# Mechanistic Models of Chemical Exchange Induced Relaxation in Protein NMR

Simon Olsson* and Frank Noé*

Computational Molecular Biology, FB Mathematik und Informatik, Freie Universität Berlin, Berlin 14195, Germany

**S** *Supporting Information*

**ABSTRACT:** Long-lived conformational states and their interconversion rates critically determine protein function and regulation. When these states have distinct chemical shifts, the measurement of relaxation by NMR may provide us with useful information about their structure, kinetics, and thermodynamics at atomic resolution. However, as these experimental data are sensitive to many structural and dynamic effects, their interpretation with phenomenological models is challenging, even if only a few metastable states are involved. Consequently, approximations and simplifications must often be used which increase the risk of missing important microscopic features hidden in the data. Here, we show how molecular dynamics simulations analyzed through Markov state models and the related hidden Markov state models may be used to establish mechanistic models that provide a microscopic interpretation of NMR relaxation data. Using ubiquitin and BPTI as examples, we demonstrate how the approach allows us to dissect experimental data into a number of dynamic processes between metastable states. Such a microscopic view may greatly facilitate the mechanistic interpretation of experimental data and serve as a next-generation method for the validation of molecular mechanics force fields and chemical shift prediction algorithms.
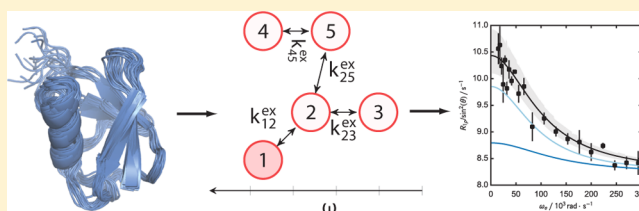
## INTRODUCTION

The basic functions of biological systems, including self-assembly, regulation, and signal transduction, are increasingly well understood in terms of protein structure and dynamics through experimental and computational studies.[1−7] This fact has sparked much innovation in computational methodology aimed at combining experiments with computer simulations in order to generate models of proteins as conformational ensembles.[8−10] However, while all of these methods allow us to map out the conformational landscape of proteins, they are not tailored to resolve the conformational exchange processes on this landscape temporally.[11] This latter fact has stifled the integration of dynamic experimental data. Unbiased molecular dynamics (MD) simulations on the other hand have a long history in aiding the interpretation of these dynamic quantities.[12−16] This approach has recently enjoyed a renewed interest with increasingly powerful computers and methodological innovations[17−22] and led to the characterization of microsecond chemical-exchange dynamics.[23−28]

The chemical shift measured by nuclear magnetic resonance (NMR) spectroscopy is a sensitive structural probe, and in biomolecules, they may yield information about a wide range of geometrical features including dihedral angles and the proximity to ring-currents at the atomic level.[29] The fluctuations of the isotropic chemical shift due to exchange between different conformational states (chemical exchange) give rise to an additive contribution to the measurable relaxation rates as $R_2^{ex}$.[30] With specially tailored NMR experiments, we may isolate $R_2^{ex}$[31−33] or attenuate it in a gradual, controllable manner.[34−37] The latter case, often

referred to as relaxation dispersion (RD) experiments, allows us to probe the spectral density of conformational or chemical exchange process. These experiments are therefore sensitive to the chemical structures, their thermodynamic weight, and their mutual exchange rates. As these experiments typically allow us to probe the dynamics of multiple sites simultaneously, they are potentially exquisitely information dense. However, due to their complex dependence on structural, kinetic, and thermodynamic parameters, their analysis is equally intricate. commonly used analysis framework is based upon the Bloch−McConnell equations[38−40] in which the chemical exchange process is described by an $N$-site lattice jump model.[41,42] The limitation of this approach is that if multiple conformational states undergo mutual chemical exchange, these are typically difficult to deconvolute from the averaged experimental signal. This problem makes it necessary to adopt simpler models and often invoke experiment-specific assumptions in order to make data analysis tractable,[43,44] robust and to minimize the risk of overfitting. Such regularization comes at the cost of the detail of the model.[40] Still, even under these simplifying assumptions, detailed characterization can be difficult unless experimental data are measured at multiple magnetic fields.[23,45]

The dramatic improvement in high-performance and high-throughput MD simulation has generated a need for rigorous, systematic analysis of the large amounts of MD simulation data. One popular framework to analyze large amounts of distributedly generated simulation trajectories are Markov

state models (MSMs).[46−52] MSMs are generated by dividing the conformational space into microstates and estimating a matrix of transition rates (or probabilities) between these states from MD simulation data. If such a model is properly constructed, it can accurately predict long time-scale dynamics and stationary probabilities/conformational free energies from simulations which are individually much shorter than the slowest relaxation time scales of the system. In this way, we can obtain a complete description of the conformational exchange processes observed in our simulated system represented as a rate matrix and structural representations for each microstate; the remaining problem is to map this microscopic model to the experimental observation. The more recently established hidden Markov state models (HMSM) are coarse-grained Markov state models which model the kinetics in terms of the transitions between a few metastable states, and a mapping from each metastable state to the microstates of the conformational ensemble.[53] HMSMs thus provide a coarse-grained view that is more intuitive and easy to interpret.

In the present work, we show how to use MSMs and HMSMs as mechanistic models of the chemical exchange contribution to protein NMR relaxation. We show how this approach allows us to resolve individual processes that contribute to the measured chemical exchange signals. Thus, the present framework avoids the conventional approach to fit phenomenological models to experimental data and thus also avoids the associated problems with overfitting and model unidentifyability. This is achieved as the models are obtained bottom-up from microscopic transitions observed in the MD data and are connected to experimentally observed signals by predicting the spectroscopic observable. While the approach is sensitive to systematic errors in both current models for chemical shift prediction and current molecular mechanics force fields, it is a principled approach to link microscopic simulations to NMR experiments and will directly benefit from improvements in chemical shift prediction models and force fields made in the future. Despite current limitations, we show that our approach can provide useful qualitative and sometimes quantitative predictions. We envisage that this framework improves the mechanistic analysis of complex chemical exchange data and will serve as a diagnostic tool that can facilitate the improvement of molecular mechanics force fields and chemical shift prediction algorithms.

## ■ THEORY

**Chemical Shift Autocorrelation.** The chemical shift is a scalar measure of a spin's local structural (chemical) environment and is sensitive to secondary structure, hydrogen bonding, proximity of ring currents, and other effects. It can be measured for NMR active isotopes, in particular $^1$H, $^{13}$C, and $^{15}$N. A more thorough discussion of the chemical shift and methods for its computation (i.e., 'forward models') can be found in extensive literature on the subject, e.g., ref 29. Here, $\delta_i(\mathbf{x}(t))$ is the chemical shift of spin $i$ given the molecular configuration $\mathbf{x}(t)$ at time $t$. Its autocorrelation function, $g_i(\tau)$ takes the form:[23]

$$g_i(\tau) = \langle (\delta_i(\mathbf{x}(t)) - \overline{\delta_i})(\delta_i(\mathbf{x}(t+\tau)) - \overline{\delta_i}) \rangle \quad (1)$$

where $\overline{\delta_i}$ is the time average of the chemical shift of a spin, and the angular brackets denote ensemble averaging. Thus, if we are able to evaluate or predict $\delta_i(\cdot)$ and have a description of the time evolution of $\mathbf{x}$ (e.g., from a sufficient amount of MD simulation data), we can evaluate $g_i(\tau)$. In practice, $\delta_i(\cdot)$ will be

some approximate forward model: a chemical shift prediction algorithm. In what follows, we will drop the subscript $i$ for simplicity, but assume that we can compute such a function for each spin of interest.

**Exchange Induced Relaxation Rate $R_2^{\text{ex}}$.** The transverse relaxation rate $R_2$ is a sum of multiple contributions arising from relaxation mechanisms due to dipole−dipole interactions, chemical shift anisotropy, chemical (conformational) exchange, and other effects. In this work, we are exclusively concerned with the contribution from conformational changes. In the fast-exchange limit, the exchange induced contribution $R_2^{\text{ex}}$ to the transverse relaxation rate $R_2$ may be expressed in terms of the chemical shift autocorrelation function $g(\tau)$:[30]

$$R_2^{\text{ex}} = (2\pi\nu_0)^2 \int_0^\infty d\tau \; g(\tau) \quad (2)$$

where $\nu_0$ is the Larmor frequency of the observed nuclei in the particular magnetic field. This theory may be readily extended to describe the quenching of exchange induced relaxation in spin-lock experiments as a function of spin-lock frequency, $\nu_1$:[54]

$$R_{1\rho,\text{SL}}^{\text{ex}}(\nu_1) = (2\pi\nu_0)^2 \int_0^\infty d\tau \; g(\tau) \; \cos(2\pi\nu_1\tau) \quad (3)$$

or the interpulse delay, $\tau_{\text{CPMG}} = 1/4\nu_{\text{CPMG}}$, in Carr−Purcell−Meiboom−Gill (CPMG) experiments:[23]

$$R_{1\rho,\text{CPMG}}^{\text{ex}}(\nu_{\text{CPMG}}) = (2\pi\nu_0)^2 \int_0^\infty d\tau \; g(\tau)\text{tri}(2\pi\nu_{\text{CPMG}}\tau) \quad (4)$$

where tri($\cdot$) is a triangle-wave function of linear segments connecting the points $(n\pi, (-1)^n)$ in the $(x, y)$ -plane $\forall \; n \in \mathbb{N}^0$. In both eqs 3 and 4, the experimental modulation of the autocorrelation function allows us to measure the spectral density by adjusting an experimental gauge − $\tau_{\text{CPMG}}$ or $\nu_1$. Similar experiments and expressions exist for multiple quantum relaxation rates.[23,55−58]

**Markov Models and Prediction of Stationary and Dynamic Observables.** A MSM is a discrete approximation of the full phase space Markovian dynamics.[46,51,59] An MSM consists of (1) a discretization of the phase space into $N$ microstates that are often found by clustering simulation data after projecting them onto a reduced-dimensional set of slow collective coordinates,[51,60,61] and (2) after mapping the simulation data to the microstates, a transition probability matrix $\mathbf{T}(\Delta t)$ is estimated (usually by maximum-likelihood or Bayesian inference). The elements $T_{ij}(\Delta t)$ of $\mathbf{T}(\Delta t)$ denote conditional transition probabilities of a system arriving in state $j$ at time $t + \Delta t$ given that it has been in state $i$ at time $t$.[51] For molecular systems in equilibrium, $\mathbf{T}(\Delta t)$ is often estimated so as to fulfill detailed balance.[62] We note that in order to have a self-consistently valid MSM, the state space discretization and the lag-time, $\Delta t$, must be chosen in such a manner that the dynamics described by $\mathbf{T}(\Delta t)$ satisfies the Chapman−Kolmogorov equation, $\mathbf{T}(k\Delta t) = \mathbf{T}^k(\Delta t)$, within statistical uncertainty.[51]

For practical applications, the number of microstates $N$ can be on the order of hundreds to thousands, which makes an intuitive understanding and analysis of the resulting MSM difficult. Different approaches have been suggested to obtain a coarse-grained MSM by lumping microstates into $M \ll N$ macrostates (in a deterministic or fuzzy assignment).[53,63−65] Here we use HMSMs for this purpose.[53] HMSMs describe the molecular kinetics as exchange between metastable distribu-

tions of microstates, and we thus also have structural representations for each metastable state. Besides greatly facilitating analysis, HMSMs have number of attractive properties over a direct coarse-graining of the transition matrix discussed in detail elsewhere.[53] In particular, they can provide significantly better estimates of the molecular kinetics than regular MSMs itself, as they are not limited by Markovianity in the microstate space, i.e., they can often work well despite suboptimal state space discretizations.

In order to compute experimental observables from MSMs, one can proceed following the approach suggested in.[66−68] If $F(\mathbf{x})$ is a forward model of an experimental observable, e.g., a model that predicts the chemical shift of a given atom from the protein configuration $\mathbf{x}$, then the average value of this observable in the set of configurations $S_i$ that comprise the $i$th microstate is computed by averaging the sampled configurations in this state:

$$\mathbf{f}_i = \frac{1}{N_i} \sum_{\mathbf{x}_i \in S_i} F(\mathbf{x}_i) \tag{5}$$

where $N_i$ is the number of samples in set $S_i$.

In all cases considered here, the transition probability matrix $\mathbf{T}(\Delta t)$ is estimated enforcing detailed balance and therefore has a unique stationary distribution $\boldsymbol{\pi}$ with

$$\boldsymbol{\pi}^{\mathrm{T}} = \mathbf{T}(\Delta t)\boldsymbol{\pi}^{\mathrm{T}}$$

This allows us to compute stationary expectation values (ensemble averages) of any experimental observable $O^{\mathrm{calc}}$ provided we are able to compute $\mathbf{f}$ (eq 5), as

$$O^{\mathrm{calc}} = \sum_{i=1}^{N} \pi_i \mathbf{f}_i = \boldsymbol{\pi} \cdot \mathbf{f} \tag{6}$$

where $\cdot$ denotes the scalar product. Such a prediction of an ensemble average can be directly compared with NMR observables, such as $^3J$-couplings and residual dipolar couplings.

However, we can go beyond stationary ensemble averages and use the eigenvalues and eigenvectors of the MSM transition probability matrix $\mathbf{T}(\Delta t)$ with lag-time $\Delta t$ to express the autocorrelation function $g(\tau)$ of a particular experimental observable (i.e., a chemical shift) as a multiexponential sum[41,49,66] (see Supporting Information for a derivation):

$$g(\tau) = \sum_{i=2}^{m} c_i \exp(-\tau/\tau_i^{\mathrm{ex}}) \tag{7}$$

with

$$\tau_i^{\mathrm{ex}} = -\frac{\Delta t}{\log|\lambda_i|}, \qquad c_i = (\mathbf{f} \cdot \mathbf{l}_i)^2 \tag{8}$$

where $\lambda_i$ is the $i$th eigenvalue, associated with the left eigenvector $\mathbf{l}_i$ of $\mathbf{T}(\Delta t)$ respectively, $m \leq N$, and $\tau_i^{\mathrm{ex}}$ is the relaxation time scale of the $i$th eigenprocess, that is sometimes referred to as the *implied time scale*. The direction of conformational exchange on the exchange network associated with a eigenprocess is encoded by the associated eigenvector. We use the standard convention where the eigenvalue-eigenvector pairs are sorted from largest to smallest. In this manner, the slowest relaxing processes have the smallest indices. Furthermore, note that the first eigenvalue is always one for Markov models and corresponds to the equilibrium process.

If we are able to compute chemical shifts of each of our microstates, the MSM directly provides us with a model of the chemical shift autocorrelation function $g(\tau)$ as a sum of exponential decays, with amplitudes $c_i$ and implied time scales $\tau_i^{\mathrm{ex}}$. It is well-known that analytical solutions for eqs 3 and 4 are easily accessible for autocorrelations functions of the form (eq 7):[54]

$$R_{1\rho,\mathrm{SL}}^{\mathrm{ex}}(\nu_1) = (2\pi\nu_0)^2 \left( \sum_{i=2}^{m} c_i \frac{\tau_i^{\mathrm{ex}}}{1 + (\tau_i^{\mathrm{ex}}\nu_1)^2} \right) \tag{9}$$

and[23,69]

$$R_{1\rho,\mathrm{CPMG}}^{\mathrm{ex}}(\nu_{\mathrm{CP}}) = (2\pi\nu_0)^2 \left( \sum_{i=2}^{m} c_i \tau_i^{\mathrm{ex}} \left( 1 - \frac{\tau_i^{\mathrm{ex}}}{\tau_{\mathrm{CP}}} \tanh \frac{\tau_{\mathrm{CP}}}{\tau_i^{\mathrm{ex}}} \right) \right) \tag{10}$$

Similarly for eq 2 we get

$$R_2^{\mathrm{ex}} = (2\pi\nu_0)^2 \sum_{i=2}^{m} c_i \tau_i^{\mathrm{ex}} \tag{11}$$

Consequently, the approach presented here is analogous to what has been approached previously,[23] however, with a key difference: the amplitudes and time-scales are obtained from a Markov model that has been computed from MD simulation data, rather than through fitting of directly computed autocorrelation functions using multiexponential decays. This approach has a much broader scope because the Markov model approach allows complex multiexponential kinetics to be resolved by taking all microscopic transitions into account, and it allows multiple short trajectories started from a nonequilibrium distribution to be used instead of single long "equilibrium" trajectories. Furthermore, our approach connects microscopic conformational exchange observed in molecular simulations to experimental observables, thus providing a detailed mechanistic model of the data. A key to interpretation here is the connection between Markov model eigenvectors and relaxation time scales (eq 8), which allows to assign specific structural rearrangements to experimentally measurable relaxation times.

## ■ RESULTS

**Prediction of Correlation Functions and Comparison of Different Chemical Shift Prediction Algorithms.** Comparison to experiments requires accurate prediction of microscopic observables or forward models. In the case of chemical exchange contributions to NMR, relaxation the observable of interest is the chemical shift. Ab initio calculation of chemical shift is a computationally expensive endeavor which has sparked the development of a wide range of efficient, empirical chemical shift predictors.[70−73] First, we evaluate the consistency of predictions when these algorithms are used to compute chemical shift correlation function $g(\tau)$ from the MSM (eq 7) versus the direct computation from the MD trajectory. Apart from serving as a comparison of these prediction algorithms, this step also acts a first internal consistency check: Does the MSM reproduce the dynamics observed in the MD simulations? To achieve this, we used a previously published ~1 ms MD simulation of BPTI[74] for which a 101 state Bayesian MSM was built (see Methods and ref 62). The chemical shifts were predicted using four methods displaying similar performance in prediction benchmarks.[70−73]

In Figure 1 we show the results for $^{15}N$ backbone amide nitrogens of residues at position 15 and 39 which are adjacent
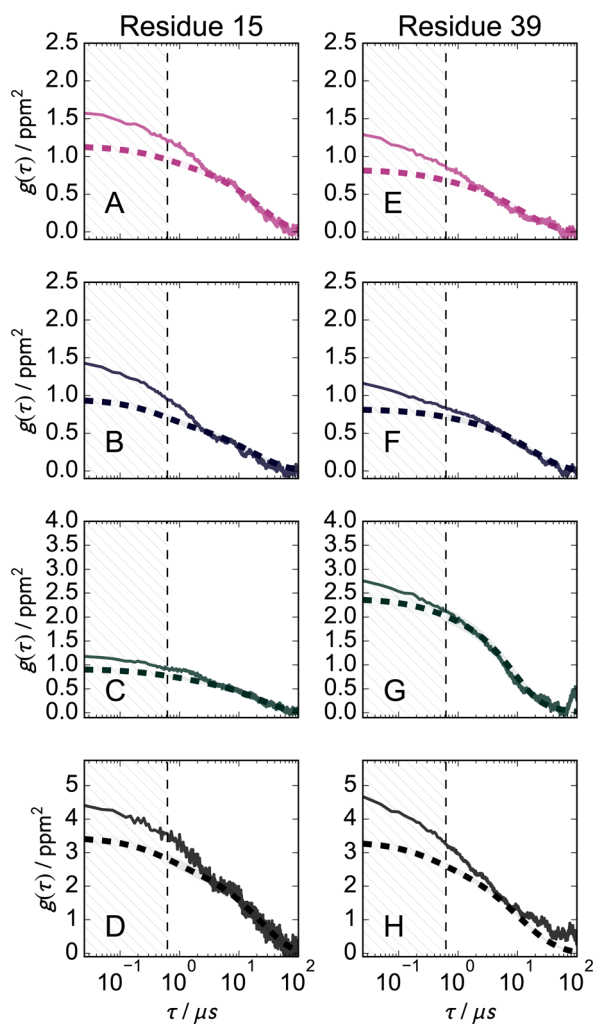


**Figure 1.** Comparison of autocorrelation functions $g(\tau)$ computed using four chemical shift predictors either through direct computation (full line) or using a 101 state Bayesian MSM (dashed line). The lag-time ($\Delta t$) is represented by the hatched area up to a vertical dashed black line. Predictions were performed with ShiftX2 (A, E), Sparta+ (B, F), Camshift (C, G), and PPM (D, H). The shaded areas correspond to a 95% confidence interval calculated from the Bayesian MSM. ShiftX2 was used without the SHIFTY option.

to a pair of cysteine residues that form a disulfide bond previously shown to undergo significant chemical exchange.[74−76] Corresponding plots for the two cysteines are shown in Figure S1.

We see the MSM accurately predicts the directly computed correlation functions for processes slower than the slow nanosecond time-scale in all cases, and we find the contributing processes have exchange time-scales in agreement with previous analyses of the same MD data.[23,74] We emphasize that MSMs predict long time-scale behavior, and their time resolution is around the lag time $\Delta t$. Therefore, we see systematic differences between computed and predicted autocorrelation function for time-scales faster than the MSM lag-time in most cases (Figures 1 and S1). However, the (H)MSM lag-time will in most practical cases be much faster than the temporal resolution of the experiments considered here, and we therefore

do not anticipate this to affect predictions made with this approach.

Note that the different chemical shift prediction algorithms provide qualitatively, but not quantitatively, similar dynamic responses. The PPM algorithm predicts systematically larger amplitudes. This is likely due to the fact that PPM, unlike the other algorithms, explicitly accounts for the ensemble averaged nature of the training data. In turn, this may involve an increased sensitivity of PPM compared to the other methods. Nonetheless, the averaged contributions of different processes are very similar for different prediction algorithms (see Figure S2). This suggests that, although the absolute scale of the chemical shift variation is not correctly predicted, it appears that the relative contributions are fairly consistent, on average. In other words, overall the scale of the computed chemical shift variance will generally not be exact, limiting us to qualitative comparisons with experimental data unless we invoke further assumptions. In cases where multiple experimental observables are available that depend on the same chemical shift variance, we expect to be able to recover the correct absolute scale by estimating a correction factor, $\alpha$ (see Methods).

**Multiple Fast Microsecond Contributions to Exchange Induced Relaxation in BPTI.** To assess the predictive power of the MSM of BPTI established above and to benchmark the different chemical-shift prediction algorithms, we back-computed partial CPMG relaxation dispersion amplitudes and compared them to experimental data (see Figure 2). We find good qualitative agreement for all chemical-shift predictors, with the relative amplitudes being surprisingly close. Again, PPM generally predicts the largest amplitudes. The degree of agreement is similar to that achieved in a previous analysis.[23] The experimental data set which is primarily sensitive to slower dynamics (ms) (Figure 2E−H) is not described as well as the data set that is also sensitive to faster dynamics ($\sim\mu s$−ms) (Figure 2A−D). We anticipate this, as the slowest implied time-scale observed in our MSM is in the range 10−100 $\mu s$ (see Figure S3), outside the sensitivity range of the former experiment. This observation is consistent with previous analyses of this BPTI simulation[23,74,77] and demonstrates that (H)MSMs can provide models limited to fast exchange dynamics, until more extensive molecular simulations become feasible. As discussed in the previous section, we do not expect to have quantitative agreement with this class of experimental data, as the employed prediction algorithms have systematic inaccuracies in the absolute amplitudes for this class of experiments.

There are four processes associated with the predicted experimental data (Figure 2) which all predict exchange close both C14 and C38 for all four prediction algorithms (see Figure S2). This suggests that multiple independent, correlated dynamic modes exist in BPTI, and none of the relaxation is due to localized conformational changes only affecting one of the sites. We characterize these motions structurally by coarse-graining the 101 state MSM into a 4 state HMSM (see Methods and Figure 3). The HMSM reveals a kite-shaped network of three mutually interconnected states of which one of these is connected to a fourth, low-population state. The two most populous states (blue, ~68%, and purple, ~20%) involve a discrete switch in the backbone residues 12−15 and subtle repopulation of the distribution rotameric states in C14 and C38. The third most populous state (yellow, ~8%) has significant flexibility in the residues surrounding C14. Finally, the least populated state (orange, ~3%) involves pivoting of
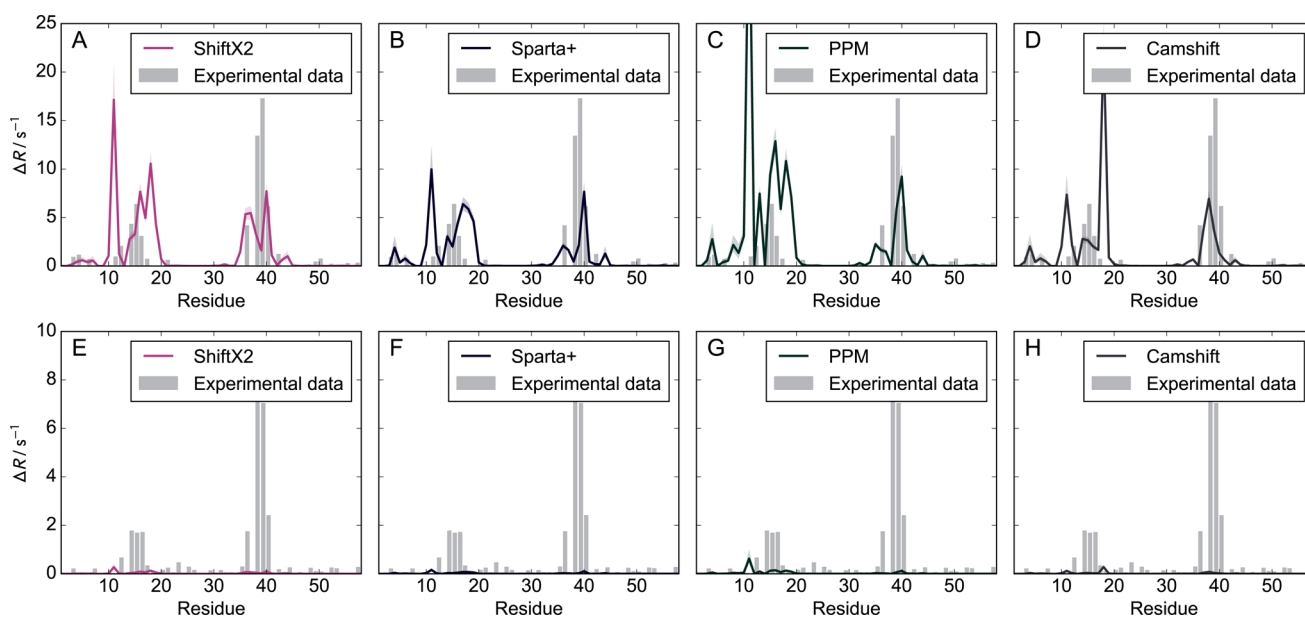
**Figure 2.** Comparison of predictions from a Bayesian MSM (full lines, shaded area: 95% CI) of (A–D) partial CPMG relaxation dispersion amplitudes: $\Delta R_2^{ex} = R_{1\rho,CPMG}^{ex}(15.4\ s^{-1}) - R_{1\rho,CPMG}^{ex}(\infty) \approx R_2^{ex}$ [78] and (E–H) $\Delta R_2^{ex} = R_{1\rho,CPMG}^{ex}(10\ ms^{-1}) - R_{1\rho,CPMG}^{ex}(1\ ms^{-1})$.[79] Predicted curves are based on contributions from the four slowest processes in the MSM.
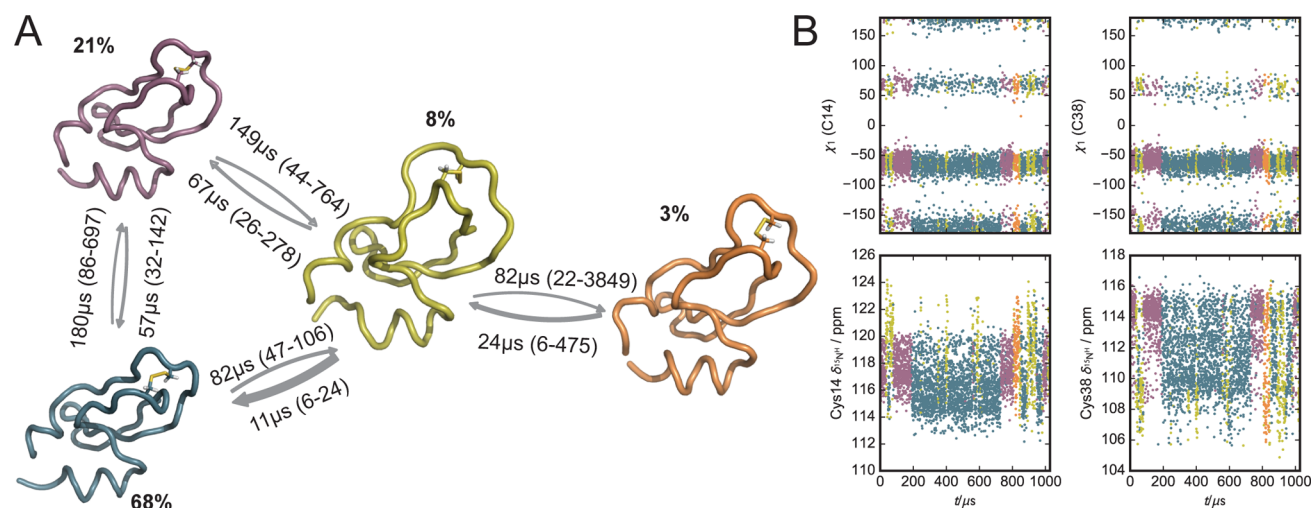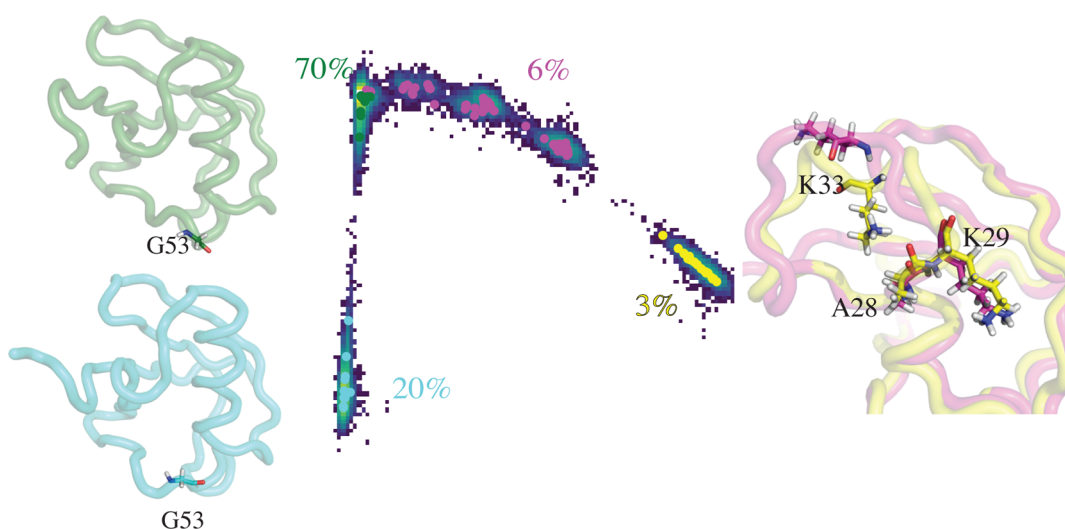


**Figure 3.** Illustration of a four-state HMSM coarse-graining of the 101 state Bayesian MSM of BPTI. (A) a network plot of four metastable configurations and their kinetic connectivity. Transition rates are shown by the arrow and are averages and the 95% confidence interval both in $\mu s$. (B) Color coded time traces of $\chi_1$ angles (upper) and $^{15}N^H$ chemical shifts (lower) in C14 (left) and C38 (right). Color coding was done according to the assignment of the MD frame in the coarse-grained HMSM, colors are as shown in (A). Chemical shifts shown were predicted using ShiftX2 (without SHIFTY option). Molecular structures were rendered using PyMOL (DeLano Scientific LLC).

residues 36−42, considerable flexibility of N-terminal residues 1−15, and again a slight repopulation of $\chi_1$ angles in C14 and C38 (Figure 3B). The slowest process involves repopulation between the purple, yellow, and orange states, and the second slowest process involves repopulation between the orange, yellow, and blue states. The latter of these has the largest contribution (see Figures S2, S3, and 3A). The third slowest process involves exchange of the yellow state with either of the other states. Finally, we observe the four HMSM states capture transitions observed in the MD trajectories of $^{15}N^H$ chemical shifts and $\chi_1$ angles in C14 and C38 (Figures 3B and S5). Overall, our HMSM is very similar to the coarse-grained description presented by Shaw and co-workers in their original study.[74] Although both finer and more coarse kinetic models of

the same MD trajectory have been presented in the literature (e.g., ref 80), we here limit the analysis to a level of detail warranted for by the experimental sensitivity of the observables considered (Figures S2 and S4).

As has been suggested previously,[23,75] it appears that the relaxation mechanism in BPTI may be more complex than the phenomenologically derived three-state models based on $^{15}N$ and $^{13}C$ relaxation data.[75,81] Specifically, instead of directly connecting the relaxation mechanism to repopulations in the rotamer distributions or jumps in-between rotameric states, we here find that it depends on multiple modes of concerted rearrangements in the backbone and side-chains. However, the implied time-scales observed in these simulation do not match experiments exactly, which does not exclude the possibility of

**Figure 4.** Illustration of a four-state HMSM coarse-graining of a 128 state MSM of ubiquitin. Major conformational features of the HMSM illustrated by molecular renders of each of the four states: cyan, green, magenta, and yellow. In the center, a rendering of the free energy landscape projected on the first two time-lagged independent components and cluster-centers color-coded according to the metastable assignment in the HMSM (see Methods). Molecular structures were rendered using PyMOL (DeLano Scientific LLC).

an even more complex mechanism that depends on conformational transitions that have not been sampled in the given simulation.

**Dissection of Fast Microsecond Chemical Exchange in Ubiquitin.** To further assess the potential of Markov models to obtain microscopic and mechanistic descriptions of chemical exchange induced relaxation, we chose ubiquitin as an example. Ubiquitin is a well-characterized protein where the presence of fast microsecond chemical exchange has been reported in multiple studies.[26,58,82−86] We used a recently published 1 ms trajectory at a temperature of 300 K[85] to build a 128-state Bayesian MSM.[62] This model was used in all analyses shown below. For convenience, we limit ourselves to chemical shift predictions using Camshift.[71] The resulting MSM reveals three processes with relaxation times in the range of tens to hundreds of microseconds which have significant amplitudes for $^{15}$N and $^{1}$H backbone resonances (see Figures S2 and S3). We note that average contributions of the different processes to predicted $^{15}$N relaxation are very similar to the other three prediction algorithms considered above (see Figure S2).

The time-scales predicted by the MSM are slower (slowest implied time scale: ∼80 $\mu$s) than those recently reported in experiments at temperatures close to 300 K (308 K, ∼5 $\mu$s), but match time-scales reported at 277 K fairly well (∼50 $\mu$s).[87] Consequently, we used an average of previously reported Arrhenius activation energies[84] to extrapolate the apparent relaxation time scales (277 K) prior to comparing to experiments. This is a rather small intervention considering different experimental and computational studies at varying sample conditions have reported a relatively broad range of microsecond time-scales dynamics to affect the same regions in ubiquitin.[26,58,85,87−89] Consequently, these rates may sensitive to the exact experimental conditions and the presence of different cosolutes.[90] However, we test this intervention thoroughly below with recently reported data sets at multiple temperatures.[87]

To obtain a model that facilitates a mechanistic interpretation of the predicted experimental observable, we first build a coarse-grained four-state HMSM (Figure 4A). This model reveals a linear exchange scheme where each state has distinct

structural characteristics. The overall features of the HMSM are very similar to what was described in the original paper by Piana et al.[85] The two most populous (green, ∼70%, and blue, ∼20%) describe a flip of the loop at 50−54 which repopulates the preference of the H−N vector of G53 between inward and outward pointing configurations (Figure 4A), respectively, as described in previous computational[91] and experimental studies.[87,92,93] Two alternative states (magenta, ∼6%, and yellow, ∼3%) both involve a partial unwinding of the C-terminal end of helix 1 (ca. residues 31−38). The less populated state (yellow) stabilizes an alternative configuration of the loop downstream of the unwound helix 1, through hydrogen bonding of H$\epsilon$s of K33 to exposed carbonyl oxygens in residues 28 and 29, whereas this loop displays increased flexibility in the other partially, unwound state (magenta) (Figure 4A). The yellow state is also characterized by a kink in the most C-terminal $\beta$-strand following I71 whose side-chain is flipped outward and thus solvent exposed.

Interestingly, visualizing the direction of the slowest process in the MSM on this network as a gradient (Figure S7) reveals that it involves exchange between the yellow and green and blue states. This slowest process is the origin of a major false-positive prediction of $^{15}$N$^{H}$−$R_2^{ex}$ values (see Figure S6). Furthermore, it was previously suggested that the population of the yellow state is overemphasized in the simulation trajectory due to insufficient sampling, force field imbalances, or a combination of these.[85] Since exchange to/form this state is the key feature of the slowest process (Figure S7), we decided to continue the analysis below leaving out this process resulting in what we here refer to as a reduced observable model.

To validate the reduced observable model, we compared its predictions to an extensive set of high-power RD $^{1}$H$^{N}$ $R_{1\rho}^{SL}$ and CPMG experiments at multiple temperatures (see Figures 5 and S8−14). Using Arrhenius' equation to extrapolate the time-scales, we find that the reduced observable model agrees well with the majority of these experiments when compared to the full MSM (see Figures S8−14). These comparisons involve estimation of an $R_{2,intrinsic}$ (intersection) and a overall scaling factor ($\alpha$, slope), for each RD profile (see Methods). This is
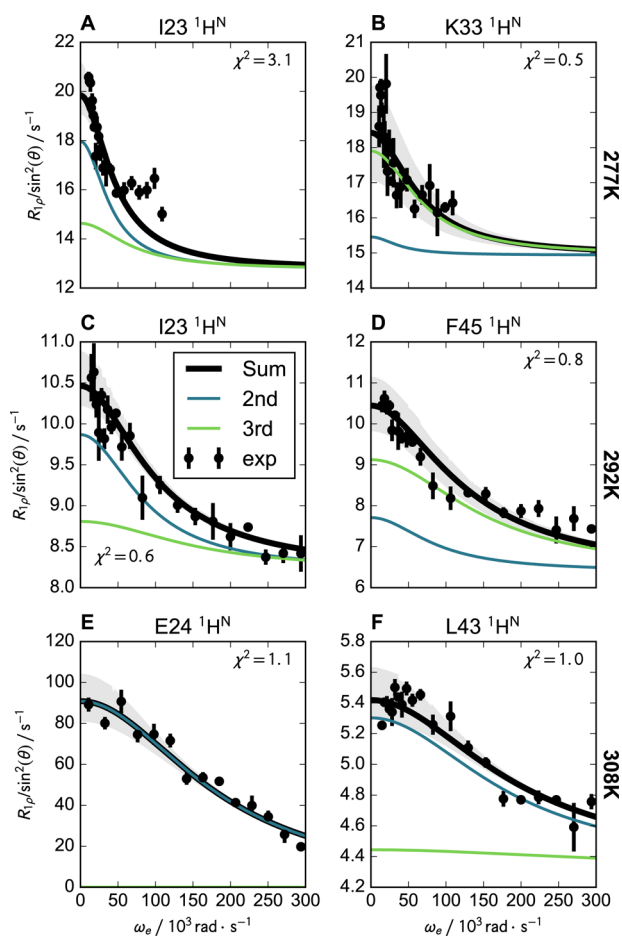
**Figure 5.** Representative set of $^1H^N$ $R_{1\rho}$ relaxation dispersion profiles at 277 K (A, B), 292 K (C, D), and 308 K (E, F) predicted from a reduced observable model of ubiquitin and corresponding experimental data.[87] Rates at temperatures 292 and 308 K were extrapolated from apparent low-temperature kinetics (ca. 277 K) of the MSM using an average of previously reported activation energies of 35 kJ mol$^{-1}$.[84] Bold black line represents the sum of the predicted dispersion profiles of the second (blue) and third (green) slowest processes, scaled by $\alpha$ and with an intersect of $R_{2,intrinsic}$ (estimated as described in Methods). The blue and green lines have the same intersect ($R_{2,intrinsic}$) and are scaled by the same $\alpha$, as was estimated for the sum to ensure comparable scales. Gray shaded area represents the 95% confidence interval of the predicted profile. The $\chi^2$ values shown are computed using eq 14.

necessary, as we here only predict the $R_2^{ex}$ contribution up to a scaling constant, as discussed above, and the experimental data potentially contain contributions from other relaxation processes. Instead, correlating these data to the MSM or either of the three processes independently yields very comparable agreements (see Table S1). This latter comparison consequently provides additional validation independent of the fitting parameters $R_{2,intrinsic}$ and $\alpha$. Collectively these tests suggest that the reduced observable model is a good consensus model. These comparisons allow us to take a closer look at the second and third slowest processes in our MSM and study how they contribute to the respective experimental dispersion profiles.

Next, we evaluate the influence of the second and third slowest processes of the MSM (the reduced observable model) on the predictions of $^1H^N R_2^{ex}$ amplitudes and relate them to structural changes in ubiquitin. We color code residues

according to which process dominates the predicted profiles, e.g., which of the processes predicts a larger value of $R_{1\rho}/\sin^2(\theta)$ for $\omega_e \rightarrow 0$, i.e., independently of fitting parameters (see Figure 6). The two processes have substantial overlap but
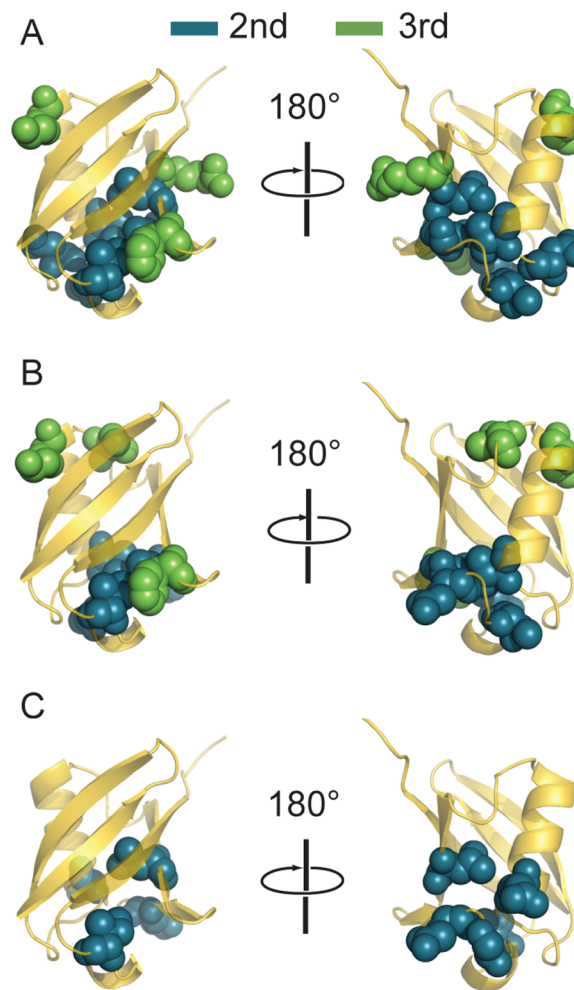


**Figure 6.** Visualization of the largest contributing process to predicted $^1H^N R_2^{ex}$ values at temperatures 277 K (A), 292 K (B), and 308 K (C). Residues with dominant contribution to the experimental observable at the respective relaxation process are highlighted as green or blue spheres. Molecular structures were rendered using PyMOL (DeLano Scientific LLC).

appear to form two lobes: the second slowest process mainly affects a small cluster of residues involving loops 2 (residues 18−22) and 5 (residues 50−55), whereas the third slowest process affects residues 45−48 through the $\beta$-sheet to the C-terminal cap of $\alpha$-helix 1 (residues 30−38). The slower of these two processes is observable by high-power RD experiments across the temperature range previously reported, whereas the faster process is below experimental resolution at the higher temperatures. Previous studies of ubiquitin found correlated motions across the $\beta$-sheet as well in loops 2 (residues 18−22) and 5 (residues 50−55). These motions were suggested to be correlated with ubiquitin binding.[92,94]

Interestingly, Smith et al. report mutagenesis of residues (E24A and G53A) results in a significant attenuation of observed relaxation dispersion amplitudes in most residues.[87] Both of these mutations are situated at positions dominated by

the second slowest process of the reduced observable model. Conversely, the relaxation dispersion of I36 was reported to be almost unaffected in either of these two mutants. I36 is spatially proximate to a cluster of residues whose relaxation dispersion is predicted to be dominated by the third slowest relaxation process. This specific spatial attenuation of the exchange induced relaxation could suggest that there are two independent processes inducing the signals observed, and the two modes of correlated motion previously described are independent.[94] It appears that the reduced observable model presented here correlates well with this, and thus the two processes of the model may be regarded as a possible mechanistic description of the observed data. Since the time-scales of these processes are very close ($4.56 \pm 0.34$ $\mu$s and $3.59 \pm 0.41$ $\mu$s, extrapolated values at 308 K), they are difficult to deconvolute using conventional phenomenological fitting approaches.

## DISCUSSION

Analyzing MD simulations through a (H)MSM yields a network model of the conformational kinetics of the system. A key feature of (H)MSMs is that they can be constructed using multiple short simulations started out-of-equilibrium and predict the long-time scale dynamics. These models can be used to compute stationary and dynamic experimental observables for comparison with experimental data and thereby serve as validation of both force field and the MSM as well as a framework for detailed mechanistic interpretations of data. While stationary experimental data have been used extensively in the literature to validate MSMs, the use of dynamic observables to the same purpose has been rare.[66,95] Here we establish a link between MSMs and the chemical exchange induced relaxation measured in NMR experiments.

Predicting autocorrelation functions is the first step to predicting $R_2^{ex}$ contributions. Doing this, we find that chemical shift autocorrelation functions directly computed from MD simulations compare well with those computed from the corresponding MSMs, which self-consistently validates the long-time scale predictions of the MSM. However, we also observe that although different chemical shift prediction algorithms provide similar qualitative responses, they significantly differ in the absolute amplitudes predicted ($g(0)$). This indicates that they have different levels of sensitivity. Indeed, the PPM algorithm systematically predicts higher amplitudes and has been shown to provide improved predictions for several spin-types, such as side-chain methyl $^1$H and $^{13}$C resonances, particularly prone to fast motional averaging.[70] Since predicted chemical shift amplitudes are in general only correct up to a scaling constant, we found that back-computed partial CPMG relaxation dispersion amplitudes also only agree with experimental data qualitatively. However, for relaxation dispersion profiles, we are able to improve the agreement by a estimating a slope and an intersect. The estimated slopes ($\alpha$) are quite large, which suggests that the empirical chemical shifts will, as a rule, provide a lower bound for the chemical shift variance. A possible remedy to this is quantum mechanics based predictions. These are expected to have a larger overall sensitivity.[96,97] A part from this, ab initio approaches will be attractive in cases where nonstandard amino acids, post-translational modifications, small-molecule ligands, or nucleic acids are present, as most empirical methods currently do not describe such systems. Since (H)MSMs reduce the overall number of prediction calculations needed to compute the

autocorrelation functions (as little as one representative computation per microstate), we anticipate this route to become accessible in the near future.

The usefulness of the MSM framework in the present context depends on their ability to accurately predict experimental observables. For two test systems, BPTI and ubiquitin, we establish models using previously published MD trajectories.[74,85] In both cases we find good qualitative agreement with experimental data, and for ubiquitin, we also achieve good quantitative agreement with relaxation dispersion profiles following the procedure described above. We show how we can facilitate the analysis of the MSM by coarse-graining them to HMSMs and thereby obtain models between a few conformational states that provide a straightforward microscopic interpretation of the experiments. In the case of ubiquitin, we find two independent correlated processes which contribute to the observed relaxation dispersion profile: one which involves a previously described main-chain switch (around G53)[87,92,93] and another process which involves residues abut to this switch through the $\beta$-sheet to residues 31−38. Both of these sites previously been reported to undergo motion,[58,87,89,92−94,98] however, to which extent these motions were correlated has been unclear. We find mutagenesis within a cluster of residues involved in the second slowest process primarily affects experimental observables predicted by the most strongly affected by this process. While residues primarily affected by the third slowest process are also affected in these mutants, the only unaffected residue, I36, is predicted to be in the region affected by the third slowest process. These results illustrate the potential of the framework to deconvolute multiple dynamic processes in proteins and thereby facilitate a mechanistic interpretation of experimental data and potentially aid the selection of residues as targets for mutagenesis. Still, as the methodology is sensitive to current inaccuracies in the chemical shift model and the force field, we stress the importance to experimentally test hypotheses generated using this framework.

The method presented here is applicable in a narrow but important range of experimental parameters. First, we only consider the contribution of conformational changes to the relaxation rate, and we only treat this contribution in the fast-exchange limit, but also only for dynamics slow compared to the MSM lag-time, which is usually much faster than the overall correlation time $\tau_c$ (typically in the few nanosecond range for small proteins). However, a (H)MSM can be combined with models of spin-relaxation to account for other contributions, even outside the fast-exchange regime. For instance, the transition matrix and the chemical shifts of the microstates can be integrated into an analysis based upon the stochastic Liouville equation.[99,100] As it is becoming increasingly feasible to obtain models of millisecond and slower dynamics in proteins, such an approach may become a keystone to rigorous mechanistic data analysis in the near future.

## CONCLUSION

We have presented a new framework for the prediction of chemical exchange induced relaxation. The approach uses MD simulations analyzed in terms of an *N*-state exchange master-equation type model, a (hidden) Markov state model, which we demonstrate how to use to compute a number of chemical exchange sensitive NMR relaxation observables. While the presented approach is subject to current limitations such as sampling and force field inaccuracies, it has a number of distinct

advantages over previous efforts. We are not limited to detecting the metastable states in through the chemical shift, but we can use any metric computable from a MD simulation. This means that we can detect processes and metastable states which are not well resolved by the chemical-shift data. Consequently, we are able to dissect multiple independent processes, each of which have contributions to the observed signals which otherwise could have been interpreted as a single process with higher amplitude. Second, since the model is estimated for the molecular system as a whole, and not for individual resonances at a time, we can trace the influence of global dynamical processes and thereby connect correlated structural motions to specific experimental observables. This is in strong contrast to recent studies using long MD simulations to predict these experiments.[23,24] Finally, high-quality MSMs can be built without the need to explicitly simulating the time-scales of these experiment, which is likely to remain infeasible for the foreseeable future. With improving strategies for the construction of MSMs[101] and incorporating experimental data as restraints in simulations[102−105] as well as more sophisticated chemical shift predictors, we hope to see the role of the current limitations reduce in the near future.

## ■ METHODS

**Bayesian Markov State Models and Hidden Markov State Models.** The previously published ~1 ms trajectory of BPTI[74] was strided in 25 ns steps and used to construct a Bayesian MSM.[62] To partition the configuration space, we used all pairwise $C\alpha$ distances (1653) as input to time-lagged independent component analysis (TICA, (lag time: 2.5 $\mu$s) to project the data to a five-dimensional space of slowly relaxing collective variables.[106] This projection was transformed to a kinetic map in which Euclidean distance corresponds to kinetic distance.[107] This space was clustered into 101 states using the *k*-means algorithm and a count matrix was determined(lag-time: 625 ns). Subsequently, we sampled 50 transition probability matrices using the Bayesian formalism for reversible Markov models previously described[62] and validated the models by using the Chapman−Kolmogorov test and testing the convergence ($\Delta t$ independence) of the implied time-scales[51] (see Figure S3). These two validation steps test Markovianity and self-consistency of the models. The sampled transition probability matrices were used to evaluate all sample averages and confidence intervals shown in this paper and the Supporting Information. The MSM was used to seed the estimation of a four-state HMSM with the same lag-time as the count matrix.[53]

We built a Bayesian MSM using the previously published 1 ms MD simulation of ubiquitin in the CHARMM22* force field.[85] The trajectory was strided into 5 ns steps, and we used the same procedure as for BPTI to conduct the state segmentation count matrix estimation and validation (see Figure S3). We used the same protocol as above with the following parameters: number of pairwise $C\alpha$ distances: 2850, TICA lagtime: 55 ns, TICA space dimensionality: 3, number of clusters: 128, and MSM/count matrix lag-time: 125 ns. This MSM was used to seed the estimation of a four-state HMSM with the same lag-time. All analyses were performed using PyEMMA 2.2.[80]

**Prediction of High-Power Relaxation Dispersion Experiments.** Comparison with $R_{1\rho,\mathrm{SL}}$ and $R_{2,\mathrm{eff}}$ (see Figures 5 and S8−14) experiments involved the estimation of two positive constants, $R_{2,\mathrm{intrinsic}}$ and $\alpha$, using

$$R_{1\rho}^{\mathrm{calc}}/\sin^2(\theta) = R_{2,\mathrm{intrinsic}} + \alpha R_{1\rho,\mathrm{SL}}^{\mathrm{ex}}(\omega_e) \tag{12}$$

for spin-lock experiments where $R_{1\rho,\mathrm{SL}}^{\mathrm{ex}}(\omega_e)$ is computed from the MSM using eq 9, and using

$$R_{2,\mathrm{eff}}^{\mathrm{calc}} = R_{2,\mathrm{intrinsic}} + \alpha R_{2,\mathrm{eff}}(\nu_{\mathrm{CPMG}}) \tag{13}$$

for CPMG experiments where $R_{2,\mathrm{eff}}(\nu_{\mathrm{CPMG}})$ is computed from the MSM using eq 10. Unless otherwise stated, the second and third

slowest processes were used to compute the terms $R_{1\rho,\mathrm{SL}}^{\mathrm{ex}}(\omega_e)$ and $R_{2,\mathrm{eff}}(\nu_{\mathrm{CPMG}})$. The tilt angle in the rotating frame is given by $\tan\theta = \omega_1/\Omega$, where $\bar{\Omega}$ is the average chemical shift, and $\omega_1$ is the spin-lock field strength.[40] The fitted coefficients ($R_{2,\mathrm{intrinsic}}$ and $\alpha$) for each residue and temperature are repored in the Supporting Information. These fits assume that all auxiliary relaxation contributions (including any minor $R_1$ contributions) are described by $R_{2,\mathrm{intrinsic}}$ and that $\alpha R_{1\rho,\mathrm{SL}}^{\mathrm{ex}}(\omega_e)$ or $\alpha R_{2,\mathrm{eff}}(\nu_{\mathrm{CPMG}})$ provide the full exchange contribution to the observed relaxation rate. Alternatively, if all nonchemical exchange contributions were known, they could readily be included in the formalism, resulting in an elimination of the free parameter $R_{2,\mathrm{intrinsic}}$. The shown $\chi^2$ values were computed as

$$\chi^2 = \frac{1}{N}\sum_{i<N}(R_2^{\mathrm{pred},i} - R_2^{\mathrm{exp},i})^2/(\sigma_{\mathrm{exp},i} + \sigma_{\mathrm{pred},i})^2 \tag{14}$$

where $R_2^{\mathrm{pred}}$ is a predicted value corresponding to the experimental value (i.e., $R_{1\rho}^{\mathrm{calc}}/\sin^2(\theta)$ or $R_2^{\mathrm{exp},i}$ with uncertainty $\sigma_{\mathrm{exp},i}$). The prediction error is $\sigma_{\mathrm{pred},i} = \alpha\cdot\mathrm{std}(R_X^{\mathrm{ex}}(\nu))$, where $\mathrm{std}(R_X^{\mathrm{ex}}(\nu))$ is the standard deviation of the predicted exchange contribution from the Bayesian MSM and X = 1$\rho$,SL or X = 2,CPMG for spin-lock and CPMG type experiments, respectively.

**Prediction of Chemical Shifts.** For BPTI prediction was performed using the following software ShiftX2,[72] Sparta+,[73] AL-MOST[108] (Camshift),[71] and PPM.[70] All software was used with default parameters, except ShiftX2 which was used without the SHIFTY option. For ubiquitin only Camshift (almost) predictions were performed for all main-chain resonances (only $^{15}N^H$ and $^1H^N$ were used here). In addition, we carried out $^{15}N^H$ chemical shifts prediction using ShiftX2,[72] Sparta+,[73] and PPM[70] for 25 randomly sampled structures of each of the 128 microstates of the Bayesian MSM to generate Figure S2.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b09460.

> A derivation of the spectral decomposition of the transition probability matrix, MSM validation plots, and other figures comparing MSMs to experimental data and mutually comparing chemical shift prediction algorithms (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

*simon.olsson@fu-berlin.de
*frank.noe@fu-berlin.de

### ORCID Ⓞ

Simon Olsson: 0000-0002-3927-7897

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

## ■ REFERENCES

(1) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. *Science* **2006**, *313*, 1638−1642.

(2) Tang, C.; Iwahara, J.; Clore, G. M. *Nature* **2006**, *444*, 383−386.

(3) Teilum, K.; Smith, M. H.; Schulz, E.; Christensen, L. C.; Solomentsev, G.; Oliveberg, M.; Akke, M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18273−18278.

(4) Kjaergaard, M.; Teilum, K.; Poulsen, F. M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12535−12540.

(5) Li, Y.; Altorelli, N. L.; Bahna, F.; Honig, B.; Shapiro, L.; Palmer, A. G. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16462−16467.

(6) Kunze, M. B. A.; Wright, D. W.; Werbeck, N. D.; Kirkpatrick, J.; Coveney, P. V.; Hansen, D. F. *J. Am. Chem. Soc.* **2013**, *135*, 17862−17868.

(7) Olsson, S.; Strotz, D.; Vögeli, B.; Riek, R.; Cavalli, A. *Structure* **2016**, *24*, 1464−1475.

(8) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. *PLoS Comput. Biol.* **2014**, *10*, e1003406.

(9) van den Bedem, H.; Fraser, J. S. *Nat. Methods* **2015**, *12*, 307−318.

(10) Schröder, G. F. *Curr. Opin. Struct. Biol.* **2015**, *31*, 20−27.

(11) Salvi, N.; Abyzov, A.; Blackledge, M. *J. Phys. Chem. Lett.* **2016**, *7*, 2483−2489.

(12) Levy, R. M.; Karplus, M.; Wolynes, P. G. *J. Am. Chem. Soc.* **1981**, *103*, 5998−6011.

(13) Kessler, H.; Griesinger, C.; Lautz, J.; Mueller, A.; Van Gunsteren, W. F.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393−3396.

(14) Palmer, A. G.; Case, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 9059−9067.

(15) Chatfield, D. C.; Szabo, A.; Brooks, B. R. *J. Am. Chem. Soc.* **1998**, *120*, 5301−5311.

(16) Peter, C.; Daura, X.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *20*, 297−310.

(17) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. Ø.; Xu, H.; Trbovic, N.; Friesner, R. A.; Palmer, A. G.; Shaw, D. E. *J. Phys. Chem. B* **2008**, *112*, 6155−6158.

(18) Xue, Y.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2011**, *133*, 14614−14628.

(19) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Am. Chem. Soc.* **2012**, *134*, 3787−3791.

(20) Persson, F.; Halle, B. *J. Am. Chem. Soc.* **2013**, *135*, 8735−8748.

(21) Xia, J.; Deng, N.-j.; Levy, R. M. *J. Phys. Chem. B* **2013**, *117*, 6625−6634.

(22) Yuwen, T.; Xue, Y.; Skrynnikov, N. R. *Biochemistry* **2016**, *55*, 1784−1800.

(23) Xue, Y.; Ward, J. M.; Yuwen, T.; Podkorytov, I. S.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2012**, *134*, 2555−2562.

(24) Calligari, P.; Abergel, D. *J. Phys. Chem. B* **2014**, *118*, 3823−3831.

(25) Ma, P.; Xue, Y.; Coquelle, N.; Haller, J. D.; Yuwen, T.; Ayala, I.; Mikhailovskii, O.; Willbold, D.; Colletier, J.-P.; Skrynnikov, N. R.; Schanda, P. *Nat. Commun.* **2015**, *6*, 8361.

(26) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Phys. Chem. B* **2016**, *120*, 8313−8320.

(27) Wang, Y.; Papaleo, E.; Lindorff-Larsen, K. *eLife* **2016**, *5*, e17505.

(28) Vallurupalli, P.; Chakrabarti, N.; Pomès, R.; Kay, L. E. *Chem. Sci.* **2016**, *7*, 3602−3613.

(29) Case, D. A. *Curr. Opin. Struct. Biol.* **2013**, *23*, 172−176.

(30) Abragam, A. *The Principles of Molecular Magnetism*; Claredon Press: Oxford, 1961.

(31) Kroenke, C. D.; Loria, J. P.; Lee, L. K.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **1998**, *120*, 7905−7915.

(32) Hansen, D. F.; Feng, H.; Zhou, Z.; Bai, Y.; Kay, L. E. *J. Am. Chem. Soc.* **2009**, *131*, 16257−16265.

(33) Toyama, Y.; Osawa, M.; Yokogawa, M.; Shimada, I. *J. Am. Chem. Soc.* **2016**, *138*, 2302−2311.

(34) Akke, M.; Palmer, A. G. *J. Am. Chem. Soc.* **1996**, *118*, 911−912.

(35) Konrat, R.; Tollinger, M. *J. Biomol. NMR* **1999**, *13*, 213−221.

(36) Massi, F.; Johnson, E.; Wang, C.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **2004**, *126*, 2247−2256.

(37) Mangia, S.; Traaseth, N. J.; Veglia, G.; Garwood, M.; Michaeli, S. *J. Am. Chem. Soc.* **2010**, *132*, 9979−9981.

(38) McConnell, H. M. *J. Chem. Phys.* **1958**, *28*, 430.

(39) Trott, O.; Palmer, A. G., III *J. Magn. Reson.* **2004**, *170*, 104−112.

(40) Palmer, A. G.; Massi, F. *Chem. Rev.* **2006**, *106*, 1700−1719.

(41) Wittebort, R. J.; Szabo, A. *J. Chem. Phys.* **1978**, *69*, 1722.

(42) Brüschweiler, R.; Case, D. A. *Prog. Nucl. Magn. Reson. Spectrosc.* **1994**, *26*, 27−58.

(43) Chao, F.-A.; Byrd, R. A. *J. Am. Chem. Soc.* **2016**, *138*, 7337−7345.

(44) Abergel, D.; Palmer, A. G. *ChemPhysChem* **2004**, *5*, 787−793.

(45) Millet, O.; Loria, J. P.; Kroenke, C. D.; Pons, M.; Palmer, A. G. *J. Am. Chem. Soc.* **2000**, *122*, 2867−2877.

(46) Schütte, C.; Fischer, A.; Huisinga, W.; Deuflhard, P. *J. Comput. Phys.* **1999**, *151*, 146−168.

(47) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. *J. Chem. Phys.* **2007**, *126*, 155102.

(48) Chodera, J. D.; Dill, K. A.; Singhal, N.; Pande, V. S.; Swope, W. C.; Pitera, J. W. *J. Chem. Phys.* **2007**, *126*, 155101.

(49) Buchete, N.-V.; Hummer, G. *J. Phys. Chem. B* **2008**, *112*, 6057−69.

(50) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. *J. Chem. Phys.* **2009**, *131*, 124101.

(51) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. *J. Chem. Phys.* **2011**, *134*, 174105.

(52) An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. *Advances in Experimental Medicine and Biology*; Bowman, G. R., Pande, V. S., Noé, F., Eds.; Springer: Heidelberg, 2014; Vol. 797.

(53) Noé, F.; Wu, H.; Prinz, J.-H.; Plattner, N. *J. Chem. Phys.* **2013**, *139*, 184114.

(54) Deverell, C.; Morgan, R.; Strange, J. *Mol. Phys.* **1970**, *18*, 553−559.

(55) Kloiber, K.; Konrat, R. *J. Biomol. NMR* **2000**, *18*, 33−42.

(56) Wang, C.; Palmer, A. G. *J. Biomol. NMR* **2002**, *24*, 263.

(57) Orekhov, V. Y.; Korzhnev, D. M.; Kay, L. E. *J. Am. Chem. Soc.* **2004**, *126*, 1886−1891.

(58) Salvi, N.; Ulzega, S.; Ferrage, F.; Bodenhausen, G. *J. Am. Chem. Soc.* **2012**, *134*, 2481−2484.

(59) Sarich, M.; Noé, F.; Schütte, C. *Multiscale Model. Simul.* **2010**, *8*, 1154−1177.

(60) Schwantes, C. R.; Pande, V. S. *J. Chem. Theory Comput.* **2013**, *9*, 2000−2009.

(61) Perez-Hernandez, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. *J. Chem. Phys.* **2013**, *139*, 015102.

(62) Trendelkamp-Schroer, B.; Wu, H.; Paul, F.; Noé, F. *J. Chem. Phys.* **2015**, *143*, 174101.

(63) Kube, S.; Weber, M. *J. Chem. Phys.* **2007**, *126*, 024103.

(64) Hummer, G.; Szabo, A. *J. Phys. Chem. B* **2015**, *119*, 9029−9037.

(65) Orioli, S.; Faccioli, P. Dimensional Reduction of Markov State Models from Renormalization Group Theory. 2016, arXiv:1604.07614 [cond-mat.stat-mech] (accessed April 26, 2016.

(66) Noe, F.; Doose, S.; Daidone, I.; Lollmann, M.; Sauer, M.; Chodera, J. D.; Smith, J. C. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 4822−4827.

(67) Keller, B. G.; Prinz, J.-H.; Noé, F. *Chem. Phys.* **2012**, *396*, 92−107.

(68) Chodera, J. D.; Noé, F. *J. Chem. Phys.* **2010**, *133*, 105102.

(69) Allerhand, A.; Thiele, E. *J. Chem. Phys.* **1966**, *45*, 902.

(70) Li, D.-W.; Brüschweiler, R. *J. Biomol. NMR* **2012**, *54*, 257−265.

(71) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 13894−5.

(72) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. *J. Biomol. NMR* **2011**, *50*, 43−57.

(73) Shen, Y.; Bax, A. *J. Biomol. NMR* **2010**, *48*, 13−22.

(74) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341−346.

(75) Grey, M. J.; Wang, C.; Palmer, A. G. *J. Am. Chem. Soc.* **2003**, *125*, 14324−14335.

(76) Otting, G.; Liepinsh, E.; Wuethrich, K. *Biochemistry* **1993**, *32*, 3571−3582.

(77) Long, D.; Brüschweiler, R. *J. Am. Chem. Soc.* **2011**, *133*, 18999−19005.

(78) Wang, C.; Grey, M. J.; Palmer, A. G., III *J. Biomol. NMR* **2001**, *21*, 361−366.

(79) Loria, J. P.; Rance, M.; Palmer, A. G., III *J. Biomol. NMR* **1999**, *15*, 151−155.

(80) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.

(81) Takeda, M.; Miyanoiri, Y.; Terauchi, T.; Kainosho, M. *J. Biomol. NMR* **2016**, *66*, 37−53.

(82) Charlier, C.; Khan, S. N.; Marquardsen, T.; Pelupessy, P.; Reiss, V.; Sakellariou, D.; Bodenhausen, G.; Engelke, F.; Ferrage, F. *J. Am. Chem. Soc.* **2013**, *135*, 18665−18672.

(83) Massi, F.; Grey, M. J.; Palmer, A. G., III *Protein Sci.* **2005**, *14*, 735−742.

(84) Ban, D.; Funk, M.; Gulich, R.; Egger, D.; Sabo, T. M.; Walter, K. F. A.; Fenwick, R. B.; Giller, K.; Pichierri, F.; de Groot, B. L.; Lange, O. F.; Grubmüller, H.; Salvatella, X.; Wolf, M.; Loidl, A.; Kree, R.; Becker, S.; Lakomek, N.-A.; Lee, D.; Lunkenheimer, P.; Griesinger, C. *Angew. Chem., Int. Ed.* **2011**, *50*, 11437−11440.

(85) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 5915−5920.

(86) Mills, J. L.; Szyperski, T. *J. Biomol. NMR* **2002**, *23*, 63−67.

(87) Smith, C. A.; Ban, D.; Pratihar, S.; Giller, K.; Paulat, M.; Becker, S.; Griesinger, C.; Lee, D.; de Groot, B. L. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 3269−3274.

(88) Dittmer, J.; Bodenhausen, G. *J. Am. Chem. Soc.* **2004**, *126*, 1314−1315.

(89) Kitahara, R.; Yokoyama, S.; Akasaka, K. *J. Mol. Biol.* **2005**, *347*, 277−285.

(90) Wong, M.; Khirich, G.; Loria, J. P. *Biochemistry* **2013**, *52*, 6548−6558.

(91) Bottaro, S.; Boomsma, W.; Johansson, K. E.; Andreetta, C.; Hamelryck, T.; Ferkinghoff-Borg, J. *J. Chem. Theory Comput.* **2012**, *8*, 695−702.

(92) Huang, K. Y.; Amodeo, G. A.; Tong, L.; McDermott, A. *Protein Sci.* **2011**, *20*, 630−639.

(93) Sidhu, A.; Surolia, A.; Robertson, A. D.; Sundd, M. *J. Mol. Biol.* **2011**, *411*, 1037−1048.

(94) Fenwick, R. B.; Esteban-Martín, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. *J. Am. Chem. Soc.* **2011**, *133*, 10336−10339.

(95) Lindner, B.; Yi, Z.; Prinz, J.-H.; Smith, J. C.; Noé, F. *J. Chem. Phys.* **2013**, *139*, 175101.

(96) Christensen, A. S.; Linnet, T. E.; Borg, M.; Boomsma, W.; Lindorff-Larsen, K.; Hamelryck, T.; Jensen, J. H. *PLoS One* **2013**, *8*, e84123.

(97) Larsen, A. S.; Bratholm, L. A.; Christensen, A. S.; Channir, M.; Jensen, J. H. *PeerJ* **2015**, *3*, e1344.

(98) Ermolenko, D. N.; Dangi, B.; Gvritishvili, A.; Gronenborn, A. M.; Makhatadze, G. I. *Biophys. Chem.* **2007**, *126*, 25−35.

(99) Abergel, D.; Palmer, A. G. *Concepts Magn. Reson.* **2003**, *19A*, 134−148.

(100) Kubo, R. *J. Math. Phys.* **1963**, *4*, 174.

(101) Wu, H.; Paul, F.; Wehmeyer, C.; Noé, F. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E3221−E3230.

(102) Olsson, S.; Ekonomiuk, D.; Sgrignani, J.; Cavalli, A. *J. Am. Chem. Soc.* **2015**, *137*, 6270−8.

(103) Olsson, S.; Vögeli, B. R.; Cavalli, A.; Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K.; Hamelryck, T. *J. Chem. Theory Comput.* **2014**, *10*, 3484−3491.

(104) Cavalli, A.; Camilloni, C.; Vendruscolo, M. *J. Chem. Phys.* **2013**, *138*, 094112.

(105) Olsson, S.; Frellsen, J.; Boomsma, W.; Mardia, K. V.; Hamelryck, T. *PLoS One* **2013**, *8*, e79439.

(106) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. *J. Chem. Phys.* **2013**, *139*, 015102.

(107) Noé, F.; Clementi, C. *J. Chem. Theory Comput.* **2015**, *11*, 5002−5011.

(108) Fu, B.; Sahakyan, A. B.; Camilloni, C.; Tartaglia, G. G.; Paci, E.; Caflisch, A.; Vendruscolo, M.; Cavalli, A. *J. Comput. Chem.* **2014**, *35*, 1101−1105.