Journal of Magnetic Resonance 338 (2022) 107196

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Motional clustering in supra-τ_c conformational exchange influences NOE cross-relaxation rate

Christopher Kolloff [a,b], Adam Mazur [a], Jan K. Marzinek [c], Peter J. Bond [c,d], Simon Olsson [b,*], Sebastian Hiller [a,*]

[a] Biozentrum, Universität Basel, Spitalstrasse 41, Basel 4056, Switzerland
[b] Department of Computer Science and Engineering, Chalmers University of Technology, Rännvägen 6, Göteborg 412 58, Sweden
[c] Bioinformatics Institute (A*STAR), 30 Biopolis Street, #07-01 Matrix, Singapore 138671, Singapore
[d] National University of Singapore, Department of Biological Sciences, 14 Science Drive 4, Singapore 117543, Singapore

ARTICLE INFO

Article history:
Received 15 December 2021
Revised 1 March 2022
Accepted 13 March 2022
Available online 23 March 2022

Keywords:
Protein dynamics
Nuclear Overhauser Effect
Methyl spectroscopy
Slow conformational exchange
Malate Synthase G
Molecular dynamics simulations

ABSTRACT

Biomolecular spin relaxation processes, such as the NOE, are commonly modeled by rotational τ_c-tumbling combined with fast motions on the sub-τ_c timescale. Motions on the supra-τ_c timescale, in contrast, are considered to be completely decorrelated to the molecular tumbling and therefore invisible. Here, we show how supra-τ_c dynamics can nonetheless influence the NOE build-up between methyl groups. This effect arises because supra-τ_c motions can cluster the fast-motion ensembles into discrete states, affecting distance averaging as well as the fast-motion order parameter and hence the cross-relaxation rate. We present a computational approach to estimate methyl–methyl cross-relaxation rates from extensive ($> 100 \times \tau_c$) all-atom molecular dynamics (MD) trajectories on the example of the 723-residue protein Malate Synthase G. The approach uses Markov state models (MSMs) to resolve transitions between metastable states and thus to discriminate between sub-τ_c and supra-τ_c conformational exchange. We find that supra-τ_c exchange typically increases NOESY cross-peak intensities. The methods described in this work extend the theory of modeling sub-µs dynamics in spin relaxation and thus contribute to a quantitative estimation of NOE cross-relaxation rates from MD simulations, eventually leading to increased precision in structural and functional studies of large proteins.

1. Introduction

The importance of the NOE for the field of biomolecular NMR spectroscopy cannot be understated. To this day, it remains the central method to infer distances between spins and to use them as restraints for solving three-dimensional structures of biological macromolecules [1–4]. In its most simple form, an isolated spin pair approximation is used to obtain an upper-distance limit between two spins from the linear regime of the NOE build-up. In this regime, magnetization is transferred directly between spin pairs and is not significantly modulated by the presence of other nearby spins, simplifying the analysis [5]. This property thereby sidesteps the need for more complex relaxation pathways in the analysis of dipolar couplings. More advanced approaches center around the idea to obtain exact distances between spins and to use those interpretations to determine the structure of proteins [6–8] and account for spin diffusion [9,10].

One limitation of solution NMR is the size of the systems under study. In recent years, there have been many advances, both on the spectroscopic [11–14] as well as on the biochemical side [15–19], to study the structure, function, and dynamics of ever-larger systems using methyl labeling and spectroscopy [20–24]. The study of those systems is mainly possible because of the favorable spectroscopic properties of methyl groups even in cases of slow molecular tumbling. In addition, methyl-labeled proteins on a highly deuterated background have a reduced number of spins, which leads to fewer overall signals and hence, a diminished spin diffusion effect. Importantly, since the spin density in methyl-labelled systems is low, also long NOESY mixing times τ_m can be used, leading to substantial cross-peak intensities even for long interspin distances. In such experiments, the isolated spin pair approximation breaks down and second-order magnetization effects have to be accounted for. A quantitative interpretation of the resulting NOESY cross-peaks therefore not only requires full relaxation matrix analysis but also a model that accounts for the dynamics of the system.

A major reason for the versatility of protein function and regulation is their conformational plasticity [25,26]. This also means that most biologically relevant systems have complicated energy land-

* Corresponding authors.
E-mail addresses: kolloff@chalmers.se (C. Kolloff), adam.mazur@unibas.ch (A. Mazur), marzinekj@bii.a-star.edu.sg (J.K. Marzinek), peterjb@bii.a-star.edu.sg (P.J. Bond), simonols@chalmers.se (S. Olsson), sebastian.hiller@unibas.ch (S. Hiller).

https://doi.org/10.1016/j.jmr.2022.107196
1090-7807/

scapes with multiple states [27]. Prominent computational methods to sample the conformational space of biomolecules are molecular dynamics (MD) simulations, which are, however, limited by the timescale involved [28–30]. In experiments, on the other hand, it is often difficult to observe conformational changes at the single-molecule level [30,31]. Nuclear spin relaxation experiments are often used for parameterizing and assessing the quality of the force field employed in the MD simulation [32–34], and a quantitative interpretation of NOESY spectra provides a valuable tool. However, dynamics on a wide range of timescales from pico- to milliseconds (i.e., $10^{-12} - 10^{-3}$ s), modulate NOE signals, complicating their analysis. Early efforts found that MD simulations are a useful tool to analyze the motional effects of proteins, as they permit the direct calculation of correlation functions to determine the effects of motional averaging [35]. While there have been attempts to model the internal dynamics of proteins using MD, the simulation timescale was often a limiting factor to model the cross-relaxation rate accurately [36,37]. Despite the extensive theoretical knowledge that is available on cross-relaxation in nuclear spin systems [38–43], we currently have very limited means to quantitatively interpret NOE cross-peaks of biomolecular systems. In methyl-labeled systems, it is particularly important to obtain a comprehensive understanding of the underlying dynamics since much of the state-of-the-art work on the structure, function, and dynamics of large protein relies on an interpretation of methyl NOE cross-peaks [23,44,8,45]. In addition, NOE spectroscopy is one of the central methods for the assignment of methyl-labelled systems [46,18,47,4].

In this work, we investigate the effect of sub-microsecond dynamics on cross-relaxation rate averaging in methyl-labelled systems using explicitly solvated MD simulations with multiple force fields, taking the well-studied large protein Malate Synthase G as a model system. We distinguish between a sub-τ_c ($\tau_{ex} \ll \tau_c$) and a supra-τ_c ($\tau_c \ll \tau_{ex}$) regime relative to the rotational correlation time of the protein and demonstrate how the presence of exchange in either regime affects the cross-peak build-up. The discrimination between those two regimes is done using hidden Markov state models (hMSMs). These models estimate the Markovian operator of the conformational dynamics via statistical analysis of MD simulations. Through these models, we get access to structures of important metastable states and the kinetics of conformational exchange, thus providing a quantitative yet interpretable description of the underlying dynamics. We then compare the cross-relaxation rates and show how their different averaging can have tremendous effects on the resulting NOE build-up.

2. Theory

Consider a system of n spins 1/2. NOE signal intensities can be well approximated by [42]

$$\mathbf{I}(\tau_m) = \mathbf{I}(0) \cdot \exp(-\mathbf{R}\tau_m), \tag{1}$$

where $\mathbf{I}(\tau_m)$ is the matrix of signal intensities after mixing time τ_m, $\mathbf{I}(0)$ is the vector of initial signal intensities, and $\exp(-\mathbf{R}\tau_m)$is the matrix exponential of the $n \times n$ relaxation matrix $\mathbf{R}$ for the Solomon equations [39]. The symmetric matrix $\mathbf{R}$ contains the auto- and cross-relaxation rate constants, ρ and σ, respectively

$$\mathbf{R} = \begin{bmatrix} \rho_i & \sigma_{ij} & \cdots & \sigma_{in} \\ \sigma_{ji} & \rho_j & & \vdots \\ \vdots & & \ddots & \\ \sigma_{ni} & \cdots & & \rho_n \end{bmatrix}. \tag{2}$$

For a pair of protons i and j, the cross- and auto-relaxation rate constants can be described as

$$\sigma_{ij} = \frac{1}{10}K^2\left[6J_{ij}(2\omega_H) - J(0)\right] \tag{3}$$

and

$$\rho_{ij} = \frac{1}{10}K^2\left[3J_{ij}(2\omega_H) + \frac{2}{3}J_{ij}(\omega_H) + \frac{1}{2}J(0)\right]. \tag{4}$$

Consequently,

$$\rho_i = \sum_{j=1 j\neq i}^{n} \rho_{ij} \tag{5}$$

with the constant K

$$K = \frac{\mu_0}{4\pi}\hbar\gamma_H. \tag{6}$$

$J(\omega)$ is the spectral density function for a molecule undergoing isotropic tumbling (see Section 2.1).

Due to the particular chemical structure of methyl groups, the auto-relaxation rate of methyl groups consists of a linear combination of intra- and inter-methyl relaxation terms to account for the effects of fast internal motion and the close proximity of the three protons. A detailed description of these terms can be found in the appendix, Section A.2.

2.1. ^{1}H – ^{1}H time correlation functions and S^2 order parameters

A common framework to account for fast internal dynamics is the Lipari-Szabo model-free formalism [48], which defines the order parameter S^2 as

$$S^2_{ij} := \lim_{t\to\infty} C^{internal}_{ij}(t), \tag{7}$$

where $C^{internal}_{ij}$ is the internal correlation function between methyl groups i and j. Describing the cross-relaxation of methyl groups, we are mainly interested in the internal motion of intermethyl ^{1}H–^{1}H vectors. For dipolar relaxation, the autocorrelation function is determined by the time-dependent dipolar coupling Hamiltonian, which we can express in terms of the internuclear distance r and the second-rank spherical harmonics functions $Y^m_2(\theta, \varphi)$ [41]:

$$C(\tau) = 4\pi\sum_{m=-2}^{2}\left\langle\frac{Y^m_2(\theta(t),\varphi(t))Y^m_2(\theta(t+\tau),\varphi(t+\tau))^*}{r^3(t)r^3(t+\tau)}\right\rangle, \tag{8}$$

where θ and φ are the polar and azimuthal angles of the inter-spin vector in the laboratory frame, at time points t and $t+\tau$, respectively. $\langle\cdot\rangle$ denotes time averaging. There are several particularities that need to be considered when calculating the NOESY build-up of methyl groups: [41] proposed to model the internal motion of methyl groups by assuming that the protons randomly execute rotational jumps about the methyl axis between the three sites. This means that the position of the methyl protons is explicitly taken into account instead of placing a pseudo-atom that represents all three protons. Hence, when calculating the distance averaging, there are 9 different vectors involved. Since the cross-relaxation rate heavily depends on the internuclear distance, a proper averaging is crucial [36,49]. Eq. 8 for methyl–methyl correlations then becomes the normalized time correlation function in the molecular frame:

$$C^{internal}_{ij}(\tau) = \frac{4\pi}{5}\left\langle\frac{1}{9}\sum_{q=1}^{3}\sum_{p=1}^{3}r^{-6}_{pq}(t)\right\rangle^{-1}\left|\sum_{m=-2}^{2}\left\langle\frac{1}{9}\sum_{q=1}^{3}\sum_{p=1}^{3}\frac{Y^m_2\left(\theta^{mol}_{pq}(t),\varphi^{mol}_{pq}(t)\right)Y^m_2\left(\theta^{mol}_{pq}(t+\tau),\varphi^{mol}_{pq}(t+\tau)\right)^*}{r^3_{pq}(t)r^3_{pq}(t+\tau)}\right\rangle\right|. \tag{9}$$

Here, we make the assumption that the internal correlation time of the methyl group τ_{internal} is much faster than the overall rotational correlation time τ_c. Given that τ_{internal} is usually in the timescale of tens of picoseconds [50–52] and τ_c is at least one order of magnitude slower for large proteins, where methyl labelling is usually employed, this assumption is reasonable. Similarly, we can calculate S^2 for fast internal motions as

$$S_{ij}^2 = \frac{4\pi}{5}\left\langle \frac{1}{9}\sum_{q=1}^{3}\sum_{p=1}^{3} r_{pq}^{-6}(t)\right\rangle^{-1} \times \sum_{m=-2}^{2}\left|\left\langle \frac{1}{9}\sum_{q=1}^{3}\sum_{p=1}^{3}\frac{Y_2^m\left(\theta_{pq}^{\text{mol}}(t),\varphi_{pq}^{\text{mol}}(t)\right)}{r_{pq}^3(t)}\right\rangle\right|^2. \tag{10}$$

For isotropically tumbling molecules undergoing Brownian motion, the overall correlation $C^{\text{overall}}(t)$ can be approximated by a single exponential of the form

$$C^{\text{overall}}(t) = \exp\left(\frac{-t}{\tau_c}\right), \tag{11}$$

where τ_c is the rotational correlation time of the protein. Following the Lipari-Szabo formalism and the decorrelation assumption [53], the total correlation function C^{tot} is expressed as the product of the correlation functions of overall tumbling $C^{\text{overall}}(t)$ and internal motion $C^{\text{internal}}(t)$:

$$\begin{aligned} C_{ij}^{tot}(t) &= C^{\text{overall}}(t)\langle r^{-6}\rangle C_{ij}^{\text{internal}}(t) \\ &= \langle r_{ij}^{-6}\rangle S_{ij}^2 e^{-t/\tau_c} + \langle r_{ij}^{-6}\rangle\left(1-S_{ij}^2\right)e^{-t/\tau_c}, \end{aligned} \tag{12}$$

where the internal motion lies below the rotational correlation time of large proteins. The spectral density function is then given by

$$\begin{aligned} J_{ij}(\omega) &= \int_0^\infty C_{ij}^{\text{tot}}(t)\cos(\omega t)\,\mathrm{d}t \\ &= \left\langle r_{ij}^{-6}\right\rangle S_{ij}^2\frac{\tau_c}{1+\omega^2\tau_c^2} + \left\langle r_{ij}^{-6}\right\rangle\left(1-S_{ij}^2\right)\frac{\tau_e}{1+\omega^2\tau_e^2} \end{aligned} \tag{13}$$

with $\tau_e^{-1} = \tau_c^{-1} + \tau_{\text{internal}}^{-1}$.

2.2. Kinetic exchange in the sub- and supra-τ_c limit

NOE signals are modulated by conformational dynamics covering timescales over several orders of magnitude. Here, we examine the effect of dynamics that occur above the timescale of the rotational correlation time (supra-τ_c) but are a subclustering of different states, each with fast internal motions. Consider two dipolar-coupled spins 1/2, *i* and *j*. Fig. 1 shows how spin *i* visits two macroscopic states, A and B. In the first case, the time constant of conformational exchange τ_{ex} between the two states is below the rotational correlation time of the protein, i.e., $\tau_{ex} \ll \tau_c$. In this case, the resulting spectral density function is [36]

$$J_{ij}(\omega) = \left\langle\frac{1}{r_{ij}^6}\right\rangle S_{ij}^2\frac{\tau_c}{1+\omega^2\tau_c^2} + \left\langle\frac{1}{r_{ij}^6}\right\rangle\left(1-S_{ij}^2\right)\frac{\tau_e}{1+\omega^2\tau_e^2} \tag{14}$$

with the cross-relaxation rate

$$\langle\sigma_{ij}\rangle = \sigma\left(S_{ij}^2, \left\langle r_{ij}^{-6}\right\rangle\right) \tag{15}$$

These expressions imply that the distances between *i* and *j* are ensemble averaged and there is no distinction between the two macroscopic states. On the other hand, if the exchange is above the timescale of τ_c and each of the macroscopic state is a cluster of fast internal dynamics, we have to consider the two states separately. This means calculating the distance averaging as well as the order parameter for the $^1H-{}^1H$ correlation independently. These two signal regimes are shown schematically in Fig. 1.

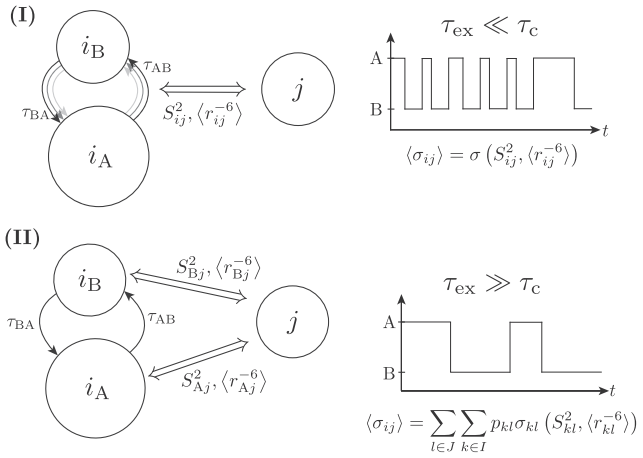

Fig. 1. Schematic overview of the two different exchange regimes. Markov state models can be built from MD simulations to estimate the exchange rate between different states. (**I**) Situation with a time constant of exchange τ_{ex} between conformations A and B that is much faster than the rotational correlation time of the protein. The resulting S^2 order parameter and distance will be averaged over the whole ensemble. (**II**) Situation in which the exchange is slower than τ_c. Each state, A and B, consists of a subclustering of fast motion ensembles with a separate S^2 order parameter and distance average. The resulting cross-relaxation rate becomes the linear combination of the weighted contributions from each state. $I = \{k,\ldots\}$ and $J = \{l,\ldots\}$ are the sets of states for methyl groups *i* and *j*, respectively.

The resulting cross-relaxation rate is a linear combination of cross-relaxation rates of the different states $\{k,l,\ldots\}$ scaled by their respective thermodynamic weights *p*

$$\langle\sigma_{ij}\rangle = \sum_{l\in J}\sum_{k\in I} p_{kl}\sigma_{kl}\left(S_{kl}^2, \left\langle r_{kl}^{-6}\right\rangle\right), \tag{16}$$

where $I = \{k,\ldots\}$ and $J = \{l,\ldots\}$ are the set of states of methyl groups *i* and *j*, respectively. For the within-state exchange, Eq. 14 and 15 apply. These equations hold up to the µs to ms timescale, when line broadening effects and subsequently line splitting starts occurring. Notably, the descriptions fall short in situations of intramolecular motions near the frequency of the rotational tumbling, where interference effects may occur [53].

2.3. Estimation of exchange rates using MD simulations and Markov modeling

We can use MD simulations and Markov models to directly compute dynamic observables [54,55,49]. A number of methods and tools are available to estimate the kinetics observed in the trajectory [56–58]. A popular framework for the estimation of discrete state kinetics are a class of statistical methods called Markov state models [59,60]. This approach typically entails feature selection and dimensionality reduction followed by clustering and finally estimation of the transition matrix, **T** [61,62]. These models encode transitions $p(j \to i;\tau)$ between conformational states *i* and *j* in element $\mathbf{T}_{ij}$. These Markov models are therefore time-discrete models of conformational exchange. We can relate **T** to a rate matrix **K** with the rates (in time^{-1}) of exchange between the states *i* and *j* in $\mathbf{K}_{ij}$ and a lag time, τ via the expression

$$\mathbf{T}(\tau) = \exp(-\mathbf{K}\tau). \tag{17}$$

Here, exp denotes the matrix exponential. Consequently, we can compute the relaxation timescales of the molecular system using the eigendecomposition of **T**:

$$\tau_{ex,i} = -\frac{\tau}{\ln|\lambda_i|}, \tag{18}$$

where λ is the eigenvalue of the ith process [61]. The eigenvectors associated with each eigenvalue λ_i allow us to link the relaxation timescale to the exchange processes between different metastable conformational states. In the MSM field, the relaxation timescales are referred to as 'implied timescales.'

Here, we use Markov state models to discriminate between sub- and supra-τ_c exchange. We classify a methyl pair to be in supra-τ_c exchange if the estimated timescale exceeds $\approx 10 \times \tau_c$. Note that we do not rely on the exact estimation of the exchange rates. Instead, we use a bandwidth filter of $>10 \times \tau_c$ and consider exchange processes below as fast, sub-τ_c. The corresponding state assignments are then used for the calculation of the dynamics parameters, $\langle r^{-6} \rangle$ and S^2.

2.4. Calculation of the NOESY time correlation function from MSMs

Markov state models allow us to compute stationary and dynamic expectation values for comparison with experiment [55,63]. To compute the NOESY signal we need to compute the appropriate time correlation function (Eq. 9): a sum over the five second-rank spherical harmonic terms averaged over the ensemble and over the methyl spins. In general, we compute dynamic equilibrium correlation function (time correlation function) of an observable function $\mathbf{o}(\mathbf{x})$ from an MSM via the expression

$$C_{\mathbf{o}}(k\tau) = \mathbb{E}[\mathbf{o}(t)\mathbf{o}(t+k\tau)^*] = \mathbf{o}^* \mathbf{\Pi} \mathbf{T}^k(\tau)\mathbf{o}, \tag{19}$$

where $\mathbf{\Pi}$ is a diagonal matrix with the stationary probabilities of the Markov states with transition probabilities $\mathbf{T}(\tau)$, $\mathbf{o}$ is a complex-valued vector of the function, $\mathbf{o}(\mathbf{x})$, evaluated in the Markov states and k is a positive integer value. In practice, we compute $\mathbf{o}$ as an average function value over all configurations i in each Markov state $\mathscr{S}_j$. For the five rank-two spherical harmonics we get,

$$\mathbf{o}_m^j = \left\langle \frac{1}{9} \sum_{q=1}^{3} \sum_{p=1}^{3} \frac{Y_2^m\left(\theta_{pq}^{\mathrm{mol}}(i), \varphi_{pq}^{\mathrm{mol}}(i)\right)}{r_{pq}(i)^3} \right\rangle_{\mathscr{S}_j} \tag{20}$$

for every state, $\mathscr{S}_j$ in the MSM. Eq. (9) then becomes

$$C_{ij}^{\mathrm{internal}}(\tau) = \frac{4\pi}{5} \left\langle \frac{1}{9} \sum_{q=1}^{3} \sum_{p=1}^{3} r_{pq}^{-6}(t) \right\rangle^{-1} \left| \sum_{m=-2}^{2} \mathbf{o}_m^* \mathbf{\Pi} \mathbf{T}^k(\tau) \mathbf{o}_m \right|. \tag{21}$$

In practice, we can compute the correlation functions (Eq. 19) using a sum of exponential decays emerging from the expressions spectral form [55,63].

3. Results and Discussion

We ran MD simulations of Malate Synthase G, a 723-residue protein with a molecular weight of 82.5 kDa using five different force fields: a99SB-disp [64], CHARMM-36m [65], Amber99SB*-ILDN-q [66], DES-Amber and DES-Amber-SF1.0 [67]. Even though it would require tens of milliseconds of simulation times for MSG to converge globally [14], it is possible for local subsystems to converge on a much shorter timescale [68]. In order to classify the kinetic exchange of methyl groups into sub- and supra-τ_c, we ran 5 µs simulations in all force fields. With a rotational correlation time τ_c of 37 ns for MSG [69], we consider the supra-τ_c regime as conformational exchange with a time constant of exchange τ_{ex} of more than $10 \times \tau_c$, i.e., >370 ns. The kinetic exchange was modeled using hidden MSMs, which were estimated for every methyl–methyl contact of up to 10 Å.

As a first step, we want to discuss a specific example in detail to illustrate the effect. Fig. 2 shows the two methyl groups from the simulation using the DES-Amber force field [67], Ile346$^{\delta 1}$ and Leu362$^{\delta 2}$, that are within NOE distance to each other. In order to identify metastable states, we performed time-lagged independent component analysis (tICA) [70,71] on the side chain and main chain dihedrals as well as the internuclear distance over the course of the trajectory and projected the space onto the slowest independent components, resulting in four clusters (Fig. 2A). Similar to principal component analysis (PCA), tICA performs a linear transformation of the given feature vectors. In tICA, autocovariance at a given lag time is maximized, contrary to PCA, where for each projection of the data, variance is maximized. We then proceeded with the use of hidden MSMs to model the exchange timescale between the four clusters [72]. The implied timescale (ITS) plot in Fig. 2B shows the slowest process of the system, and thus give an estimate of the timescale of exchange at a given lag time. We can see that the timescale for the slowest process (yellow line) is consistently above the $10 \times \tau_c$ threshold at any lag time and converges after 0.5 ns. This means that all other processes are sub-τ_c, including the exchange within the clusters (see supplementary material, figure S8). Therefore, we only have two metastable states that are in supra-τ_c exchange with each other. From the implied timescale plot, we see that the estimated exchange process is invariant of the lag time, further confirming that the processes we are interested in modelling are very slow. The Markov state model was validated by checking whether the model we have built makes predictions which are consistent with the estimates from our simulation data. This was done using a Chapman-Kolmogorov (CK) test (figure S1). In this test, k MSMs, which are integer multiples of lag time τ are estimated. The predicted populations at lag time τ can then be compared with those at multiples of the lag time $k\tau$ (estimated). In this manner, we test self-consistency of the MSM. After statistical validation of the Markov state model, we can therefore safely assume that the exchange between the two clusters is supra-τ_c and the NOE signal that builds up between the two methyl groups averages according to Eq. 16. In Fig. 2C, we show five samples from each of those states. By visual inspection, it can be seen that the distances between the methyl groups (dashed lines) in state 2 are much shorter than in state 1, where the inter-methyl distance clusters around 6 Å. This is also reflected more quantitatively in the distance histogram (Fig. 2D). After validating the Markov state model and identifying two states that are in supra-τ_c exchange, we can calculate the state-specific dynamics parameters, S^2 and $\langle r^{-6} \rangle$. Hence, we obtain a cross-relaxation rate constant that considers the two states separately instead of an average over the whole ensemble. The resulting NOESY cross-peak build-up between Ile346$^{\delta 1}$ and Leu362$^{\delta 2}$ is shown in Fig. 2E as a function of mixing time τ_m. The plot contains two curves. The dark-violet curve shows the build-up that results from treating the two states as supra-τ_c exchange. In contrast, we show the light-violet curve that would result from not discriminating between the two states and instead, averaging the dynamics parameters over the whole trajectory, i.e., treating the ensemble as sub-τ_c exchange (Eq. 15). At all mixing times, we see an increase in signal intensity of a factor of around 2 by accounting for supra-τ_c exchange. It is important to note that for the calculations, in both scenarios, the whole trajectory was considered. The key difference between sub- and supra-τ_c averaging lies in the assignment of each time-step to a state. In this example, we saw how discriminating between the two exchange regimes can lead to a drastic increase in the expected NOE signal. When expanding the same type of analysis to the entire protein, we found that around 100 of the $\approx 1,500$ local systems that are within a 10-Å distance cut-off where affected by supra-τ_c exchange. These around 100 methyl pairs are located throughout the protein without apparent clustering in specific regions (see figure S9). This is, however, not surprising given that dynamics on the supra-τ_c timescale often comprise side chain rotations which can occur independent of the protein's

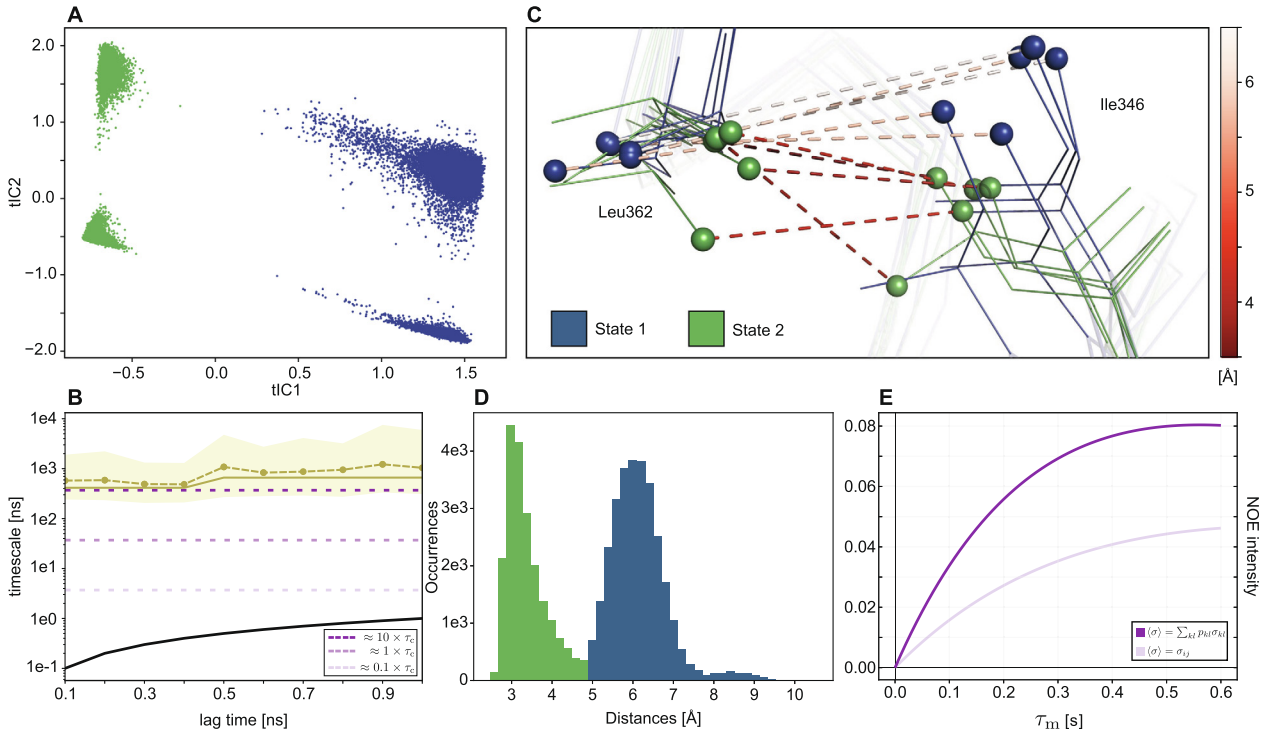

Fig. 2. Dynamics of Ile346$^{\delta 1}$ and Leu362$^{\delta 2}$ of MSG. **A**: Plot of the slowest two time-lagged independent components (tICs) of the trajectory. The conformers cluster in two metastable states that are in supra-τ_c exchange (blue and green). The exchange within each of the two cluster was determined to be sub-τ_c. **B**: Implied Time Scale (ITS) plots showing the slowest process that corresponds to the exchange between state 1 and 2 (yellow line) as a function of lag time (Eq. 18). The different exchange regimes are shown in purple dashes for orientation. The estimated timescales and the predictions from the data using the Chapman-Kolmogorov test are shown in dashed lines and solid lines, respectively. The 95 % confidence interval is shown in the shaded area. The black line at the bottom represents the lag time threshold. **C**: 5 representative structures were sampled from each of the two clusters (shown in blue and green, respectively). The distances between the two methyl groups (spheres) are shown in dashes and color coded according to the color bar shown on the right. In state 2, the intermethyl distances are visibly shorter than in state 1. **D**: Distance histogram over the whole trajectory. The two peaks were assigned to each of the states and color coded respectively. **E**: The resulting NOE build-up between the two moieties using supra-τ_c distance averaging (dark purple, Eq. 16) and, for comparison, disregarding supra-τ_c exchange and using sub-τ_c averaging (light purple, Eq. 15).

function. The dynamics of MSG related to its function have been reported to occur mostly on even slower timescales (micro- to milliseconds) [73,52,74].

3.1. Supra-τ_c dynamics in the internal $^1H - ^1H$ correlation function

After having inspected a first example for a spin system that is affected by supra-τ_c motions, we now want to examine specifically how supra-τ_c dynamics affect the correlation function as well as the S^2 order parameter and how the approach to separate the clusters can model supra-τ_c motions. An suitable example is given by Val611$^{\gamma 2}$ and Leu712$^{\delta 2}$ (Fig. 4). The correlation function for this spin system was calculated from the trajectory with force field DES-Amber-SF1.0 using Eq. 8. $^1H - ^1H$ correlation functions are typically characterized by a sharp initial drop due to the rapid rotation of the methyl group about its symmetry axis followed by a decay that converges to a stationary plateau value, around which it oscillates. By definition [48], this plateau value is the order parameter S^2 (Eq. 7). In Fig. 3A, it can be seen that state 1 and state 2 each have two distinct plateaus at 0.46 and 0.63, respectively, showing that the internal correlation is indeed different for the two clusters. Furthermore, we see that each of the two processes converges rapidly after the initial drop. The S^2 order parameters of the two states (shown in dashes) were calculated according to Eq. 10. In stark contrast, when the two clusters are not considered and the whole trajectory is averaged, the correlation function does not converge to a stationary value. Furthermore, we show the correlation function calculated from the hidden Markov model of the system for comparison, calculated using Eqs. (19)–(21). In theory, direct calculation of NOEs from correlation functions should always yield the correct result [75]. A consistent MSM should then accurately reproduce the results of the direct method. However, due to imperfect sampling, statistical noise, and finite trajectory length, the time correlation functions obtained from MD simulations may not converge for slow processes. In particular, we expect in the direct method large statistical errors for dynamic processes of the order of $\approx 1/10$ of the trajectory length and slower. As can be seen in panel B and C, only six transitions between the two states were observed within the trajectory, which evidently represents an undersampling of the process. Furthermore, the two states are distinctly separated with effective distances r_{eff} of 3.6 and 6.2 Å, respectively. MSMs can therefore add a layer of interpretability to the relaxation processes of the system. With our method, we can show that for cases where we can deconvolute discrete states of fast-exchanging subensembles (such as the one shown here), we can use MSMs and the formalism derived here to calculate NOE build-ups from MD trajectories. Furthermore, we want to point out that fitting the time correlation function of the whole trajectory using, for example, the extension of the Lipari-Szabo formalism by [76] to obtain an order parameter of 'fast' (S_f^2) and slow (S_s^2) motion would not lead to a valid result. Here, we show how through a simple deconvolution of the two motional states that are in supra-τ_c exchange, it is possible to obtain dynamics parameters that accurately reflect the underlying conformational changes. The time correlations of all systems outlined in this study are provided in the appendix.

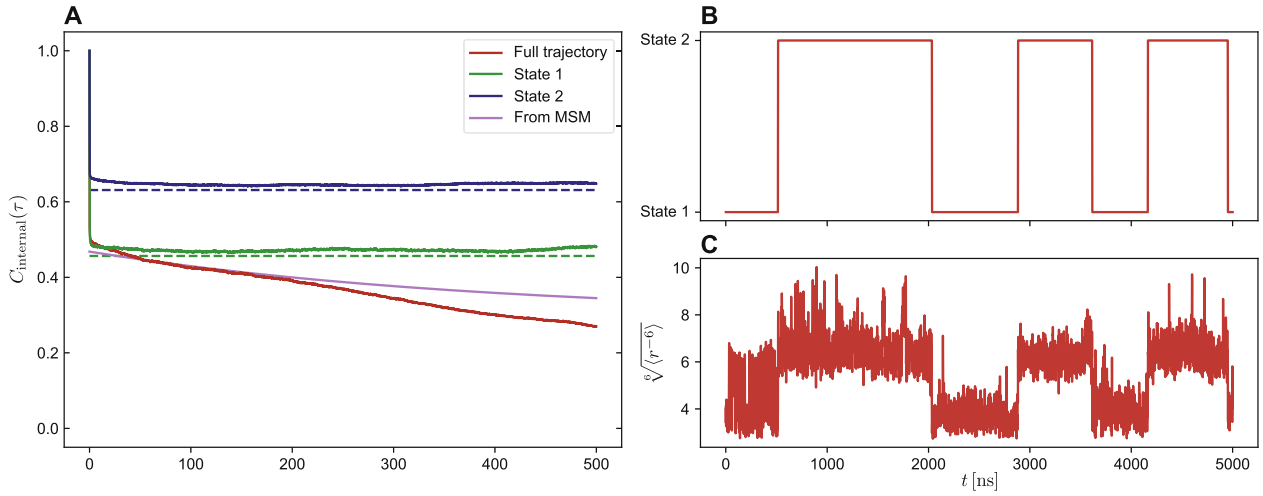

Fig. 3. Time correlation function between Val611$^{\gamma 2}$ and Leu712$^{\delta 2}$. **A**: Separating the two states, it can be seen that the time correlation function for the respective states reaches a plateau at 0.46 (green) and 0.63 (blue). Those correspond to the S^2 order parameter of the two states (shown in dashes). The full trajectory significantly underestimates the correlation between the two ^{1}H–^{1}H vectors and does not reach a plateau value. The trajectories were averaged over 500 ns. Correlation functions calculated from MSMs directly are shown in purple. **B**: The transitions between the two states as a function of trajectory length. In total, six transitions can be observed. **C**: The effective distance, i.e., $\sqrt[6]{\langle r^{-6}\rangle}$ as a function of the trajectory length. The two states have distinct effective distances of of 3.6 and 6.2 Å, respectively.

3.2. Effect of supra-τ_c dynamics on spin diffusion

In a next step, we show how supra-τ_c conformational exchange affects the magnetization pathways and hence the cross-relaxation rates that are mediated by spin diffusion. In Fig. 4, we show two examples of how, depending on the local geometry, NOESY cross-peak intensities can either be enhanced or diminished by the indirect transfer of magnetization. Panel A shows the system Ile268$^{\delta 1}$–Val118$^{\gamma 2}$–Met508$^{\varepsilon}$, where Ile268 experiences a rotamer change on the supra-τ_c timescale (ITS plot is provided in the supplementary material, figure S2). Since the direct magnetization transfer between Ile268$^{\delta 1}$ and Met508$^{\varepsilon}$ is less efficient than the transfer via Val118$^{\gamma 2}$, the resulting cross-peak includes spin diffusion. Separating the slow states, however, leads to increased cross-peak intensity compared to sub-τ_c averaging because of two reasons. Firstly, the effective distances of the states are 7.1 and 8.8 Å, respectively compared to 7.7 Å for the fast averaging. Secondly, state two is in greater proximity to Val118 than state 1, leading to

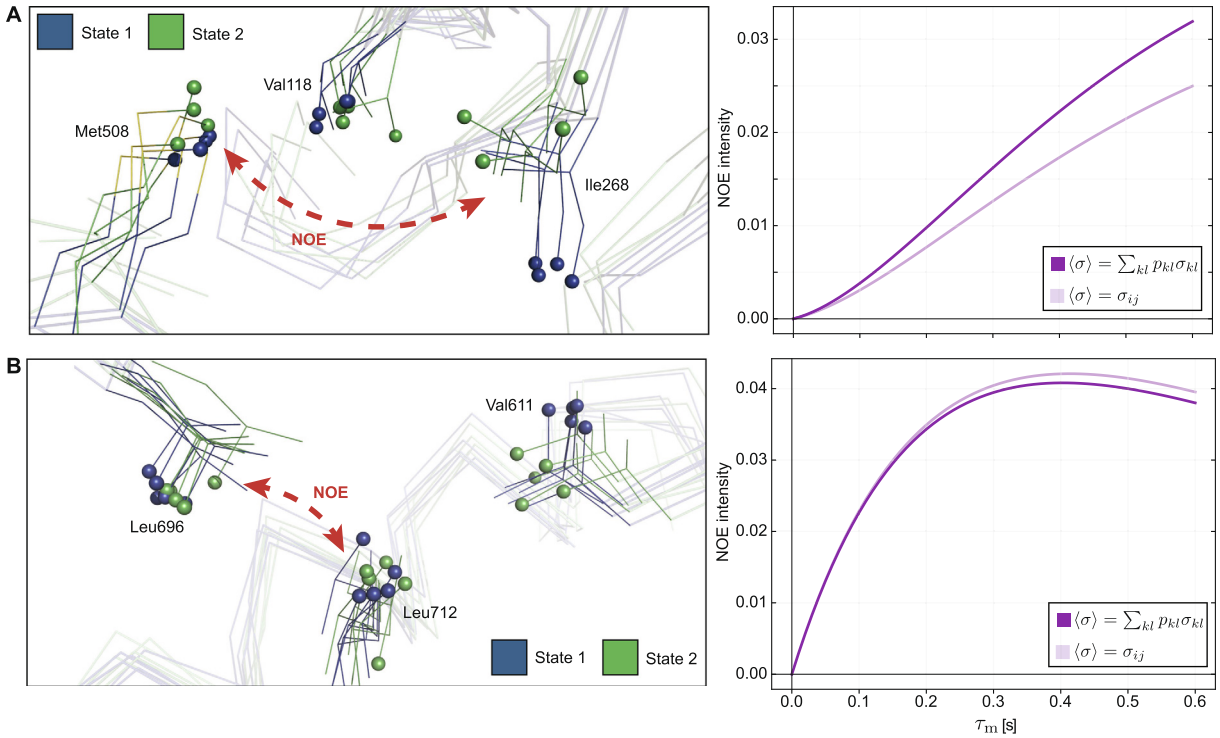

Fig. 4. Different effects of supra-τ_c dynamics on spin diffusion. Two examples of local systems, where cross-relaxation is mediated by spin diffusion, are shown. **A**: System of Val118$^{\gamma 2}$–Ile268$^{\delta 1}$–Met508$^{\varepsilon}$. The cross-peak build-up between Ile268$^{\delta 1}$ and Met508$^{\varepsilon}$ is shown. **B**: System Val611$^{\gamma 2}$–Leu712$^{\delta 2}$–Leu696$^{\delta 2}$ with the NOE build-up between Leu696$^{\delta 2}$ and Leu712$^{\delta 2}$ is shown.

an increased spin diffusion effect. The second example (Fig. 4B) shows the system Val611$^{\gamma 2}$–Leu712$^{\delta 2}$–Leu696$^{\delta 2}$, where Val611 is the residue that experiences a rotameric exchange on the supra-τ_c timescale. The example now highlights the cross-peak between Leu696$^{\delta 2}$ and Leu712$^{\delta 2}$. Albeit the difference being only minor in this example, it illustrates that the discrimination between slowly exchanging states can also lead to a decrease in cross-peak intensity. The decrease comes from the fact that, while the presence of Leu712$^{\delta 2}$ enhances the signal between Leu696$^{\delta 2}$ and Val611$^{\gamma 2}$ (shown in the supplement figure S4), it is this increase in signal intensity that reduces the amount of magnetization that is exchanged between Leu696$^{\delta 2}$ and Leu712$^{\delta 2}$. The justification for the supra-τ_c treatment of Val611$^{\gamma 2}$–Leu712$^{\delta 2}$ is shown in the supplementary material (figure S3). This example shows that for a quantitative analysis of NOESY spectra, not only is it necessary to account for all magnetization pathways by means of full relaxation matrix analysis, it is also important to consider that supra-τ_c exchange has an influence in how magnetization propagates through the spin system.

4. Conclusion

Here, we have studied the effects of supra-τ_c dynamics on the build-up of NOE cross-peaks in methyl-labelled proteins. Using MD simulations and Markov state models, we could identify conformational exchange on the timescale well above the rotational correlation time τ_c of MSG (37 ns). We find that the supra-τ_c motions are not always decorrelated from sub-τ_c motions, such that they can be treated separately, but that there exist convolutions of supra- and sub-τ_c motions in a way that requires their joint analysis.

In particular, such analysis becomes essential if the correlation functions of the slow processes do not reach a stationary point during the course of the trajectory. This criterion is regularly met in contemporary simulations of typical systems, as shown in this work. If supra-τ_c exchange can be deconvoluted into discrete states, one can obtain state-specific order parameters and averaged distances, which can then be used to calculate the cross-relaxation rate.

The work thus highlights the importance of considering molecular motions in the supra-τ_c regime and extends the theory of modeling spin relaxation to that timescale. In addition, our contribution also adds to the repertoire of methods to parameterize MD force fields, given the high sensitivity of cross-relaxation rates to minor structural changes. We envision improving MD force-fields by adjusting the force field parameters to match the experimental observables. Alternatively, one might adjust inaccurate simulation data to match experimental observables *post hoc [63]*.

Our method provides a way to model the cross-relaxation rate between methyl groups that cluster in metastable states and exchange on the supra-τ_c timescale and thus contributes to a quantitative description of NOE cross-peaks.

5. Materials and Methods

5.1. Molecular dynamics simulations

The initial structure was obtained from [77] with 723 residues. SWISS-MODEL [78] was used to model the residues missing from the crystal structure. Protein hydrogen atoms were added according to neutral pH with one of the five different force fields (FFs): a99SB-disp [64], CHARMM36m [65], Amber99SB*-ILDN-q [66], DES-Amber and DES-Amber-SF1.0 [67]. In the case of DES-Amber and DES-Amber-SF1.0, the TIP4P-D [79] water model used while a99SB-disp employed TIP4P-D with modified protein and water van der Waals interaction terms [64]. For Amber99SB*-ILDN-q, two water models were tested, either TIP3P or TIP4P [80]. For, CHARMM36m, TIP3P was used. In total there were six systems studied. For each system, the protein was placed in the center of dodecahedron box with a 11 nm edge. Approximately 31,000 water molecules were added to each box. The overall charge of each system was neutralized by the addition of $\approx$ 150 mM sodium chloride. Energy minimization was performed using steepest descents for $\leqslant 5,000$ steps with a 0.01 nm step size. Equilibrations in the NVT and NPT ensembles were performed for $\approx$ 0.01 µs in total with position restraints applied to protein backbone atoms, while no restraints were applied to the solvent. All unrestrained production simulations were run in the NPT ensemble starting from different velocities and were performed for 5 µs each, using GROMACS2018 [81]. For all FFs, equations of motion were integrated via the Verlet leapfrog algorithm with a 2 fs time step. All bonds connected to hydrogens were constrained with the LINCS algorithm. The cutoff distance was 1.2 nm in case of CHARMM36m and 1.0 nm in case all Amber-based FFs, for the short-range neighbor list and van der Waals interactions. The Particle Mesh Ewald method [82] was applied for long-range electrostatic interactions with a 1.2 and 1.0 nm real space cutoff for CHARMM36m FF and Amber based FFs respectively. The velocity rescaling thermostat was used to maintain the temperature at 310 K. Pressure was maintained at 1 bar using the Parinello-Rahman [83] barostat. Simulations were performed on an in-house Linux cluster composed of 8 nodes containing 2 GPUs (Nvidia GeForce RTX 2080 Ti) and 24 CPUs (Intel® Xeon® Gold 5118 CPU @ 2.3 GHz) each.

5.2. Markov state models

Local Markov state models (MSMs) of methyl–methyl contacts of MSG were estimated to discriminate between sub- and supra-τ_c kinetic exchange. In this work, the following methyl groups were considered: Ala$^{\beta 1}$, Ile$^{\delta 1}$, Leu$^{\delta 2}$, Val$^{\gamma 2}$, Thr$^{\gamma 2}$, Met$^{\varepsilon}$ (AILVTM), a total of 288 methyl groups. MSMs were then build for all methyl–methyl pairs that are below 10 Å, resulting in a total of $\approx 1,500$ MSMs per force field. Features included side chain and main chain torsion angles of both residues as well as the distances between them. We then used time-lagged independent component analysis (tICA) [84,71] with a lag time of 1 ns to reduce dimensionality to the two slowest independent components. Clustering was then performed in the reduced space using the *k*-means algorithm with 100 centers. Bayesian hidden Markov state models (hMSMs) were estimated for 2 to 3 hidden states depending on the amount of processes that are $>10 \times \tau_c$ with a lag time of 1 ns. Implied time scale and tICA plots of the slowest methyl–methyl pairs for the respective force fields can be explored using the dashboard and database provided on GitHub (see Section 5.5). All analyses were done in Python using MDAnalysis (v. 2.0.0) [56,85] and PyEMMA (v. 2.5.7) [60].

5.3. NOESY build-up calculation

For the calculation of NOESY intensities, first, the dynamics parameters ($\langle r^{-6} \rangle$, S^2, and weight p of each state) were calculated for all AILVTM combinations following the calculations outlined in chapter 2. Subsequently, **R** was obtained by calculating the auto- and cross-relaxation rates using the pairwise-calculated dynamics parameters mentioned above. Calculations assumed τ_c of 37 ns [69], as well as the experimentally determined methyl rotational correlation times τ_{internal} determined by [52]. All calculations were done using an in-house written software implemented in Python.

5.4. Visualization

All figures displaying molecular structures were made with PyMol [86]. Plots were generated using Matplotlib [87].

5.5. Data availability

A dashboard, designed using Dash by Plotly [88] and written in Python, is available on GitHub for visualizing all correlation functions, ITS and tICA plots as well as NOESY build-up curves for all force fields of all methyl–methyl pairs within 5 Å. For systems that are in supra-τ_c exchange, data are provided for contacts of $\leqslant 10$ Å of distance. The dashboard can be found under the following link.

The processed MD data used for MSM estimation as well as all NOESY build-ups are available on Mendeley Data and can be freely accessed under the following link (DOI: 10.17632/d9r96x4c9x.3).

CRediT authorship contribution statement

Christopher Kolloff: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – Original Draft, Visualization. **Adam Mazur:** Conceptualization, Methodology, Supervision, Writing – Review & Editing, Validation, Project administration. **Jan K. Marzinek:** Data curation, Writing – Original Draft, Visualization. **Peter J. Bond:** Writing – Review & Editing, Funding acquisition. **Simon Olsson:** Supervision, Conceptualization, Writing – Review & Editing, Funding acquisition. **Sebastian Hiller:** Conceptualization, Resources, Writing – Review & Editing, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Swiss National Science Foundation (grant 185388 to S.H.) as well as by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation (C.K. and S.O.). We also acknowledge BII (A*STAR) core funding. Finally, the authors would like to thank Dr. Guillaume Mas and Dr. Thomas Müntener for providing their perspective and for the interesting discussions. This work is dedicated to Prof. Kurt Wüthrich to acknowledge his exceptional and unparalleled contributions to the field of biomolecular NMR spectroscopy. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Appendix A

A.1. Correlation function and supra-τ_c exchange

Here we show how, similar to [48,76], the correlation function can be expressed in terms of a sum of exponentials of exchange processes. The total correlation C^{tot} of a spin system in an molecule undergoing isotropic Brownian motion can be factored as the product between the overall correlation $C^{\text{overall}}(t)$ and the internal correlation C^{internal}:

$$C^{\text{tot}}(t) = C^{\text{overall}}(t)\langle r^{-6}\rangle C^{\text{internal}}(t). \tag{22}$$

Following the rationale of [48]:

$$\begin{aligned} C^{\text{tot}}(t) &= C^{\text{overall}}(t)\sum_{i=1}^{n}\langle r_i^{-6}\rangle C_i^{\text{internal}}(t) \\ &= C^{\text{overall}}(t)\sum_{i=1}^{n}\langle r_i^{-6}\rangle p_i \underbrace{\exp\left(-\frac{t}{\tau_{\text{ex},i}}\right)}_{\lim_{t\to\infty} C_i^{\text{internal}} := S_i^2}, \end{aligned} \tag{23}$$

with p being the weight and $\sum p = 1, \tau_{\text{ex}}$ being the exchange time constant, and S^2 being the order parameter of state i. From Eq. 23, it follows that

$$\begin{aligned} J(\omega) &= \int_0^\infty C^{\text{tot}}(t)\cos(\omega t)\,\mathrm{d}t \\ &= \int_0^\infty \underbrace{C^{\text{overall}}(t)}_{\exp\left(-\frac{t}{\tau_c}\right)} \sum_{i=1}^{n}\langle r_i^{-6}\rangle \underbrace{C_i^{\text{internal}}(t)}_{S_i^2} \cos(\omega t)\,\mathrm{d}t \\ &\approx \sum_{i=1}^{n}\langle r_i^{-6}\rangle p_i S_i^2 \frac{\tau_c}{1+\omega^2\tau_c^2}. \end{aligned} \tag{24}$$

A.2. Autorelaxation rate

The autorelaxation rate constant ρ for a methyl moiety consists of the following contributions:

1. intramethyl relaxation,
   (a) proton–proton relaxation (ρ^{HH})
   (b) proton–carbon relaxation (ρ^{HC})
2. intermethyl relaxation (ρ^{inter}),
3. external relaxation (ρ^{ext}),

where ρ^{ext} accounts for other relaxation mechanisms, e.g., intermolecular dipole–dipole relaxation caused by dissolved oxygen or other paramagnetic species present in the sample, chemical shift anisotropy, etc.) It follows that ρ is given by the linear combination of the different terms [89]:

$$\rho = \rho^{\text{HH}} + \rho^{\text{HC}} + \rho^{\text{inter}} + \rho^{\text{ext}}. \tag{25}$$

A.2.1. Intramethyl relaxation

The close proximity of the three protons in the methyl group causes them to relax, adding to the diagonal entries of **R** [90]. The spectral density function for two intramethyl protons, J^{intra} is given by [49,90,40]:

$$J^{\text{intra}}(\omega) = \frac{1}{r_{\text{methyl}}^6}\left(\frac{1}{4}\frac{\tau_c}{1+\omega^2\tau_c} + \frac{3}{4}\frac{\tau_e}{1+\omega^2\tau_e}\right), \tag{26}$$

where

$$\frac{1}{\tau_e} = \frac{1}{\tau_c} + \frac{1}{\tau_{\text{internal}}}.$$

τ_{internal} and r_{methyl} are the internal rotational correlation time and the proton–proton distance of a methyl group 1.7 Å, respectively [49].

Consider the three protons of a methyl group: p, q and r. The auto- and cross-relaxation terms of the protons within the methyl group can be described as:

$$\boldsymbol{\rho}^{\text{HH}} = \begin{pmatrix} 2\rho_{pp} & \sigma_{pq} & \sigma_{pr} \\ \sigma_{qp} & 2\rho_{qq} & \sigma_{qr} \\ \sigma_{rp} & \sigma_{rq} & 2\rho_{rr} \end{pmatrix}.$$

Due to the chemical equivalence of the three protons, i.e., $p = q = r$, the intramethyl proton–proton relaxation can be considered as the average, reducing $\boldsymbol{\rho}_i^{\text{HH}}$ to the scalar quantity

$$\begin{aligned}\rho^{\mathrm{HH}} &= \frac{6\rho + 6\sigma}{3} \\ &= 2\rho + 2\sigma.\end{aligned} \tag{27}$$

Thus, the proton–proton relaxation contribution is given by:

$$\begin{aligned}\rho^{\mathrm{HH}} &= 2\rho + 2\sigma \\ &= \tfrac{1}{5}K_{\mathrm{HH}}^2 \cdot \left[6J^{\mathrm{intra}}(2\omega_{\mathrm{H}}) + 3J^{\mathrm{intra}}(\omega_{\mathrm{H}}) + J^{\mathrm{intra}}(0)\right] \\ &\quad + \tfrac{1}{5}K_{\mathrm{HH}}^2 \cdot \left[6J^{\mathrm{intra}}(2\omega_{\mathrm{H}}) - J^{\mathrm{intra}}(0)\right] \\ &= \tfrac{1}{5}K_{\mathrm{HH}}^2 \cdot \left[12J^{\mathrm{intra}}(2\omega_{\mathrm{H}}) + 3J^{\mathrm{intra}}(\omega_{\mathrm{H}})\right].\end{aligned} \tag{28}$$

The second term in Eq. 25 accounts for the proton–carbon relaxation:

$$\rho^{\mathrm{HC}} = \frac{1}{10}K_{\mathrm{HC}}^2 \cdot [6J(\omega_{\mathrm{H}} + \omega_{\mathrm{C}}) + 3J(\omega_{\mathrm{H}}) + J(\omega_{\mathrm{H}} - \omega_{\mathrm{C}})], \tag{29}$$

where the distance in J is r_{HC}, the proton–carbon distance that corresponds to 1.09 Å.

A.2.2. Intermethyl relaxation

Additional to the intramethyl relaxation terms, all other j protons in the surrounding methyl groups k also contribute to the autorelaxation:

$$\rho_{ij}^{\mathrm{inter}} = \sum_{k,k\neq i}\sum_{j=1}^{3}\sum_{i=1}^{3}\frac{1}{10}K_{\mathrm{HH}}^2 \cdot \left[6J_{ij}(2\omega_{\mathrm{H}}) + 3J_{ij}(\omega_{\mathrm{H}}) + J_{ij}(0)\right]. \tag{30}$$

Appendix B. Supplementary Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jmr.2022.107196.

References

[1] I. Pritišanac, J.M. Würz, T.R. Alderson, P. Güntert, Automatic structure-based NMR methyl resonance assignment in large proteins, Nat. Commun. 10 (2019) 4922.
[2] P. Güntert, C. Mumenthaler, K. Wüthrich, Torsion angle dynamics for NMR structure calculation with the new program DYANA, J. Mol. Biol. 273 (1997) 283–298.
[3] Y.R. Monneau, P. Rossi, A. Bhaumik, C. Huang, Y. Jiang, T. Saleh, T. Xie, Q. Xing, C.G. Kalodimos, Automatic methyl assignment in large proteins by the MAGIC algorithm, J. Biomol. NMR 69 (2017) 215–227.
[4] S. Nerli, V.S. De Paula, A.C. McShan, N.G. Sgourakis, Backbone-independent NMR resonance assignments of methyl probes in large proteins, Nat. Commun. 12 (2021).
[5] K. Wüthrich, Nobel lecture: NMR studies of structure and function of biological macromolecules, Biosci. Rep. 23 (2003) 119–168.
[6] B. Vögeli, P. Güntert, R. Riek, Multiple-state ensemble structure determination from eNOE spectroscopy, Mol. Phys. 111 (2013) 437–454.
[7] B. Vögeli, S. Olsson, P. Güntert, R. Riek, The Exact NOE as an Alternative in Ensemble Structure Determination, Biophys. J . 110 (2016) 113–126.
[8] C.N. Chi, D. Strotz, R. Riek, B. Vögeli, NOE-Derived Methyl Distances from a 360 kDa Proteasome Complex, Chem. - A Eur. J. 24 (2018) 2270–2276.
[9] B.R. Leeflang, L.M. Kroon-Batenburg, CROSREL: Full relaxation matrix analysis for NOESY and ROESY NMR spectroscopy, J. Biomol. NMR 2 (1992) 495–518.
[10] J. Orts, B. Vögeli, R. Riek, Relaxation matrix analysis of spin diffusion for the NMR structure calculation with eNOEs, J. Chem. Theory Comput. 8 (2012) 3483–3492.
[11] J.E. Ollerenshaw, V. Tugarinov, L.E. Kay, Methyl TROSY: Explanation and experimental verification, Magn. Reson. Chem. 41 (2003) 843–852.
[12] C. Zwahlen, K.H. Gardner, S.P. Sarma, D.A. Horita, R.A. Byrd, L.E. Kay, An NMR experiment for measuring methyl-methyl NOEs in 13C-labeled proteins with high resolution, J. Am. Chem. Soc. 120 (1998) 7617–7625.
[13] V. Tugarinov, R. Muhandiram, A. Ayed, L.E. Kay, Four-dimensional NMR spectroscopy of a 723-residue protein: Chemical shift assignments and secondary structure of malate synthase G, J. Am. Chem. Soc. 124 (2002) 10025–10035.
[14] D.M. Korzhnev, K. Kloiber, V. Kanelis, V. Tugarinov, L.E. Kay, Probing Slow Dynamics in High Molecular Weight Proteins by Methyl-TROSY NMR Spectroscopy: Application to a 723-Residue Enzyme, J. Am. Chem. Soc. 126 (2004) 3964–3973.
[15] K.H. Gardner, L.E. Kay, Production and incorporation of 15N, 13C, 2H (1H-δ1 methyl) isoleucine into proteins for multidimensional NMR studies, J. Am. Chem. Soc. 119 (1997) 7599–7600.
[16] I. Ayala, R. Sounier, N. Usé, P. Gans, J. Boisbouvier, An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein, J. Biomol. NMR 43 (2009) 111–119.
[17] R. Kerfah, M.J. Plevin, R. Sounier, P. Gans, J. Boisbouvier, Methyl-specific isotopic labeling: A molecular tool box for solution NMR studies of large proteins, Curr. Opin. Struct. Biol. 32 (2015) 113–122.
[18] P. Rossi, Y.R. Monneau, Y. Xia, Y. Ishida, C.G. Kalodimos, Toolkit for NMR Studies of Methyl-Labeled Proteins, volume 614, 1 ed., Elsevier Inc., 2019. doi: 10.1016/bs.mie.2018.08.036. doi: 10.1016/bs.mie.2018.08.036.
[19] I. Ayala, L. Chiari, R. Kerfah, J. Boisbouvier, P. Gans, Asymmetric Synthesis of Methyl Specifically Labelled L - Threonine and Application to the NMR Studies of High Molecular Weight Proteins, ChemistrySelect 5 (2020) 5092–5098.
[20] G. Mas, E. Crublet, O. Hamelin, P. Gans, J. Boisbouvier, Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins, J. Biomol. NMR 57 (2013) 251–262.
[21] G. Mas, J.Y. Guan, E. Crublet, E.C. Debled, C. Moriscot, P. Gans, G. Schoehn, P. Macek, P. Schanda, J. Boisbouvier, Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle, Sci. Adv. 4 (2018) 1–9.
[22] V. Tugarinov, W.Y. Choy, V.Y. Orekhov, L.E. Kay, Solution NMR-derived global fold of a monomeric 82-kDa enzyme, Proc. Natl. Acad. Sci. USA 102 (2005) 622–627.
[23] R. Sprangers, L.E. Kay, Quantitative dynamics and binding studies of the 20S proteasome by NMR, Nature 445 (2007) 618–622.
[24] Z.K. Boswell, M.P. Latham, Methyl-Based NMR Spectroscopy Methods for Uncovering Structural Dynamics in Large Proteins and Protein Complexes, Biochemistry 58 (2019) 144–155.
[25] S. Hiller, Chaperone-Bound Clients: The Importance of Being Dynamic, Trends Biochem. Sci. 44 (2019) 517–527.
[26] B.M. Burmann, C. Wang, S. Hiller, Conformation and dynamics of the periplasmic membrane-protein-chaperone complexes OmpX-Skp and tOmpA-Skp, Nat. Struct. Mol. Biol. 20 (2013) 1265–1272.
[27] D. Thirumalai, S.A. Woodson, Kinetics of Folding of Proteins and RNA, Acc. Chem. Res. 29 (1996) 433–439.
[28] H. Lei, C. Wu, H. Liu, Y. Duan, Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations, Proc. Natl. Acad. Sci. USA 104 (2007) 4925–4930.
[29] Y. Duan, P.A. Kollman, Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, Science 282 (1998) 740–744.
[30] C.D. Snow, H. Nguyen, V.S. Pande, M. Gruebele, Absolute comparison of simulated and experimental protein-folding dynamics, Nature 420 (2002) 102–106.
[31] G. Cecconi, E.A. Shank, C. Bustamante, S. Marqusee, Direct observation of the three-state folding of a single protein molecule, Science 309 (2005) 2057–2060.
[32] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, Proteins: Struct. Funct. Genet. 65 (2006) 712–725.
[33] S.A. Showalter, R. Brüschweiler, Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB force field, J. Chem. Theory Comput. 3 (2007) 961–975.
[34] M. Buck, S. Bouguet-Bonnet, R.W. Pastor, A.D. MacKerell, Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme, Biophys. J. 90 (2006) L36–L38.
[35] E.T. Olejniczak, C.M. Dobson, M. Karplus, R.M. Levy, Motional Averaging of Proton Nuclear Overhauser Effects in Proteins. Predictions from a Molecular Dynamics Simulation of Lysozyme, J. Am. Chem. Soc. 106 (1984) 1923–1930.
[36] R. Brüschweiler, B. Roux, M. Blackledge, C. Griesinger, M. Karplus, R.R. Ernst, Influence of Rapid Intramolecular Motion on NMR Cross-Relaxation Rates. A Molecular Dynamics Study of Antamanide in Solution, J. Am. Chem. Soc. 114 (1992) 2289–2302.
[37] R. Brüschweiler, D.A. Case, Characterization of biomolecular structure and dynamics by NMR cross relaxation, Prog. Nucl. Magn. Reson. Spectrosc. 26 (1994) 27–58.
[38] A. Abragam, The principles of nuclear magnetism, 1961.
[39] I. Solomon, Relaxation Processes in a System of Two Spins.pdf, Phys. Rev. 99 (1955) 559–565.
[40] D.E. Woessner, Spin Relaxation Processes in a Two- Proton System Undergoing Anisotropic Reorientation, J. Chem. Phys. 36 (1962).
[41] J. Tropp, Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: The effect of fluctuating internuclear distances, J. Chem. Phys. 72 (1980) 6035–6043.
[42] S. Macura, R.R. Ernst, Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy, Mol. Phys. 41 (1980) 95–117.
[43] B. Vögeli, The nuclear Overhauser effect from a quantitative perspective, Prog. Nucl. Magn. Reson. Spectrosc. 78 (2014) 1–46.
[44] T.L. Religa, R. Sprangers, L.E. Kay, Dynamic Regulation of Archaeal Proteasome Gate Opening As Studied by TROSY NMR, Science 328 (2010) 98–103.
[45] R. Sounier, L. Blanchard, Z. Wu, J. Boisbouvier, High-accuracy distance measurement between remote methyls in specifically protonated proteins, J. Am. Chem. Soc. 129 (2007) 472–473.

[46] I. Pritišanac, M.T. Degiacomi, T.R. Alderson, M.G. Carneiro, E. Ab, G. Siegal, A.J. Baldwin, Automatic Assignment of Methyl-NMR Spectra of Supramolecular Machines Using Graph Theory, J. Am. Chem. Soc. 139 (2017) 9523–9533.
[47] I. Pritišanac, T.R. Alderson, P. Güntert, Automated assignment of methyl NMR spectra from large proteins, Prog. Nucl. Magn. Reson. Spectrosc. 118–119 (2020) 54–73.
[48] G. Lipari, A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity, J. Am. Chem. Soc. 104 (1982) 4546–4559.
[49] C. Peter, X. Daura, W.F. van Gunsteren, Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations, J. Biomol. NMR 20 (2001) 297–310.
[50] E. Andrew, R. Gaspar, W. Vennart, Proton magnetic resonance investigation of solid polyamino acids, Chem. Phys. Lett. 38 (1976) 141–143.
[51] D.R. Muhandiram, T. Yamazaki, L.E. Kay, B.D. Sykes, Measurement of 2H T1 and T1ρRelaxation Times in Uniformly 13C-Labeled and Fractionally 2H-Labeled Proteins in Solution, J. Am. Chem. Soc. 117 (1995) 11536–11544.
[52] V. Tugarinov, L.E. Kay, Quantitative 13C and 2H NMR relaxation studies of the 723-residue enzyme malate synthase G reveal a dynamic binding interface, Biochemistry 44 (2005) 15970–15977.
[53] P. Luginbühl, K. Wüthrich, Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules, Prog. Nucl. Magn. Reson. Spectrosc. 40 (2002) 199–247.
[54] S. Olsson, F. Noé, Mechanistic models of chemical exchange induced relaxation in protein NMR, J. Am. Chem. Soc. 139 (2017) 200–210.
[55] F. Noé, S. Doose, I. Daidone, M. Löllmann, M. Sauer, J.D. Chodera, J.C. Smith, Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments, Proc. Natl. Acad. Sci. USA 108 (2011) 4822–4827.
[56] R. Gowers, M. Linke, J. Barnoud, T. Reddy, M. Melo, S. Seyler, J. Domański, D. Dotson, S. Buchoux, I. Kenney, O. Beckstein, MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations, in: Proceedings of the 15th Python in Science Conference, 2016, pp. 98–105.
[57] A.-R. Allouche, Software News and Updates Gabedit – A Graphical User Interface for Computational Chemistry Softwares, J. Comput. Chem. 32 (2012) 174–182.
[58] R.T. McGibbon, K.A. Beauchamp, M.P. Harrigan, C. Klein, J.M. Swails, C.X. Hernández, C.R. Schwantes, L.P. Wang, T.J. Lane, V.S. Pande, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories, Biophys. J. 109 (2015) 1528–1532.
[59] C. Schütte, A. Fischer, W. Huisinga, P. Deuflhard, A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo, J. Comput. Phys. 151 (1999) 146–168.
[60] M.K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.H. Prinz, F. Noé, PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models, J. Chem. Theory Comput. 11 (2015) 5525–5542.
[61] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J.D. Chodera, C. Schütte, F. Noé, Markov models of molecular kinetics: Generation and validation, J. Chem. Phys. 134 (2011) 174105.
[62] F. Noé, H. Wu, J.H. Prinz, N. Plattner, Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules, J. Chem. Phys. 139 (2013).
[63] S. Olsson, H. Wu, F. Paul, C. Clementi, F. Noé, Combining experimental and simulation data of molecular processes via augmented Markov models, Proc. Natl. Acad. Sci. USA 114 (2017) 8265–8270.
[64] P. Robustelli, S. Piana, D.E. Shaw, Developing a molecular dynamics force field for both folded and disordered protein states, Proc. Natl. Acad. Sci. USA 115 (2018) E4758–E4766.
[65] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. De Groot, H. Grubmüller, A.D. MacKerell, CHARMM36m: An improved force field for folded and intrinsically disordered proteins, Nat. Methods 14 (2016) 71–73.
[66] R.B. Best, D. de Sancho, J. Mittal, Residue-specific α-helix propensities from molecular simulation, Biophys. J. 102 (2012) 1462–1467.
[67] S. Piana, P. Robustelli, D. Tan, S. Chen, D.E. Shaw, Development of a Force Field for the Simulation of Single-Chain Proteins and Protein-Protein Complexes, J. Chem. Theory Comput. 16 (2020) 2494–2507.
[68] S. Olsson, F. Noé, Dynamic graphical models of molecular kinetics, Proc. Natl. Acad. Sci. USA 116 (2019) 15001–15006.
[69] V. Tugarinov, L.E. Kay, Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods, J. Am. Chem. Soc. 125 (2003) 13868–13878.
[70] L. Molgedey, H.G. Schuster, Separation of a mixture of independent signals using time delayed correlations, Phys. Rev. Lett. 72 (1994) 3634–3637.
[71] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov model construction, J. Chem. Phys. 139 (2013).
[72] F. Noé, H. Wu, J.H. Prinz, N. Plattner, Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules, J. Chem. Phys. 139 (2013) 1–17.
[73] V. Tugarinov, L.E. Kay, Stereospecific NMR assignments of prochiral methyls, rotameric states and dynamics of valine residues in malate synthase G, J. Am. Chem. Soc. 126 (2004) 9827–9836.
[74] V. Tugarinov, J.E. Ollerenshaw, L.E. Kay, Probing side-chain dynamics in high molecular weight proteins by deuterium NMR spin relaxation: An application to an 82-kDa enzyme, J. Am. Chem. Soc. 127 (2005) 8214–8225.
[75] G. Chalmers, J.N. Glushka, B.L. Foley, R.J. Woods, J.H. Prestegard, Direct NOE simulation from long MD trajectories, J. Magn. Reson. 265 (2016) 1–9.
[76] G.M. Clore, A. Szabo, A. Bax, L.E. Kay, P.C. Driscoll, A.M. Gronenborn, Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins, J. Am. Chem. Soc. 112 (1990) 4989–4991.
[77] B.R. Howard, J.A. Endrizzi, S.J. Remington, Crystal structure of Escherichia coli malate synthase G complexed with magnesium and glyoxylate at 2.0 Å resolution: Mechanistic implications, Biochemistry 39 (2000) 3156–3168.
[78] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: Homology modelling of protein structures and complexes, Nucl. Acids Res. 46 (2018) W296–W303.
[79] S. Piana, A.G. Donchev, P. Robustelli, D.E. Shaw, Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States, J. Phys. Chem. B 119 (2015) 5113–5123.
[80] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[81] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: Fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.
[82] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.
[83] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, J. Appl. Phys. 52 (1981) 7182–7190.
[84] C.R. Schwantes, V.S. Pande, Improvements in Markov State Model construction reveal many non-native interactions in the folding of NTL9, J. Chem. Theory Comput. 9 (2013) 2000–2009.
[85] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, O. Beckstein, MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations, J. Comput. Chem. 32 (2011) 174–182.
[86] L.L.C. Schrodinger, The PyMOL Molecular Graphics System, Version 1 (2015) 8.
[87] J.D. Hunter, MATPLOTLIB: A 2D GRAPHICS ENVIRONMENT, Comput. Sci. Eng. 9 (2007) 90–95.
[88] Plotly Technologies Inc, Plotly, 2015. URL https://plot.ly.
[89] L. Zhu, H.J. Dyson, P.E. Wright, A NOESY-HSQC simulation program, SPIRIT, J. Biomol. NMR 11 (1998) 17–29.
[90] S.P. Edmondson, NOE R factors and structural refinement using FIRM, an iterative relaxation matrix program, J. Magn. Reson. 98 (1992) (1969) 283–298.