the equilibrium position of individual residues, and are expressed, using eigenvalues and eigenvectors of our equation of motion, following the general relation

$$B_i = \frac{8\pi^2}{3} \langle [\mathbf{R}_i(t) - \mathbf{R}_i(0)]^2 \rangle = \frac{16\pi^2 t^2}{3} \sum_{a=0}^{N-1} \frac{(Q_{i+1,a} - Q_{i,a})^2}{\mu_2} [1 - \exp(-\sigma \lambda_a t)]. \quad (3.3)$$

If the equation of motion correctly represents the dynamics of the protein under consideration, the same equation

should also provide a good representation of short-time fluctuations as measured by temperature factors. Since the time interval of the fluctuations measured by x-ray experiments is not known, we assume in our calculations a time interval of  $\Delta t = 3.5ps$ , which provides the best agreement for the baseline between our predictions and the experimental data. The adopted value of 3.5 ps is consistent with the expected value. Furthermore, we observe that differences in  $\Delta t$  on the order of  $\pm 0.5ps$  do not significantly change the quality of the agreement between theory and experiment.

Fig. 5 (bottom panel) shows that both experiment and theory present similar regions of enhanced short-time fluctuations along the protein primary structure. Most of these regions correspond to flexible loops. The agreement between theory and experiment is good, with a correlation coefficient of 0.60. The observed agreement quality is comparable to other theoretical models used to calculate local fluctuations from x-ray structures, such as the network models pioneered by Bahar et al. and Tirion (46–48). A notable factor that affects the precision of the experimental data is the presence of intermolecular constraints due to crystal packing. Intermolecular constraints, which are absent in physiological conditions and are not included in our approach, suppress the amplitude of local fluctuations, and could be responsible for the lack of mobility observed, for example, in the region defined by residues 100–105.

In summary, Fig. 5 illustrates the utility of a general approach for protein dynamics, which describes in a unified theoretical framework data obtained from independent measurements and different experimental techniques.

## CONCLUSIONS

We propose an equation of motion for protein dynamics that allows for the calculation of NMR relaxation parameters, starting from the protein static structure. The theory is ab

initio, as it is developed using the conventional tools of nonequilibrium statistical mechanics. The equation of motion describing protein dynamics is analytically solved through matrix diagonalization, as opposed to a simulation procedure, to provide eigenvalues and eigenvectors from which the experimentally measurable quantities are expressed. Inputs to the equation of motion are structural correlations and friction coefficients obtained from molecular dynamics trajectories. Since the numerical values of all the physical parameters that enter our formalism (such as local friction, degree of exposure to solvent, bond flexibility, and statistical bond length) are determined from the analysis of computer simulation trajectories, the theory is free of adjustable parameters.

The theory is novel since it explicitly includes in the equation of motion the effects of hydrophobic regions. Hydrodynamic effects are accounted for in a completely general formalism that describes the dynamics of residues in the whole range, from totally exposed to completely screened from solvent. Interresidue interactions are mapped onto an effective pairwise potential, which is obtained from the statistical mechanical description of the intramolecular distribution. Our approach provides a realistic description of the system, since, through the site-specific hydrodynamic interaction and the effective interresidue potential, the specific nature of each amino acid is represented.

The approach is fairly accurate since good agreement with data from different experimental techniques and simulations is obtained. The theory makes predictions of NMR spin relaxation along the protein primary structure, in reasonably good agreement with experiments. We tested our approach through direct comparison to measurements of  $^1\text{H}$ - $^{15}\text{N}$  nuclear Overhauser effect, spin-lattice relaxation time, and spin-spin relaxation time for the signal transduction protein CheY, which controls the swimming behavior of *E. coli* bacteria. Since the timescale characteristic of the physical quantities measured by NMR relaxation exceed the timescale investigated in computer simulations, the analytical approach presented here projects information obtained from computer simulations into the long-time dynamical regime.

The theory is general, because its predictions are not limited to NMR relaxation experiments, but includes any dynamical quantity of interest related to protein equilibrium fluctuations. To highlight the generality of our approach we present comparisons of theoretical predictions with experimental Debye-Waller temperature factors obtained from high-resolution x-ray crystallography, with bond reorientation time correlation functions measured by molecular dynamics computer simulations, and with order parameters extracted from NMR measurements using the Lipari-Szabo model. We find that our theory agrees reasonably well with experiments and simulations for all of the dynamical properties investigated.

Because of its simplicity, our approach provides a straightforward procedure for a first assessment of dynamical properties from the static structure of a protein. In relation to

existing NMR experiments, it provides an estimate of the dynamics of bond relaxation for those regions in the protein that cannot be measured experimentally. Furthermore, our approach could be useful as a systematic method to assess the long-time dynamical properties of a protein with known structure, before NMR relaxation experiments are carried on.

Since our theory is site specific, it allows for calculation of structural and dynamical properties of interest in relation to the biological function of the protein under study. Through a direct comparison with computer simulation data, the theory identifies regions along the protein primary sequence where relevant conformational energy barriers are present. For the protein CheY, we observe a correlation between residues forming such high barrier regions, and the extent to which those residues are conserved in the large family of response regulator proteins. This fact indicates that local energetic barriers, as identified by our method, likely play an important role in the signal transduction mechanism of the CheY protein.

We thank Prof. Andrew H. Marcus for a critical reading of this article.

This material is based upon work supported by the National Science Foundation under grant No. 0509808. We acknowledge the donors of the American Chemical Society Petroleum Research Fund for partial support of this research.

## REFERENCES

1. Cavanagh, J., and R. A. Vanders. 2001. Protein dynamic studies move to a new time slot. *Nat. Struct. Biol.* 8:912–914.
2. Igumenova, T. I., K. K. Frederick, and A. J. Wand. 2006. Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem. Rev.* 106:1672–1699.
3. Palmer III, A. G. 2001. NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* 30:129–155.
4. Wand, A. J. 2001. Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nat. Struct. Biol.* 8:926–931.
5. Brooks III, C. L., M. Karplus, and B. M. Pettitt. 1988. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*. John Wiley & Sons, New York.
6. Yang, D., and L. E. Kay. 1996. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.* 263:369–382.
7. Cooper, A., and D. T. F. Dryden. 1984. Allostery without conformational change. A plausible model. *Eur. Biophys. J.* 11:103–109.
8. Karplus, M., and J. N. Kushick. 1981. Method for estimating the configurational entropy of macromolecules. *Macromol.* 14:325–332.
9. Roberts, M. F., Q. Cui, C. J. Turner, D. A. Case, and A. G. Redfield. 2004. High-resolution field-cycling NMR studies of a DNA octamer as a probe of phosphodiester dynamics and comparison with computer simulations. *Biochemistry*. 43:3637–3650.
10. Freedberg, D. I., R. M. Venable, A. Rossi, T. E. Bull, and R. W. Pastor. 2004. Discriminating the helical forms of peptides by NMR and molecular dynamic simulations. *J. Am. Chem. Soc.* 126:10478–10484.
11. Chatfield, D. C., A. Szabo, and B. R. Brooks. 1998. Molecular dynamics of staphylococcal nuclease: comparison of simulation with  $^{15}\text{N}$  and  $^{13}\text{C}$  NMR relaxation data. *J. Am. Chem. Soc.* 120:5301–5311.

12. Woessner, D. E. 1962. Spin relaxation processes in a two-proton system undergoing anisotropic reorientation. *J. Chem. Phys.* 36:1–4.
13. Wallach, D. 1967. Effect of internal rotation on angular correlation functions. *J. Chem. Phys.* 47:5258–5268.
14. Bremi, T., R. Brüschweiler, and R. R. Ernst. 1997. A protocol for the interpretation of side-chain dynamics based on NMR relaxation: application to phenylalanines in antamanide. *J. Am. Chem. Soc.* 119:4272–4284.
15. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559.
16. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* 104:4559–4570.
17. Massi, F., and A. G. Palmer III. 2003. Temperature dependence of NMR order parameters and protein dynamics. *J. Am. Chem. Soc.* 125:11158–11159.
18. Yang, D. W., Y. K. Mok, J. D. Forman-Kay, N. A. Farrow, and L. E. Kay. 1997. Contributions to protein entropy and heat capacity from bond vector motions measured by NMR spin relaxation. *J. Mol. Biol.* 272:790–804.
19. Berne, B. J., and R. Pecora. 2000. *Dynamic Light Scattering*. Dover Publications, Mineola, NY.
20. Clore, G. M., A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn. 1990. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112:4989–4991.
21. Vugmeister, L., D. P. Raleigh, A. G. Palmer III, and B. E. Vugmeister. 2003. Beyond the decoupling approximation in the model free approach for the interpretation of NMR relaxation of macromolecules in solution. *J. Am. Chem. Soc.* 125:8400–8404.
22. Chang, S.-L., A. Szabo, and N. Tjandra. 2003. Temperature dependence of domain motions of calmodulin probed by NMR relaxation at multiple fields. *J. Am. Chem. Soc.* 125:11379–11384.
23. Baber, J. L., A. Szabo, and N. Tjandra. 2001. Analysis of slow interdomain motion of macromolecules using NMR relaxation data. *J. Am. Chem. Soc.* 123:3953–3959.
24. Proper, J. J., and R. Brüschweiler. 2001. Reorientational eigenmode dynamics: a combined MD/NMR relaxation analysis method for flexible parts in globular proteins. *J. Am. Chem. Soc.* 123:7305–7313.
25. Proper, J. J., and R. Brüschweiler. 2002. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc.* 124:4522–4534.
26. Guenza, M., and K. F. Freed. 1996. Extended rotational isomeric model for describing the long time dynamics of polymers. *J. Chem. Phys.* 105:3823–3837.
27. La Penna, G., A. Perico, and D. Genest. 1999. Modeling the dynamics of the solvated SL 1 domain of HIV-1 genomic RNA. *J. Biomol. Struct. Dyn.* 17:673–685.
28. Giachetti, A., G. La Penna, A. Perico, and L. Banci. 2004. Modeling the backbone dynamics of reduced and oxidized solvated rat mitochondrial cytochrome *b<sub>5</sub>*. *Biophys. J.* 87:498–512.
29. Doi, M., and S. F. Edwards. 1986. *The Theory of Polymer Dynamics*. Oxford University Press, NY.
30. Garcia de la Torre, J., M. L. Huertas, and B. Carrasco. 2000. HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* 147:138–146.
31. Reichl, L. E. 1998. *A Modern Course in Statistical Physics*, 2nd Ed. John Wiley & Sons, NY.
32. Guenza, M. 2002. Cooperative dynamics in unentangled polymer fluids. *Phys. Rev. Lett.* 88:0259011–0259014.
33. Ansari, A., C. M. Jones, E. R. Henry, J. Hofrichter, and W. A. Eaton. 1992. The role of solvent viscosity in the dynamics of protein conformational changes. *Science*. 256:1796–1798.
34. Szurmant, H., and G. W. Ordal. 2004. Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol. Mol. Biol. Rev.* 68:301–319.
35. Hoch, J., and T. Silhavy. 1995. *Two-Component Signal Transduction*. J. Hoch and T. Silhavy, editors. American Society for Microbiology, Washington, DC.
36. Volz, K., and P. Matsumura. 1991. Crystal structure of *Escherichia coli* CheY refined at 1.7 Å resolution. *J. Biol. Chem.* 266:15511–15519.
37. Wangness, R. K., and F. Bloch. 1953. The dynamical theory of nuclear induction. *Phys. Rev.* 89:728–739.
38. Bloch, F. 1956. Dynamical theory of nuclear induction. II. *Phys. Rev.* 102:104–135.
39. Redfield, A. G. 1957. On the theory of relaxation processes. *IBM J. Res. Dev.* 1:19–31.
40. Redfield, A. G. 1965. Theory of nuclear relaxation processes. *Adv. Magn. Reson.* 1:1–32.
41. Perico, A., and M. Guenza. 1985. Viscoelastic relaxation of segment orientation in dilute polymer solutions. *J. Chem. Phys.* 83:3103–3109.
42. Zwanzig, R. 2001. *Nonequilibrium statistical mechanics*. Oxford University Press, NY.
43. Bixon, M., and R. Zwanzig. 1978. Optimized Rouse-Zimm theory for stiff polymers. *J. Chem. Phys.* 68:1896–1902.
44. Perico, A., and M. Guenza. 1986. Viscoelastic relaxation of segment orientation in dilute polymer solutions. II. Stiffness Dependence of Fluorescence Polarization. *J. Chem. Phys.* 84:510–516.
45. Portman, J. J., S. Takada, and P. G. Wolynes. 1998. Variational theory for site resolved protein folding free energy surfaces. *Phys. Rev. Lett.* 81:5237–5240.
46. Bahar, I., A. R. Atligan, M. C. Demirel, and B. Erman. 1998. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* 80:2733–2736.
47. Bahar, I., B. Erman, T. Haliloglu, and R. L. Jernigan. 1997. Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations. *Biochemistry*. 36:13512–13523.
48. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.
49. Bird, R. B., O. Hassager, R. C. Armstrong, and C. F. Curtiss. 1977. *Dynamics of Polymeric Liquids. Kinetic Theory*. John Wiley & Sons, NY.
50. Venable, R. M., and R. W. Pastor. 1988. Frictional models for stochastic simulations of proteins. *Biopolymers*. 27:1001–1014.
51. Grigoriev, I. Z., E. Z. Meilikhov, and A. A. Radzig. 1997. *Handbook of Physical Quantities*. I. Z. Grigoriev and E. Z. Meilikhov, editors. CRC Press, Boca Raton, FL.
52. Zwanzig, B., and M. Bixon. 1970. Hydrodynamic theory of the velocity correlation function. *J. Chem. Phys.* 2:2005–2012.
53. Guenza, M. 2002. Intermolecular effects in the center-of-mass dynamics of unentangled polymer fluids. *Macromol.* 35:2714–2722.
54. Guenza, M. 2003. Cooperative dynamics in semiflexible unentangled polymer fluids. *J. Chem. Phys.* 119:7568–7573.
55. Sambriski, E. J., G. Yatsenko, M. A. Nemirovskaya, and M. G. Guenza. 2007. Bridging length scales in polymer melt relaxation for macromolecules with specific local structures. *J. Phys. Condens. Matter*. 19:205115.1–11.
56. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
57. Pastor, R. W., and M. Karplus. 1988. Parametrization of the friction constant for stochastic simulations of polymers. *J. Phys. Chem.* 92:2636–2641.
58. Sanner, M. F., A. J. Olson, and J.-C. Spehner. 1996. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*. 38:305–320.
59. Miller, S., J. Janin, A. M. Lesk, and C. Chothia. 1987. Interior and surface of monomeric proteins. *J. Mol. Biol.* 196:641–656.

60. Allen, M. P., and D. J. Tildesley. 1987. *Computer Simulations of Liquids*. Oxford University Press, NY.
61. Zwanzig, R., and N. K. Ailawadi. 1969. Statistical error due to finite time averaging in computer experiments. *Phys. Rev.* 182: 280–283.
62. Chandrasekhar, I., G. M. Clore, A. Szabo, A. M. Gronenborn, and B. R. Brooks. 1992. A 500 ps molecular dynamics simulation study of interleukin 1b in water. Correlation with nuclear magnetic resonance spectroscopy and crystallography. *J. Mol. Biol.* 226:239–250.
63. Zhu, X., C. D. Amsler, K. Volz, and P. Matsumura. 1996. Tyrosine 106 of CheY plays an important role in chemotaxis signal transduction in *Escherichia coli*. *J. Bacteriol.* 178:4208–4215.
64. Formanek, M. S., L. Ma, and Q. Cui. 2006. Reconciling the old and new views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins*. 63:846–867.
65. Simonovic, M., and K. Volz. 2001. A distinct meta-active conformation in the 1.1 Å resolution structure of wild-type, apo-CheY. *J. Biol. Chem.* 276:28637–28640.