THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

From NMR Relaxation to Fractional Brownian Dynamics in Proteins: Results from a Virtual Experiment

Paolo Calligari,[†,▽] Vania Calandrini,[‡,§,⊥,▽] Gerald R. Kneller,[‡,§,#,▽] and Daniel Abergel*[,†]

[†]Département de Chimie, Ecole Normale Supérieure, UMR 7203 CNRS-UPMC-ENS, 24 rue Lhomond, 75005 Paris, France

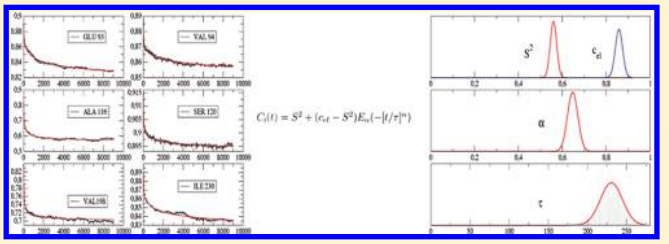
[‡]Centre de Biophysique Moléculaire, CNRS, Rue Charles Sadron, 45071 Orléans, France

[§]Synchrotron Soleil, L'Orme de Merisiers, B.P. 48, 91192 Gif-sur-Yvette, France

[⊥]Centre de Physique Théorique, CNRS, Campus de Luminy, 13288 Marseille, France

[#]Université d'Orléans, Chateau de la Source-Av. du Parc Floral, 45067 Orléans, France

ABSTRACT: In a recent simulation study [*J. Chem. Phys.* **2010**, *133*, 145101], it has been shown that the time correlation functions probed by nuclear magnetic resonance (NMR) relaxation spectroscopy of proteins are well described by a fractional Brownian dynamics model, which accounts for the wide spectrum of relaxation rates characterizing their internal dynamics. Here, we perform numerical experiments to explore the possibility of using this model directly in the analysis of experimental NMR relaxation data. Starting from a molecular dynamics simulation of the 266 residue protein 6PGL in explicit water, we construct virtual ^{15}N R_1, R_2, and NOE relaxation rates at two different magnetic fields, including artificial noise, and test how far the parameters obtained from a fit of the model to the virtual experimental data coincide with those obtained from an analysis of the MD time correlation functions that have been used to construct these data. We show that in most cases, close agreement is found. Acceptance or rejection of parameter values obtained from relaxation rates are discussed on a physical basis, therefore avoiding overfitting.



1. INTRODUCTION

Internal dynamics of proteins is recognized to play an important role in their biological function, and its investigation has been the goal of an impressive number of studies. In this context, nuclear magnetic resonance (NMR) relaxation spectroscopy has proven to be a very powerful tool for the investigation of both overall and internal motions of proteins, and due to its versatility, NMR allows one to probe protein dynamics over a wide range of time scales extending from the ps−ns regime to μs−ms or even longer.[1,2] One of the most important features of modern multiple-pulse, multidimensional NMR techniques is that spectral information can be obtained in a site-specific manner, so that molecular motions can be probed at the atomic level. A wide variety of experiments has been proposed for the investigation of protein backbone and side-chain dynamics, among which amide backbone ^{15}N relaxation experiments on isotopically labeled proteins represents the most widespread techniques.

Random modulations of nuclear spin interactions by molecular motions provide pathways to relaxation.[3,4] Thus, relaxation rates can be used to help characterize internal motions and obtain thermodynamics-related quantities, such as bond conformational entropy.[5,6] However, one faces the difficult task of unambiguously relating relaxation rates to models describing the dynamics on a sound physical basis. These difficulties are reinforced by the fact that a relatively limited number of experiments, which probe spectral density functions at few frequencies only, is available. This therefore imposes the use of interpretative models containing a restricted number of parameters. In this context, the ability of too simple models to faithfully reflect the dynamics may be questionable.

Recently, we proposed an alternative approach to model protein internal dynamics in this NMR perspective. We based our analysis on a fractional Brownian dynamics model in order to take into account the presence of multiple time scale processes through a simplified model depending on a limited number of parameters. The physical relevance of this approach, as well as its capacity to correctly model correlation functions involved in spin relaxation, has been recently demonstrated using molecular dynamics (MD) simulations.[7] In particular, we have shown that angular internal correlation functions were correctly modeled by functions of the Mittag−Leffler (ML) type,[8] which are solutions of the fractional Ornstein−Uhlenbeck (fOU) stochastic diffusion model. Indeed, a relatively simple model of subdiffusion, where a continuous distribution of time scales is involved, was shown to correctly reproduce computed correlation functions based on first principles, as implemented in MD simulations.[7] Therefore, this model seems to represent a valuable tool for interpreting NMR relaxation data in terms of microscopic dynamics.

A natural development of our previous research is to investigate the possibility to derive a methodology aiming at interpreting spin relaxation data. During the past decades, several approaches have been proposed in order to interpret spin relaxation measurements,

Received: June 8, 2011
Revised: July 25, 2011
Published: August 15, 2011

in which the underlying physical picture represents a plausible assumption that is to be confirmed by the data.[9] In contrast, the starting point of our strategy is the analysis of MD simulations, which is the most detailed microscopical description of protein internal motions available to date. Thus, we address the problem of extracting the fOU model parameters from NMR relaxation rates. This clearly represents a demanding task because of the relative scarcity of independent spectroscopic observables. Moreover, model parameters obtained with this approach should not only comply with criteria of statistical quality, but obviously, they should also be physically meaningful. The latter concern is of particular relevance in the context of NMR relaxation studies of protein dynamics, where relaxation rates are only indirectly related to molecular motions through angular correlation functions of the spatial parts of the spin Hamiltonians.[3] The rationale of this link is given by the Redfield theory of relaxation,[3,10] which does not itself provide, but rather requires, a model of correlation functions. Thus, accurate modeling of molecular motions, in particular, internal protein motions, is both critical and difficult.

In this article, we present the analysis of computed amide NH correlation functions from a 91 ns MD simulation of the protein 6-phosphogluconolactonase (6PGL).[11] We used these correlation functions to generate synthetic relaxation data and investigated conditions under which model parameters could be determined from the rates obtained from this "virtual experiment", and a realistic, unambiguous picture of the underlying microscopic dynamics could be gained. The main goal of this work was to devise a strategy to analyze relaxation rates based on fractional Brownian dynamics (fBD), thus performing a proof of principle of this approach as a method to analyze NMR relaxation data.

2. THEORETICAL BACKGROUND

NMR relaxation. The relationship between microscopic motions and spin relaxation rates is provided by Redfield's theory.[10] Amide ^{15}N relaxation results from the fluctuations of the ^{15}N−^{1}H dipole−dipole interactions with the directly attached proton and of the ^{15}N chemical shift anisotropy (CSA) tensor $\boldsymbol{\sigma}$. For a molecule in solution, NMR relaxation rates are determined by the time correlation functions of the form $C(t) = \langle P_2(\boldsymbol{\mu}(t)\cdot\boldsymbol{\mu}(0))\rangle$, where $\boldsymbol{\mu}(t)$ is a unit vector pointing along the ^{15}N ^{1}H bond and $P_2(.)$ is the second-order Legendre polynomial. The CSA tensor $\boldsymbol{\sigma}$ is assumed to be axially symmetric with its axis parallel to the NH bond. Longitudinal and transverse ^{15}N relaxation rates (R_1, R_2), and ^{15}N{^{1}H} heteronuclear Overhauser enhancement ($\eta_{\mathrm{NH}i}$) are expressed as combinations of the spectral density functions $J(\omega)$, the Fourier transforms of the $C(t)$, which are evaluated at the Larmor frequencies 0, ω_{H}, ω_{N}, and $\omega_{\mathrm{H}\pm\mathrm{N}} = \omega_{\mathrm{H}} \pm \omega_{\mathrm{N}}$

$$\begin{aligned}\eta_{\mathrm{NH}} &= 1 + \frac{\gamma_{\mathrm{H}}}{\gamma_{\mathrm{N}}}\frac{d^2}{R_1}(6J(\omega_{\mathrm{H+N}}) - J(\omega_{\mathrm{H-N}}))\\ R_1 &= d^2(3J(\omega_{\mathrm{N}}) + J(\omega_{\mathrm{H-N}}) + 6J(\omega_{\mathrm{H+N}})) + 2c^2J(\omega_{\mathrm{N}})\\ R_2 &= d^2\left(2J(0) + \frac{3}{2}J(\omega_{\mathrm{N}}) + \frac{1}{2}J(\omega_{\mathrm{H-N}}) + 3J(\omega_{\mathrm{H}}) + 3J(\omega_{\mathrm{H+N}})\right)\\ &\quad + c^2\left(\frac{4}{3}J(0) + J(\omega_{\mathrm{N}})\right) \qquad (1)\end{aligned}$$

where $d = \mu_0\hbar\gamma_{\mathrm{H}}\gamma_{\mathrm{N}}/4(10)^{1/2}\pi\langle r_{\mathrm{NH}}^3\rangle$, $c = \gamma_{\mathrm{N}}B_0\Delta\sigma_{\mathrm{N}}/(15)^{1/2}$, and r_{NH} is the NH distance . The parameters γ_{H} and γ_{N} are the gyromagnetic ratios of ^{15}N and ^{1}H atoms, respectively, μ_0 is the vacuum magnetic susceptibility, $\hbar$ is the reduced Planck constant, and $\Delta\sigma_{\mathrm{N}}$ is the ^{15}N chemical shift anisotropy. The presence of additional mobility on the μs−ms time scale appears as a contribution R_{ex} to the observed transverse relaxation rate, $R_2^{\mathrm{app}} = R_2 + R_{\mathrm{ex}}$.[12−14] The spectral density function $J(\omega)$ is given by the Fourier transform of the relevant correlation function

$$J(\omega) = \int_0^\infty C(t)\cos\omega t\,\mathrm{d}t \qquad (2)$$

Provided that statistical decorrelation of internal and overall motions is admissible, and if, in addition, the molecule undergoes isotropic overall diffusive motion, it is possible to factorize $C(t)$ into global and internal time correlation functions $C_{\mathrm{G}}(t)$ and $C_{\mathrm{I}}(t)$[15]

$$C(t) = C_{\mathrm{G}}(t)C_{\mathrm{I}}(t) \qquad (3)$$

Situations where this factorization may not be assumed can be effectively accounted for by a mode-coupling approach.[16] In the following, such a decorrelation will be assumed, and the focus will be on the determination of the internal correlation function in the presence of isotropic overall tumbling with a correlation function $C_{\mathrm{G}}(t) = \mathrm{e}^{-t/\tau_0}$.

Fractional Ornstein−Uhlenbeck Processes and Mittag−Leffler Type Correlation Functions. Analysis of experimental data requires the use of a model correlation function that should faithfully reflect the underlying dynamics. It is possible to assume certain characteristics of the latter, which is the usual approach, and to postulate more or less simple motional models, such as diffusion on a cone, jump motions of the chemical bond,[9] or more elaborate models as in the case of the mode-coupled approach.[16] Alternatively, it may be desirable to take advantage of the atomic description of protein dynamics provided by MD simulations. Thus, one may seek to interpret various dynamical quantities obtained from this microscopic picture by modeling the relevant correlation functions in the simplest possible manner. This provides a way of calculating $J(\omega)$ and relaxation rates. On the basis of several observations, we recently proposed[7,17] to model internal rotational correlation functions $C_{\mathrm{I}}(t)$ by the following expression

$$C_{\mathrm{I}} = S^2 + (c_{\mathrm{el}} - S^2)E_\alpha(-[t/\tau]^\alpha) \qquad (4)$$

where the parameter c_{el}, which is such that $S^2 < c_{\mathrm{el}} \leq 1$, will be discussed below. The ML function

$$E_\alpha(z) = \sum_{k=0}^{\infty}\frac{z^k}{\Gamma(1 + \alpha k)} \qquad (5)$$

is an entire function in the domain of complex numbers,[8] and the parameter α is, in general, complex. The function $E_\alpha(-[t/\tau]^\alpha)$ in eq 4 can be viewed as a generalization of a stretched exponential function, with α and τ being the shape and scale parameters, respectively.

For $\alpha = 1$, the stretched ML function reduces to an exponential, whereas for $0 < \alpha < 1$, it exhibits a power law decay at large times, $E_\alpha(-[t/\tau]^\alpha) \propto (t/\tau)^{-\alpha}$, and an infinitely steep decay at $t = 0$. For $0 < \alpha \leq 1$, the ML function can be expressed as the continuous superposition of exponential relaxation functions $\exp(-\lambda t)$, with the relaxation rate distribution function $p_{\alpha,\tau}(\lambda)$

$$E_\alpha(-[t/\tau]^\alpha) = \int_0^\infty \mathrm{d}\lambda\, p_{\alpha,\tau}(\lambda)\exp(-\lambda t) \qquad (6)$$

dx.doi.org/10.1021/jp205380f |J. Phys. Chem. B 2011, 115, 12370–12379

The spectrum of relaxation rates is positive and has the form[18,19]

$$p_{\alpha,\tau}(\lambda) = \frac{\tau}{\pi}\frac{(\tau\lambda)^{\alpha-1}\sin(\pi\alpha)}{(\tau\lambda)^{2\alpha} + 2(\tau\lambda)^{\alpha}\cos(\pi\alpha) + 1} \quad (7)$$

In eq 7, $p_{\alpha,\tau}(\lambda)$ satisfies the normalization condition $\int_0^\infty d\lambda\, p_{\alpha,\tau}(\lambda) = 1$, and for $\alpha = 1$, it reduces to a Dirac distribution centered at the value τ^{-1}. Moreover, the inverse of the scaling parameter τ gives the median of the distribution $p_{\alpha,\tau}(\lambda)$, so that $\lambda_{1/2} = \tau^{-1}$.[17] The stretched ML function is the solution of a fractional differential equation.[18,20]

The spectral density function associated with eqs 3 and 6 is given by[7]

$$J(\omega) = \frac{S^2\tau_0}{1 + (\omega\tau_0)^2} + (c_{el} - S^2)\frac{1}{\gamma}\frac{(\gamma\tau)^{\alpha}\cos\beta + \cos[\beta(1-\alpha)]}{(\gamma\tau)^{\alpha} + (\gamma\tau)^{-\alpha} + 2\cos\beta\alpha} \quad (8)$$

where $\cos\beta = (\tau_0\gamma)^{-1}$, $\sin\beta = \omega/\gamma$, and $\gamma = (\tau_0^{-2} + \omega^2)^{1/2}$.

As previously noted in ref 7, the initial decay of the correlation function, occurring at time lags typically shorter than ~1 ps, is due to the presence of very fast processes that give rise to rapidly damped oscillations of the correlation function. These phenomena, which cannot be described by a diffusion process, can nevertheless be empirically taken into account by introducing the parameter $c_{el} < 1$ in eq 4.[7] This parameter is just the value taken by the correlation function at the minimum time lag where the theory is assumed valid, what happens at shorter times remaining beyond the scope of the model. Therefore, the use of ML functions represents a way to account for the presence of multiple time scale internal dynamical processes while keeping at the same time the number of model parameters as small as possible.[7]

3. MATERIAL AND METHODS

MD simulations. MD simulations of the protein 6-phosphogluconolactonase (6PGL; PDB accession number 2J0E) were performed using the parallel MD code NAMD[21] with the all-atoms force field AMBER99SB[22] and with periodic boundary conditions. Electrostatic interactions were computed by using the particle mesh Ewald method (PME)[23] with a 12 Å cutoff. Production runs were performed after preliminary structure minimization and equilibration at constant temperature (290 K) and pressure (1 bar) using a Langevin thermostat[24] coupled with a Nosé–Hoover barostat.[25] MD simulations of 6PGL used a total of 9090 SPC/E water molecules and an additional 7 Na^+ ions to ensure an electrically neutral system of 31264 atoms. The lengths of protein chemical bonds involving hydrogen atoms were kept constant through the SHAKE algorithm[26] during the simulation, thus permitting increase of the integration time step to 2 fs. Snapshots of the trajectory were saved every 2 ps of the 91 ns production run.

NMR Rate Calculation from MD Simulations. Order parameters were computed by direct implementation of the following equation

$$S^2 = \frac{4\pi}{5}\sum_{m=-2}^{2}\langle Y_{2m}(\theta,\phi)\rangle\langle Y_{2m}^*(\theta,\phi)\rangle \quad (9)$$

where $Y_{2m}(\theta,\phi)$ are the second-order spherical harmonics that are relevant to the interaction. The ensemble average in eq 9 was computed by both time-averaging over the whole MD trajectory and by performing block averages on 5 ns time windows. The latter were used to calculate deviations from the value computed with the entire trajectory and thereby provided some error estimate associated with the S^2 value of each residue.

Synthetic amide ^{15}N spin relaxation rates were obtained from eq 1 by using two alternative evaluations of the spectral density function $J(\omega)$.

One strategy was to compute relaxation rates from the spectral density function given by eq 8. Model parameters used in this calculation were obtained by fitting the MD correlation functions to the fOU model. In this case, a selection of exponentially sampled points of $C_I(t)$ were fitted to eq 4 through a two-step procedure. First, the cost function $X^2 = [C_I(t) - C_I^{ML}(t)]^2$ was minimized over a limited region of the $\{\alpha,\tau,c_{el}\}$ parameter space by a grid search while keeping S^2 fixed to the value computed from MD by eq 9. Then, the obtained $\{\alpha,\tau,c_{el}\}$ values were used as inputs in the so-called bound-constrained "limited memory Broyden–Fletcher–Goldfarb–Shanno" (L-BFGS-B) optimization algorithm,[27] through which a constrained X^2 parameter minimization was performed.

Alternatively, relaxation rates were computed from the Fourier transform of the correlation functions derived from the MD simulations. In this case, $J(\omega)$ were calculated from the MD trajectory by segmentation of the integral in eq 2[28]

$$J(\omega) = \int_0^{t_m} C_I(t)C_0(t)\cos\omega t\, dt + \int_{t_m}^{\infty} C_I(\infty)C_0(t)\cos\omega t\, dt \quad (10)$$

where ω takes the values $\omega = 0$, ω_N, ω_H, and $\omega_{H\pm N}$, as indicated in eq 1. In eq 10, t_m corresponds to the maximum time at which the correlation function could be calculated with good enough statistics, that is, the first ~10% of the total trajectory length.[29] In principle, the plateau value $C_I(\infty)$ of the internal correlation function was taken to be equal to the mean value of $C_I(t)$ in the time interval $[t_m - 500\ ps; t_m]$. However, $C_I(\infty)$ was replaced in eq 10 by the value of S^2 computed from eq 9 whenever the condition $|C_I(\infty) - S^2| \le 0.005$ was fulfilled.[28,30]

Whenever $C_I(t)$ did not reach a plateau value within t_m, or $C_I(\infty)$ remained significantly different from the value of S^2 ($|C_I(\infty) - S^2| \ge 0.1$), insufficient sampling of internal motions during the 91 ns trajectory[30] was assumed, and $C_I(t)$ was discarded from further analysis.

Relaxation Rate Analysis. Estimating model parameters of eq 4 from NMR relaxation rates necessitates the acquisition of measurements at different magnetic fields to provide a number of data larger than the number of adjustable parameters. Relaxation rates were calculated for magnetic fields B_0 = 21.1 and 14.1 Tesla (^1H resonance frequencies of ν_0 = 900 and 600 MHz, respectively). The values $\Delta\sigma_N$ = 170 ppm and τ_0 = 19.9 ns for the CSA component and the protein overall diffusion correlation time were used.[31] The NH distance r_{NH} = 1.02 Å was assumed constant.[12]

Measurement uncertainties were simulated by adding realizations of noise, drawn from a Gaussian distribution $\mathcal{N}(\mu,\sigma)$, to the computed relaxation rates obtained from either of the procedures described above. The standard deviations σ_{R1j}, σ_{R2j}, and σ_{NOEj} were equal to 3% of the rate value, which is typical of experimental measurement. ML parameters were estimated for each realization of the Gaussian noise $\mathcal{N}(\mu,\sigma)$ by minimizing

the target function

$$\chi^2 = \sum_{j=1,2} \frac{(R_{2j} - R_{2j}^0)^2}{\sigma_{R_{2j}}^2} + \frac{(R_{1j} - R_{1j}^0)^2}{\sigma_{R_{1j}}^2} + \frac{(\mathrm{NOE}_j - \mathrm{NOE}_j^0)^2}{\sigma_{\mathrm{NOE}_j}^2} \quad (11)$$

where R_{1j}^0, R_{1j}^0, and NOE_j^0, on the one hand, and R_{1j}, R_{1j}, NOE_j on the other hand, denote the computed and theoretical relaxation rates at magnetic field B_{0j}. Minimization was performed using either a Levenberg−Marquardt algorithm[32] associated with a parameter grid search or the differential evolution method,[33] both implemented in the Scilab software.[34]

4. RESULTS AND DISCUSSION

MD simulations. Order parameters were computed from the MD trajectory, based on eq 9. In order to check for sufficient conformational sampling, and to get an estimate of S^2 accuracy, averages were performed on 5 ns fragments of the total trajectory. The value retained for S^2 was the mean of these partial averages, and their standard deviation was used as an error estimate. Results are presented in Figure 6.

In addition, correlation functions were calculated from the MD trajectories. Schematically, these fall into two categories. One of them includes correlation functions $C_I(t)$ that clearly reach a plateau value $C_I(\infty)$ within approximately 20% of the maximum admissible time lag (~9 ns). These correspond to residues undergoing mainly fast dynamics (with respect to $0.1 \times t_m$, the maximum admissible time lag for accurate computation of the correlation functions[29]). This is the case, for instance, for SER 120 shown in Figure 1, where $C_I(t)$ decays monotonously to its plateau value. The plateau reached by the correlation function exhibits very little oscillations, which attests to good statistical averaging of the correlation function. This is in agreement with the very small S^2 error bars as determined directly from the trajectory using eq 9.

For a second group of residues, the decay of $C_I(t)$ to a plateau is followed by large fluctuations affecting the tail of the internal correlation function. This strongly suggests the presence of slower dynamical processes affecting the corresponding residues, which may not undergo sufficient averaging during the time over which the correlation function can be calculated with good accuracy (~10% of the MD trajectory; see above[29]). Two such examples (GLY 144 and ALA 238) are shown in Figure 1.

Further justification of these statements, as well as a better insight into the dynamical processes involved, is provided by the observation of the scalar product $\mathbf{r}(t)\cdot\mathbf{r}(0)$ of the instantaneous bond vector with its initial value (see insets in Figure 1). Examples of correlation functions of the first group, where relaxation rates are predominantly in the fast regime (small $\tau = \lambda_{1/2}$), are shown in Figure 1. In the case of SER 120, the quantity $\cos\theta(t) = \mathbf{r}(t)\cdot\mathbf{r}(0)$ exhibits only small fluctuations about some average value. For other residues (e.g., ALA 116), the presence of large-amplitude angular fluctuations is associated with larger fluctuations of the plateau value $C_I(\infty)$ computed from the MD trajectory, as well as larger error bars of the S^2 obtained from eq 9. The decay of the correlation function is slower than that for SER 120, which may explain less good averaging of $C_I(t)$ and S^2.

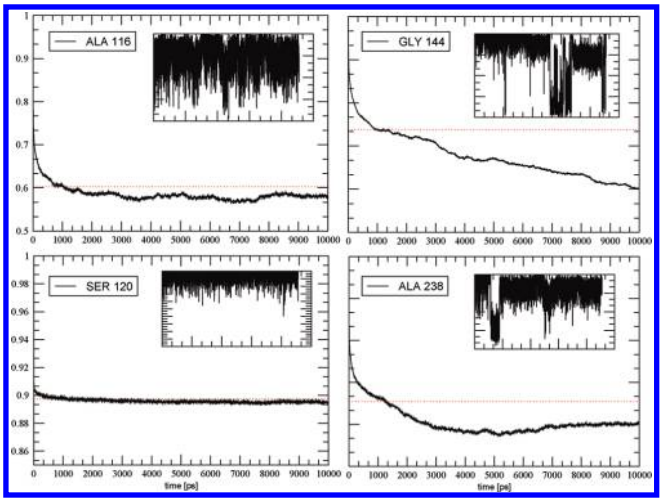


Figure 1. Correlation functions computed from the 91 ns MD trajectory of 6PGL. In the leftmost part of the figure, examples of $C_I(t)$ with good averaging are depicted; on the right, examples of incomplete averaging are shown. In these cases, the time series of $\mathbf{r}(t)\cdot\mathbf{r}(0)$ exhibits rare events that appear as jumps (insets).

Alternatively, for GLY 144 and ALA 238, $\mathbf{r}(t)\cdot\mathbf{r}(0)$ clearly exhibits jumps at the time scale of the MD simulation, which explain the large variance of S^2 parameters measured on fractions of the MD (dashed lines in the figure) and the apparent absence of convergence of $C_I(t)$ toward a plateau value. These rare events would require significantly longer simulations to be satisfactorily averaged, although they superimpose to smaller, and presumably correctly sampled, fluctuations. Longer MD simulations would therefore be necessary to achieve efficient averaging of all stochastic processes leading to decorrelation.

Thus, careful examination of the MD trajectories yielded reliable correlation functions for 173 of the 246 6PGL residues, while for 73 residues, no satisfactory averaging was obtained according to the criteria presented above, and these were therefore discarded from further analysis. The selected residues were then analyzed in the perspective of fractional Brownian dynamics. Indeed, it has been shown in previous work that correlation functions computed from MD simulations could be satisfactorily analyzed by assuming a subdiffusion model based on the fOU process.[7] A preliminary step of the present study was therefore to perform the analysis of the simulated NH bond correlation functions in terms of the ML-based model of eq 4. The agreement between the MD and the fOU model correlation functions is very good, as attested to by the several example correlation functions shown in Figure 2.

These functions exhibit an initial "instantaneous" decay, accounted for in the model by the parameter c_{el} in eq 4. The latter can be given the following justification. The internal correlation function $C_I(t)$ defines a purely diffusive process, which asumption breaks down at very short times, that is, shorter than a picosecond. It is indeed well-known (see, for instance, ref 35) that the subpicosecond evolution of the correlation functions is characterized by an oscillatory behavior that results from local correlated vibrations of torsion angles (backbone ψ, ϕ and side chain χ_1), as well as bond length and bond angle vibrations.[12,36] These phenomena thus generate the underdamped initial subpicosecond behavior, and it is useful to introduce the ad hoc parameter c_{el} that empirically determines the net decay of the correlation function at early times, which a diffusion process fails to describe.

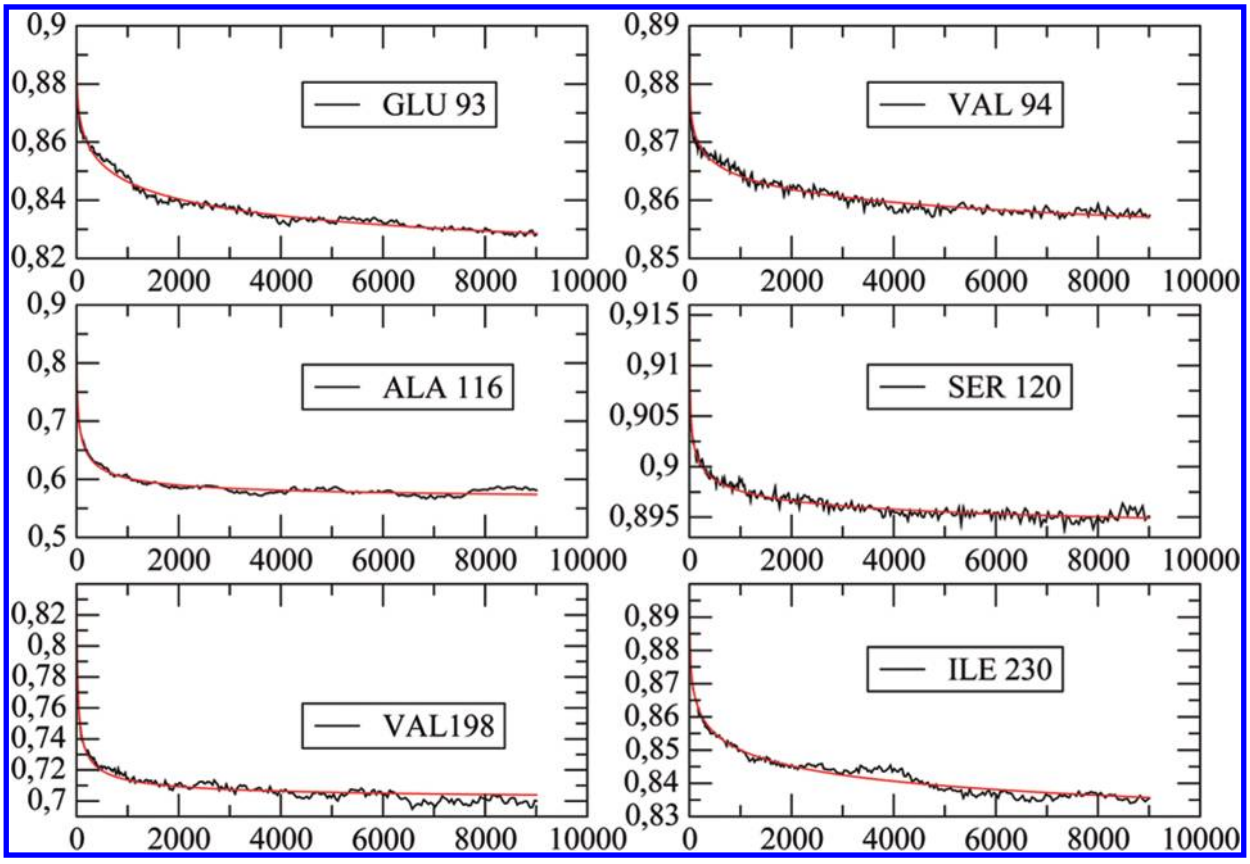


Figure 2. Examples of correlation functions from MD simulation (black). Model correlation functions based on ML functions are depicted in red.

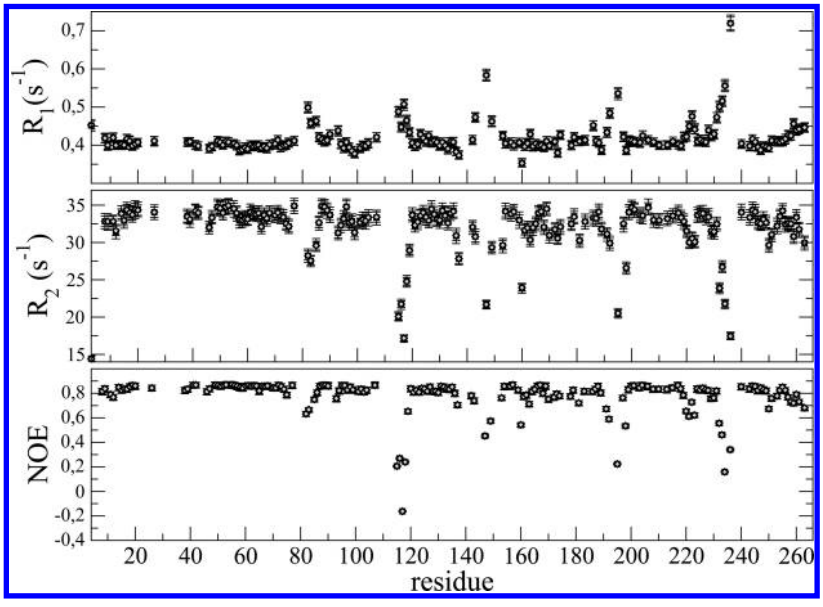


Figure 3. Comparison of simulated (black squares) and back-calculated (gray diamonds) R_1, R_2, and NOE rates for 6PGL. Calculations were performed for residues with good convergence of the MD correlation functions. Simulated rates were obtained through eq 8 by using parameter sets obtained by fitting MD correlation functions to the fOU model.

The capacity of our approach to satisfactorily reproduce internal correlation functions is therefore confirmed by our results, which extends our previous findings[7] to the case of the protein 6PGL.

Recovering ML Parameters from Relaxation Rates. As the main goal of this work is to investigate the possibility to extract the fOU model parameters from relaxation measurements, we computed sets of synthetic R_1, R_2, and NOE rates of the protein

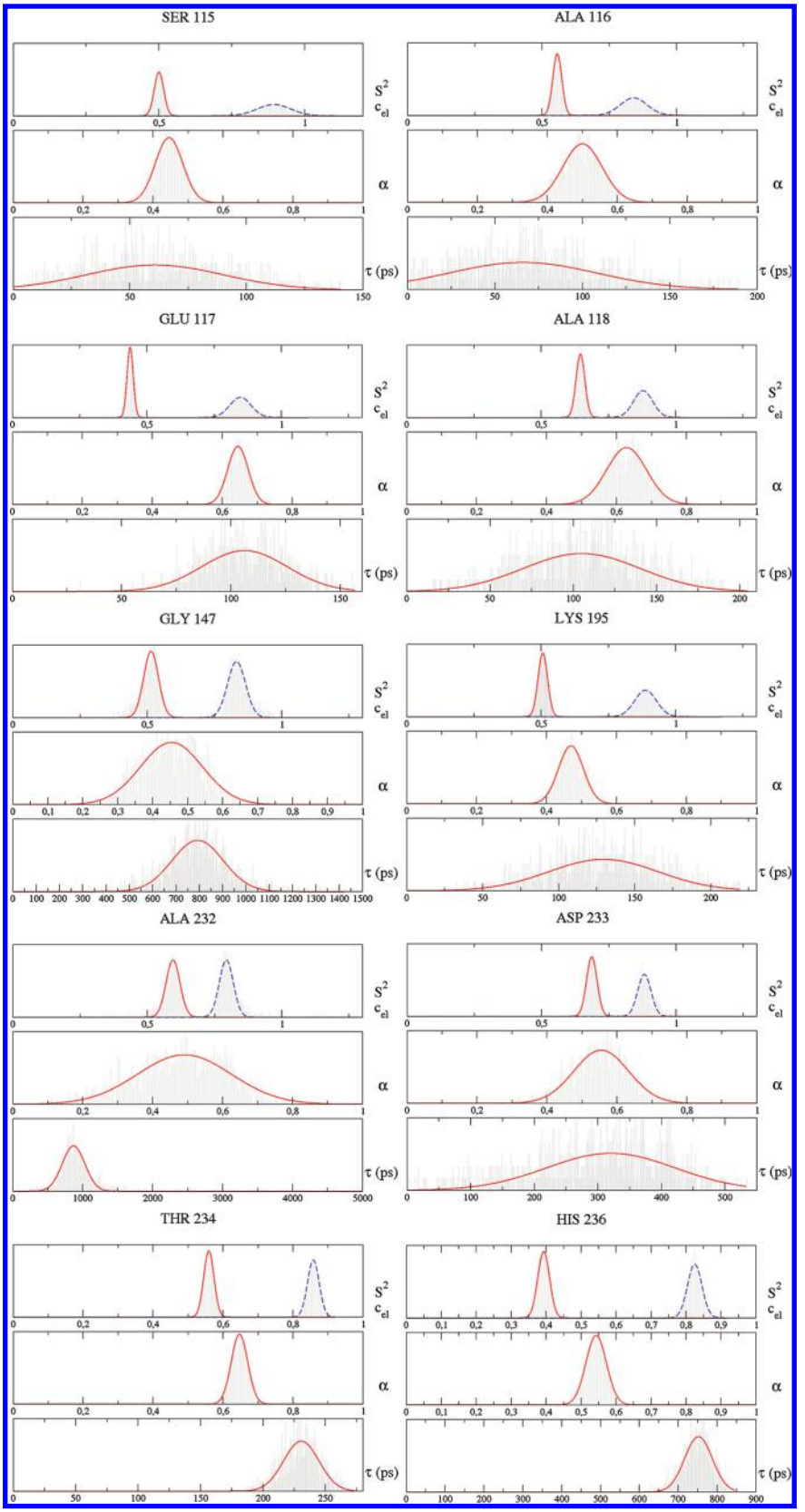


Figure 4. Histograms of ML parameters obtained from synthetic NMR rates for residues indicated in Table 1. Monte Carlo simulations were performed using 1000 realizations of the pseudoexperimental noise. Lines are depicted to guide the eye. For each residue, S^2 and c_{el} are indicated in the top graph by solid and dashed lines, respectively.

dx.doi.org/10.1021/jp205380f |*J. Phys. Chem. B* 2011, 115, 12370–12379

Table 1. Median Values of Monte-Carlo-Fitted ML Parameters Obtained from Relaxation Rates for Residues of Figure 4, Compared to Model Parameters Extracted from the MD Correlation Functions (in parentheses)[a]

residue	S^2 (rate fitting/MD)	c_{el} (rate fitting/MD)	α (rate fitting/MD)	τ (ps); (rate fitting/MD)
SER 115	0.5 [0.5;0.02] (0.50)	0.90 [0.93;0.11] (0.90)	0.45 [0.44;0.06] (0.45)	61 [63;33] (63)
ALA 116	0.56 [0.56;0.017] (0.56)	0.86 [0.91;0.19] (0.86)	0.50 [0.49;0.08] (0.50)	68 [71;44] (68)
GLU 117	0.44 [0.44;0.01] (0.43)	0.85 [0.85;0.04] (0.85)	0.65 [0.64;0.04] (0.64)	106 [106;22] (106)
ALA 118	0.65 [0.65;0.02] (0.65)	0.89 [0.90;0.08] (0.88)	0.63 [0.63;0.08] (0.63)	105 [104;40] (107)
GLU 147	0.51 [0.49;0.08] (0.51)	0.84 [0.86;0.10] (0.84)	0.44 [0.43;0.10] (0.44)	789 [797;325] (793)
LYS 195	0.50 (0.50)	0.89 (0.89)	0.47 (0.47)	129 (217)
ALA 232	0.59 [0.53;0.16] (0.59)	0.80 [0.85;0.15] (0.80)	0.47 [0.44;0.17] (0.47)	907 [21 ns;142 ns] (908)
ASP 233	0.69 [0.69;0.02] (0.69)	0.89 [0.92;0.12] (0.89)	0.55 [0.54;0.11] (0.55)	306 [286;117] (302)
THR 234	0.56 [0.56;0.01] (0.56)	0.86 [0.86;0.02] (0.86)	0.65 [0.65;0.03] (0.65)	230 [230;17] (230)
HIS 236	0.40 [0.39;0.02] (0.39)	0.82 [0.83;0.02] (0.82)	0.54 [0.54;0.04] (0.54)	750 [748;38](750)

[a] Means and standard deviations of the parameters are indicated within brackets.

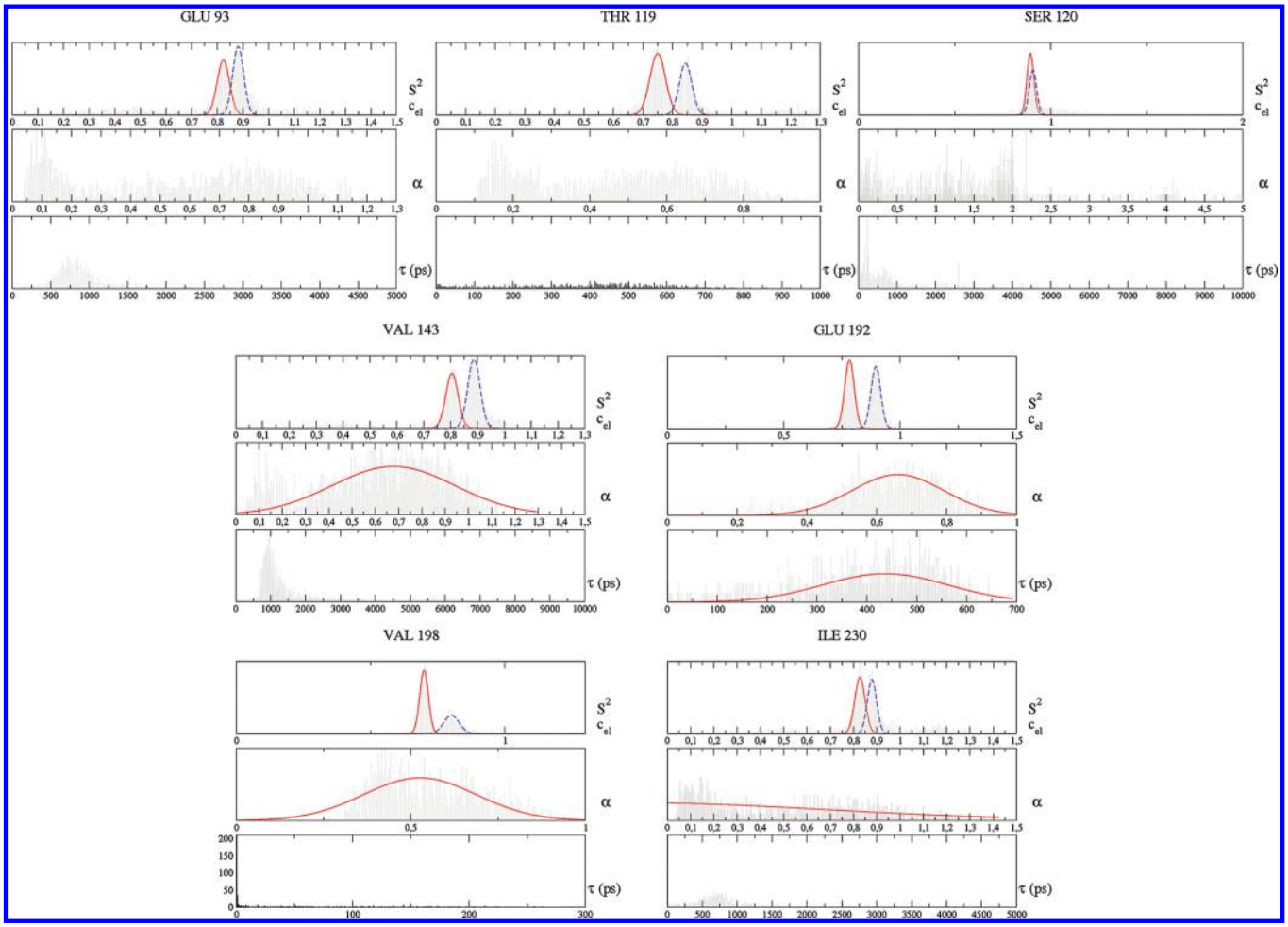


Figure 5. Histograms of ML parameters obtained from synthetic NMR rates for residues indicated in Table 2. Monte Carlo simulations were performed using 1000 realizations of the pseudoexperimental noise. The symbols used are the same as those in Figure 4.

6PGL at two different magnetic fields (B_0 = 14.1 and 21.1 Tesla). These were obtained from the spectral density function in eq 8, where parameters (S^2,c_{el},α,τ) were extracted through least-square fitting of the MD correlation functions, as detailed above. In order to assess the quality of the fits and the possibility to retrieve the original model parameters from these rates, a Monte Carlo simulation with pseudoexperimental noise was performed. A comparison of the simulated and back-calculated rates thereby obtained is depicted in Figure 3. The agreement is clearly satisfactory as rates can hardly be distinguished, and good statistics is attested to by reasonably low χ^2 values (on the order of 2, the number of degrees of freedom).

However, there appears to be no straightforward correspondence between the back-calculated relaxation rates and ML

Table 2. Same As Table *1* for Residues of Figure *5*

residue	S^2 (rate fitting/MD)	c_{el} (rate fitting/MD)	α (rate fitting/MD)	τ (ps); (rate fitting/MD)
GLU 93	0.80 [0.68;0.23] (0.80)	0.90 [0.95;0.14] (0.89)	0.45 [0.47;0.35] (0.38)	968 [1 ms;6 ms] (1389)
THR 119	0.75 [0.73;0.08] (0.75)	0.87 [1.00;0.24] (0.87)	0.45 [0.44;0.24] (0.41)	255 [1.6 μs; 47 μs] (210)
SER 120 (filtered)	0.89 (0.89)	0.98 (0.98)	0.41 (0.22)	2 (0.03)
SER 120	0.89 [0.83;0.20] (0.89)	0.92 [0.97;0.22] (0.98)	1.29 [1.69;2.45] (0.22)	498 ps [0.6 s ; 0.8 s] (0.03)
VAL 143	0.79 [0.65;0.26] (0.80)	0.89 [0.91;0.08] (0.89)	0.58 [0.53;0.30] (0.57)	1367 [217 μs ; 1183 μs] (1424)
GLU 192	0.78 [0.78;0.03] (0.78)	0.90 [0.94;0.13] (0.90)	0.65 [0.62;0.17] (0.64)	400 [366;163] (399)
VAL 198	0.70 [0.70;0.02] (0.70)	0.87 [1.18;0.53] (0.88)	0.53 [0.55;0.16] (0.51)	29 [66;81] (24)
ILE 230	0.82 [0.73;0.19] (0.83)	0.90 [0.99;0.18] (0.90)	0.34 [0.45;0.36] (0.40)	470 [1 ms;8 ms] (312)

parameters in general, and the assignment of correct values of (S^2,c_{el},α,τ) appears to be somewhat more complex.

ML Parameters Distributed about a Single Optimum. In favorable cases, fitted values are distributed about a single optimum for each parameter, which can therefore be safely extracted from the spin relaxation rates. This corresponds in practice to residues with relatively low order parameters S^2, usually associated with smaller values of the median $\lambda_{1/2} = \tau^{-1}$ of the relaxation rate distribution $p_{\alpha,\tau}(\lambda)$ in eq 7. Such a situation, which is characteristic of residues located mainly in flexible regions of the proteins such as loops, is illustrated in Figure 4 for a number of cases. It is seen that the histograms of fitted values show unimodal distributions, which therefore allow characterization of not only S^2 and c_{el} but α too. In this case, histograms of fitted parameters are rather symmetrical, as attested to by the presence of similar mean and median values. The width of the distribution was typically quite large (see Figure 4), imposing limits on the accuracy of α. However, the important point is that values of α that are significantly lower than unity can be extracted from the relaxation rates, therefore demonstrating that the multiple time scale character of the dynamical process can be retrieved. The corresponding parameter estimates for the residues presented in Figure 4 are summarized in Table 1, where it is clear that MD correlation function-fitted and relaxation rate-fitted parameters, estimated by their medians, are clearly in very good agreement. This therefore suggests that analysis of NMR relaxation along the lines of this work allows one to characterize multiple time scale processes through a single model that depends on a constant number of adjustable parameters.

Multiple Minima of the Target Function and ML Parameter Indeterminacy. At the other extreme, when $c_{el} - S^2$ is small, which occurs when S^2 is close to unity (typically $S^2 \geq 0.85$), and when the median $\lambda_{1/2} = \tau^{-1}$ of the relaxation rate distribution adopts rather higher values, (α,τ) pairs may not be determined acurately, although the values S^2 usually are. Instead, one obtains sets of (α,τ) parameters that are distributed across wide regions of the parameter space, which can likely be ascribed to the presence of multiple minima of the χ^2 function. Moreover, the closer S^2 is to c_{el}, the more pronounced the effect. Examples of residues that fall into this category are shown in Figure 5. It is seen that histograms of the fitted α and τ are not clustered about a mean value, so that these parameters seem to be, to some extent, undetermined.

Note that the arithmetic mean is not the most relevant estimator in such skewed or even multimodal distributions, and it is preferable to use the median as a statistical estimator. Parameter estimates obtained for the residues presented in Figure 5 are summarized in Table 2. Thus, for all of the residues

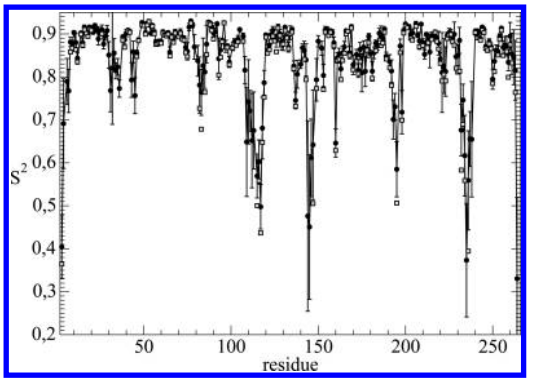


Figure 6. Comparison of the median of the order parameters S^2 extracted from synthetic NMR rates (gray squares) and computed from MD simulations (black circles). Error bars were calculated from the MD trajectory (see text for details).

in the figure, order parameters and c_{el} values are correctly reproduced. Moreover, the medians of α are also in good agreement with the true values, whereas differences are more pronounced for τ. For residues dealt with in Figure 5, fitting yields a significant number of α values above unity (despite good χ^2 values), which obviously has no physical meaning. In practice, we observed that the smaller the values of τ and of the differences $c_{el} - S^2$, the more likely this situation.

However, it is interesting to note that in some cases, when estimators are calculated on a subset of data where unphysical α values are discarded (i.e., only $0 < \alpha \leq 1$ are retained), one recovers more satisfactory results. An illustration of this is provided by residue SER 120 (see Table 2), for which c_{el} and S^2 values are very close ($c_{el} - S^2 = 0.04$) and τ is small. In contrast, this is less critical for residue VAL 198, which has a similar theoretical $\tau = 2.8$ ps, but $c_{el} - S^2 = 0.31$. For this latter residue, the median value of τ is close to the expected value (4 ps).

In sharp contrast, overfitting does not seem to affect the estimation of the order parameter. Indeed, the medians of the order parameters S^2 are, in general, in very good agreement with the exact model values in most cases, even when fitted (c_{el},α,τ) parameters turn out not to be physically relevant, although statistically satisfactory. In Figure 6, the median of the order parameters S^2 extracted from the relaxation rates calculated through eq 8 are depicted. These are essentially identical to the ones obtained from the MD correlation functions (correlation coefficient $r = 0.99$).

Various Limiting Cases of J(ω) that Explain ML Parameter Indeterminacy. The occurrence of multiple minima of the χ^2

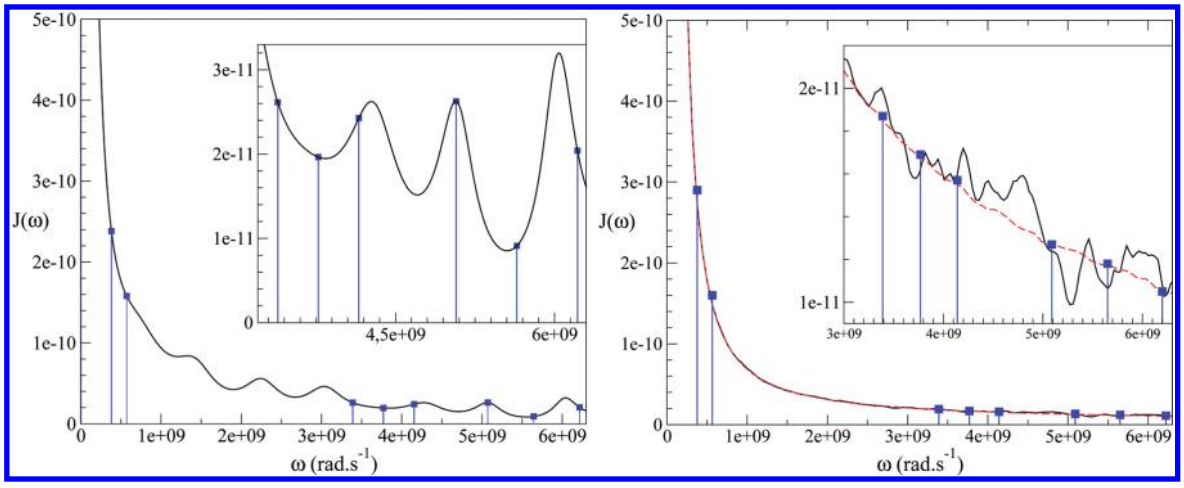


Figure 7. (Left) Spectral density function computed from MD simulation using eq 10 for residue GLU 117. (Right) Theoretical spectral density function where Gaussian noise was added to the *exact* correlation function $C_I(t)$ given by eq 4. The noise standard deviation σ was equal to 1 (solid line) and 0.1% (dashed line) of the function value. High-frequency components of the spectral densities are strongly perturbed at $J(\omega_H)$ and $J(\omega_H \pm \omega_N)$. $\omega_H/2\pi$ = 600 and 900 MHz (squares).

function, hence of possibly physically unsound parameters, can be better understood by considering various limiting expressions of the spectral density function of eq 8. For instance, when $\tau \rightarrow 0$ and for $\alpha \neq 0$, the spectral density function (eq 8) becomes

$$J(\omega) \approx \frac{S^2\tau_0}{1 + (\omega\tau_0)^2} \tag{12}$$

whereas in the limit $\alpha \rightarrow 0$, one has

$$J(\omega) \approx \frac{S^2\tau_0}{1 + (\omega\tau_0)^2} + \frac{c_{el} - S^2}{2} \frac{\tau_0}{1 + (\omega\tau_0)^2} \tag{13}$$

(this latter case includes the limits $\tau \rightarrow 0$ and $\tau \rightarrow \infty$, provided that α/τ = cnst and $\alpha \cdot \tau$ = cnst, respectively). In qualitative terms, this situation corresponds to a ML function that exhibits an abrupt decay at short times followed by a practically constant value. Note that eq 13 also corresponds to the limiting case $\tau \rightarrow \infty$ and $\alpha \neq 0$.

Obviously, these limiting expressions of $J(\omega)$ depend only on the order parameter S^2 and c_{el}, while α and τ remain undetermined. Therefore, in practice, fitting relaxation data to the fOU model whenever τ is very small leaves α unresolved, and conversely, for $\alpha \rightarrow 0$, τ may be assigned values that are equally compatible with relaxation rates. As a direct consequence, which was already mentioned, is that values of $\alpha > 1$ may be obtained through fitting in some cases. Of course, such values are not compatible with the proposed fOU model as the representation of the ML function by eqs 6 and 7 breaks down and the associated fractional differential equation fails to describe fractional Brownian diffusion.[37]

The numerical features of rate fitting can thus be related to eqs 12 and 13, which therefore provide one with a rationale for interpreting possibly unphysical model parameters. Indeed, this indicates that in the case of very short motional time scales of the NH bond, only an order parameter can be safely extracted from the data.

Interestingly, no model selection is required in our approach, but simple inspection of the data permits determination of the reliability of the model parameters. Overall, this method allows, in favorable cases, detection and characterization of the presence of multiple time scales of protein internal motions, using a model with a constant number of parameters.

We add a final comment about the possibility of extracting ML parameters from the relaxation rates obtained directly from the MD simulation. Equation 10 was also used to calculate spectral density functions for residues that fulfilled convergence criteria of both the MD simulation and the correlation function. However, MD correlation functions generally suffer from imperfections due to various limitations that are inherent to the simulation process, such as the finite length of the MD trajectory and the associated incomplete statistical averaging. As expected, these propagate to the spectral density functions and severly affect spectral densities at high frequencies. In Figure 7a, $J(\omega)$ of residue 117 is depicted. It is clear that spectral densities around ω_H are severely perturbed, although values of $J(0)$ and $J(\omega_N)$ do not seem to be affected. This situation can be easily reproduced by adding noise to the exact $J(\omega)$ given by eq 8 and is illustrated in Figure 7b, where a Gaussian white noise with σ respectively equal to 0.1 and 1% of the $C_I(t)$ value was added to the exact correlation function. It is clear that the resulting perturbations of $J(\omega)$ around ω_H and $\omega_H \pm \omega_N$ are deleterious, resulting in relative errors of the NOE values as high as ~25%, hence seriously altering the outcome of the rate fitting process. Of course, this only demonstrates that this effect is really a consequence of the nonideality of MD simulations, which prevents one from using computed spectral density functions to calculate NMR relaxation rates in a straightforward manner, despite the fact that the fOU model adequately fits MD correlation functions.

5. CONCLUSION

In this paper, we have demonstrated the feasability of a NMR relaxation rate data analysis based on a fOU model of protein internal dynamics. The physical basis of this model is rooted in a microscopic description of protein dynamics as derived from MD simulations, which showed that the stochastic properties of protein dynamics are faithfully accounted for by a fUO process.[7] We have laid the basis for further application of our approach to

the analysis of experimental data, which will be the subject of subsequent work.

AUTHOR INFORMATION

Corresponding Author

*E-mail: daniel.abergel@ens.fr.

Notes

$^{\nabla}$E-mail: paolo.calligari@ens.fr (P.C.); Vania.Calandrini@cpt.univ-mrs.fr (V.C.); gerald.kneller@cnrs-orleans.fr (G.R.K.).

ACKNOWLEDGMENT

For this work, access to the HPC resources of [CCRT/CINES/IDRIS] was granted under the allocation 2010-x2010076423 made by GENCI (Grand Equipement National de Calcul Intensif). P.C. acknowledges the Agence Nationale de la Recherche (ANR Contract ANR-COSI-2010-001-01) for funding.

REFERENCES

(1) Fisher, M. W.; Majumdar, A.; Zuiderweg, E. R. *Prog. NMR Spectrosc.* **1998**, *33*, 207.

(2) Palmer, A. G. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 129.

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

(4) Redfield, A. G. *Adv. Magn. Reson.* **1965**, *1*, 1.

(5) Akke, M.; Brüschweiler, R.; Arthur, I.; Palmer, G. *J. Am. Chem. Soc.* **1993**, *115*, 9832.

(6) Yang, D.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369.

(7) Calandrini, V.; Abergel, D.; Kneller, G. *J. Chem. Phys.* **2010**, *133*, 145101.

(8) A. Erdélyi, Magnus, W.; Oberhettinger, F.; Tricomi, F. *Higher Transcendental Functions*; McGraw Hill: New York, 1955.

(9) Daragan, V. A.; Mayo, K. H. *Progr. NMR Spectrosc.* **1997**, *32*, 63.

(10) Redfield, A. *IBM J. Res. Dev.* **1957**, *1*, 19.

(11) Duclert-Savatier, N.; Poggi, L.; Miclet, E.; Lopes, P.; Ouazzani, J.; Chevalier, N.; Nilges, M.; Delarue, M.; Stoven, V. *J. Mol. Biol.* **2009**, *388*, 1009.

(12) Korzhnev, D.; Billeter, M.; Arseniev, A.; Orekhov, V. *Prog. NMR Spectrosc.* **2001**, *38*, 197.

(13) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Oxford University Press: London, 1987.

(14) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Skelton, N. J. *Protein NMR Spectroscopy*; Academic Press, Inc.: New York, 1996.

(15) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546.

(16) Meirovitch, E.; Shapiro, Y. E.; Polimeno, A.; Freed, J. H. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *56*, 360.

(17) Calandrini, V.; Abergel, D.; Kneller, G. *J. Chem. Phys.* **2008**, *128*, 145102.

(18) Glöckle, W.; Nonnenmacher, T. *Biophys. J.* **1995**, *68*, 46.

(19) Kneller, G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2641.

(20) Kilbas, A.; Srivastava, H.; Trujillo, J. *Theory and applications of fractional differential equations*; North Holland Mathematics Studies; Elsevier: New York, 2006; Vol. 204.

(21) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.

(22) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.

(23) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(24) Izaguirre, J.; Reich, S.; Skeel, R. *J. Chem. Phys.* **1999**, *110*, 9853.

(25) Nosé, S.; Klein, M. *Mol. Phys.* **1983**, *50*, 1055.

(26) Ciccotti, G.; Ferrario, M.; Ryckaert, J. *Mol. Phys.* **1982**, *47*, 1253.

(27) Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. *SIAM J. Sci. Comput.* **1995**, *16*, 1190.

(28) P. E., I.; Jianhan Chen, W.; Brooks, C. L.; Biomol., J. *NMR* **2004**, *29*, 243.

(29) Zwanzig, R.; Ailawadi, N. K. *Phys. Rev.* **1969**, *182*, 280.

(30) Nederveen, A. J.; Bonvin, A. M. J. J. *J. Chem. Theory Comput.* **2005**, *1*, 363.

(31) Calligari, P.; Salgado, G.; Pelupessy, P.; Ouazzani, J.; Lopes, P.; Bodenhausen, G.; Abergel, D.; 2011, in revision.

(32) Press, W. H.; Flannery, B. P.; Teukolsky, S.; Vetterling, W. T. *Numerical recipes*; Cambridge University Press: New York, 1989.

(33) Storn, R.; Price, K. *J. Global Optim.* **1997**, *11*, 341.

(34) *Scilab, a free scientific software package*. INRIA ENPC, www.scilab.org, (1989−2005).

(35) Arthur, I.; Palmer, G.; Chazin, W. J. *J. Am. Chem. Soc.* **1992**, *114*, 9059.

(36) Anklin, C.; Rindlisbacher, M.; Otting, G.; Laukien, F. H. *J. Biomol. NMR* **1995**, *5*, 383.

(37) Gorenflo, R.; Mainardi, F. In *International Workshop on the Recent Advances in Applied Mathematics*; Kuwait University, Department of Mathematics & Computer Science, 1996; pp 193−208.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on October 7, 2011. Due to production error, portions of the text and graphics were not corrected. The revised paper was reposted on October 27, 2011.