

# Rotational dynamics of proteins from spin relaxation rates and molecular dynamics simulations

O. H. Samuli Ollila\*

*Institute of Biotechnology, University of Helsinki*

(Dated: June 19, 2017)

## I. INTRODUCTION

Protein conformational sampling and entropy play a significant role in protein functionality and interactions with other biomolecules. In addition to internal dynamics due to conformation sampling, the proteins also experience overall brownian tumbling. These properties are experimentally accessible through spin relaxation rates of  $^{15}\text{N}$  and  $^{13}\text{C}$  nuclei measured with nuclear magnetic resonance (NMR) techniques [1–3]. Spin relaxation rate experiments have been used to analyze, for example, conformational entropy [? ], binding entropy of ?? [? ], resolve sampled structures [? ] and validating molecular dynamics simulations [? ]. These analyses are almost exclusively based on assumption of separable motions of internal conformational sampling and overall rotational tumbling [? ]. In addition, most studies assume isotropic diffusion for overall tumbling, while inclusion of anisotropic models in the analysis is a significant complication [1? , 2].

Classical molecular dynamics simulation methods are promising tools to interpretate spin relaxation rates from NMR relaxation data for proteins with significant anisotropy or correlated internal and overall rotational motions. , because they simultaneously contain internal rotation, global brownian tumbling, anisotropy and hydration shell effects [? ]. However, practical applications have been limited by the inaccuracies in force field descriptions and available time scales in the simulations. The most used water model in rotational dynamics studies, tip3p [? ], predicts too fast rotational diffusion [? ]. On the other hand, very long simulations are needed to collect enough statistics to calculate rotational correlation functions from single molecules in MD simulations [? ]. Consequently, protein rotational diffusion coefficients are not typically calculated from simulation data. Instead, the calculated relaxation rates are fitted to experiments by assuming an isotropic diffusion model [? ].

In this work we overcome these issues by directly calculating diffusion coefficients from protein inertia axes. Diffusion coefficients can be then used to determine global rotational correlation functions from shorter simulations than with direct calculation. Furthermore, the overestimated dynamics due to water model can be anisotropically corrected by scaling the diffusion coefficients with a constant factor. The usefulness of the approach is demonstrated by interpreting the spin relaxation data from anisotropic protein constructs from *TonB* of *Helicobacter pylori* [? ] and *Pseudomonas* [? ]. These segments are considered as vital parts related to iron transport into Gram-negative bacteria [? ].

## II. METHODS

### A. Spin relaxation and rotational dynamics of molecules

Two different approaches have been commonly used to analyze rotational dynamics from NMR relaxation experiments. In "model free analysis" the fundamental idea is to separate internal dynamics from global rotation and assume exponential forms for rotational correlation functions. The parameters of rotational correlation functions are then fitted the experimental data to solve time scales and order parameters for different dynamical processes [? ]. Alternative approach is to use bead models and hydrodynamical calculations to describe protein dynamics and predict spin relaxation rates [? ]. The approaches have been successful for several proteins, but both suffer from significant limitations which limit the general applicability [? ]. Main practical issue in the "model free analysis" is that the amount of free parameters to be fitted in the experiments becomes large for anisotropic proteins experiencing complex dynamics [? ]. On the other hand, hydrodynamical calculations are sensitive for the assumptions about protein hydration shell [? ].

Spin relaxation data is usually analyzed by describing the sampled bond orientations with order parameter  $S^2$  respect to the protein reference frame and assuming that overall and internal motions are independent [2? ]. Order parameters and timescales for overall and internal motions can be then extracted by fitting various functional forms to spin relaxation data measured with different magnetic field strengths [1, 2].

Practical approaches to analyze molecular dynamics from NMR relaxation data are usually based on the connection between second order rotational correlation function  $C(t)$  of N-H bond and experimentally measured spin relaxation rates  $R_1$ ,  $R_2$  and  $R_{\text{NOE}}$  through Redfield equations [4, 5]

$$R_1 = \frac{d_{\text{NH}}^2 N_{\text{H}}}{20} \left[ J(\omega_{\text{H}} - \omega_{\text{N}}) + 3J(\omega_{\text{N}}) + 6J(\omega_{\text{N}} + \omega_{\text{H}}) \right] + \frac{(\sigma\omega_{\text{N}})^2}{15} j(\omega_{\text{N}}), \quad (1)$$

$$R_2 = \frac{1}{2} \frac{d_{\text{NH}}^2 N_{\text{H}}}{20} \left[ 4J(0) + 3j(\omega_{\text{N}}) + J(\omega_{\text{H}} - \omega_{\text{N}}) + 6J(\omega_{\text{H}}) + 6J(\omega_{\text{N}} + \omega_{\text{H}}) \right] + \frac{(\sigma\omega_{\text{N}})^2}{15 * 6} [4J(0) + 3J(\omega_{\text{N}})], \quad (2)$$

$$R_{\text{NOE}} = 1 + \frac{d_{\text{NH}}^2 N_{\text{H}}}{20} \left[ 6J(\omega_{\text{N}} + \omega_{\text{H}}) + J(\omega_{\text{H}} - \omega_{\text{N}}) \right] \frac{\gamma_{\text{H}}}{\gamma_{\text{N}} R_1}, \quad (3)$$

\* samuli.ollila@helsinki.fi; Department of Neuroscience and Biomedical Engineering, Aalto University

where  $\omega_N$  and  $\omega_H$  are the Larmor angular frequencies of  $^{15}\text{N}$  and  $^1\text{H}$  respectively, and  $N_H$  is the number of bound protons. Spectral density  $J(\omega)$  is the Fourier transformation of the second order rotational correlation function for N-H bond

$$J(\omega) = 2 \int_0^\infty C(t) \cos(\omega t) dt. \quad (4)$$

The second order rotational correlation is defined as

$$C(t) = \langle (3 \cos^2 \theta_{t';t'+t} - 1)/2 \rangle_{t'}, \quad (5)$$

where average is ensemble average and  $\theta$  is the angle between N-H bonds at times  $t'$  and  $t'+t$ . The dipolar coupling constant is given by

$$d_{\text{NH}} = -\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi \langle r_{\text{CN}}^3 \rangle},$$

where  $\mu_0$  is the magnetic constant or vacuum permeability,  $\hbar$  is the reduced Planck constant,  $\gamma_N$  and  $\gamma_H$  are the gyromagnetic constants of  $^{15}\text{N}$  and  $^1\text{H}$ , respectively. Average cubic length is  $\langle r_{\text{CN}}^3 \rangle \approx$  and the chemical shift anisotropy is  $\Delta\sigma \approx 160 * 10^{-6}$  for N-H bonds in proteins [? ]. Same equations can be used, for example, to C-H bond by changing the constants related to nitrogen to the ones corresponding carbon.

Experimental spin relaxation rates of proteins are typically interpreted by assuming that the global and internal rotational dynamics are independent. The rotational correlation function for each bond can be then written as [? ]

$$C(t) = C_I(t)C_O(t), \quad (6)$$

where  $C_I(t)$  and  $C_O(t)$  are correlation functions for internal and overall rotations, respectively. Within this approximation the internal rotational correlation function decays to a plateau, which defines the square of order parameter respect to molecular axes  $S^2$ . Timescale for internal relaxation dynamics can be estimated by using the effective internal correlation time

$$\tau_{\text{eff}} = \int_0^\infty C'_I(t) dt, \quad (7)$$

where  $C'_I(t) = (C_I - S^2)/(1 - S^2)$  is the reduced correlation function [? ].

The global rotational dynamics for fully anisotropic molecule can be described as a sum of five exponentials [? ]

$$C_O(t) = \sum_{j=1}^5 A_j e^{-t/\tau_j}, \quad (8)$$

where time constants  $\tau_j$  are related [6] to the diffusion constants around three principal axes of a molecule ( $D_{xx}$ ,  $D_{yy}$  and  $D_{zz}$ ) and prefactors  $A_j$  can be related to the directions of bond respect to the principal axes. The rotational diffusion

constants are defined as

$$\begin{aligned} \langle (\Delta\alpha_{t';t'+t})^2 \rangle_{t'} &= 2D_{xx}t \\ \langle (\Delta\beta_{t';t'+t})^2 \rangle_{t'} &= 2D_{yy}t \\ \langle (\Delta\gamma_{t';t'+t})^2 \rangle_{t'} &= 2D_{zz}t, \end{aligned} \quad (9)$$

where  $\langle (\Delta\alpha_{t';t'+t})^2 \rangle_{t'}$ ,  $\langle (\Delta\beta_{t';t'+t})^2 \rangle_{t'}$  and  $\langle (\Delta\gamma_{t';t'+t})^2 \rangle_{t'}$  are mean square angle deviations of protein inertia axes.

The internal and overall correlation functions are monoexponential for proteins with isotropic overall rotational diffusion and single timescale for internal motion. The dynamics of such proteins is described with three parameters in the original "model free analysis"; internal rotational relaxation time  $\tau_e$ , global rotational relaxation time  $\tau_c$  and the order parameter  $S^2$ . The values for these parameters can be then determined by fitting the Eqs. 1-8 to experimental spin relaxation data [? ]. However, the amount of exponentials needed to describe correlation functions, and thus the free parameters, in such fit increases if proteins experience anisotropic overall diffusion or several internal timescales. Thus the "model free analysis" becomes less applicable for proteins with significant anisotropy or more complicated internal dynamics.

## B. Rotational dynamics from molecular dynamics simulations

Classical molecular dynamics simulation gives a trajectory for each atom in a system as a function of time. These trajectories can be used to calculate rotational correlation functions for each bond from Eq. 5. The rotational correlation functions can be further used to calculate the spin relaxation times through Eqs. 1-4 and the resulting values can be compared to experimental data in order to assess simulation model quality [? ] and interpret experiments [? ]. However, the comparison is often complicated by the short simulation times [? ] and incorrect overall rotational diffusion due to water models [? ].

Here we use overall rotational diffusion constants calculated from Eq. 9 to determine the timescales of global rotational correlation function in Eq. 8. The rotational diffusion coefficients are given by a linear fit to the mean square angle deviation of inertia axes calculated from simulations (see results and discussion). Straight line has only one parameter (slope) to be fitted, in contrast to multiexponential sum in Eq. 8 with ten parameters. Thus, the calculation of timescales through overall rotational diffusion constants is numerically more robust and requires less simulation data than a direct fit of Eq. 8 to the rotational correlation function calculated from simulation. The prefactors in 8 are determined by fitting to correlation function calculated from simulations, but timescales from rotational diffusion coefficients are used. The rotational diffusion constants can be also scaled with a constant factor and new correlation functions calculated to compensate the incorrect rotational diffusion due to water model in spin relaxation rate calculations.

The analysis can be divided in essentially six steps:

1) Total rotational correlation functions  $C(t)$  for protein N-H bonds are calculated from MD simulation trajectory by applying Eq. 5.

- 2) Rotational correlation functions for internal dynamics  $C_I(t)$  are calculated from a trajectory from where the overall rotation of protein is removed.
- 3) The overall and internal motions are assumed to be independent and overall rotational correlation function is calculated as  $C_O(t) = C(t)/C_I(t)$  according to Eq. 6.
- 4) The protein axes of inertia and their mean square deviations as function of time are calculated from MD simulation trajectory.
- 5) Rotational diffusion constants  $D_x$ ,  $D_y$  and  $D_z$  are calculated by fitting a straight line to mean square angle deviations of inertia axes according to Eq. 9.
- 6) Timescales in Eq. 8 are calculated from diffusion constants and weighting factors  $A_j$  are determined by fitting the equation to rotational correlation functions of overall rotational motion  $C_O(t)$  determined in step 3.
- 7) New total rotational correlation functions based on Eqs. 6 and 8 are determined as

$$C_N(t) = C_I(t) \sum_{j=1}^5 A_j e^{-t/\tau_j}, \quad (10)$$

where internal correlation function  $C_I(t)$  is taken from step 2,  $\tau_i$  values from rotational diffusion constants (step 5) and prefactors  $A_j$  from step 6. The incorrect rotational diffusion due to water model can be compensated in new correlation functions at this point by scaling the rotational diffusion coefficients with a constant factor.

### C. Simulation and analysis details

Simulations were ran using Gromacs 5 [7] And Amber ff99SB-ILDN [8] force field for proteins. The proteins were solvated to tip3p[9], tip4p [9] or OPC4 [10] water models. NMR structures from [?] and [11] are used as initial structure for PaTonB and HpTonB-92, respectively. Temperature was coupled to desired value with v-rescale thermostat [12] and pressure was isotropically set to 1 bar using Parrinello-Rahman barostat [13]. Timestep was 2 fs, Lennart-Jones interactions were cut-off at 1.0 nm, PME [14, 15] was used for electrostatics and LINCS was used to constraint all bond lengths [16]. Simulation trajectory and related files are available at [?]. The simulated systems are listed in Table I

Rotational correlation functions are calculated with *gmx rotacf* and overall rotation for  $C_I(t)$  calculation is removed by using fit option in *gmx trjconv* from Gromacs package [17]. Inertia axes of protein for rotational diffusion calculation are calculated with *compute\_inertia\_tensor* from MDTraj python library [18]. For spectral density calculation a sum of 471 exponentials having correlation times from 1 ps to 50 ns with logarithmic spacing

$$C_N(t) = \sum_{i=1}^N \alpha_i e^{-t/\tau_i} \quad (11)$$

were fitted to the new correlation function calculated from

Eq. 10 by using the *lsqnonneg* routine in MATLAB [19]. The Fourier transform is then calculated by using analytical function for the sum of exponentials

$$J(\omega) = 4 \sum_{i=1}^N \alpha_i \frac{\tau_i}{1 + \omega^2 \tau_i^2}. \quad (12)$$

Similar approach is used previously for lamellar systems in combination with solid state NMR experiments [20, 21].

## III. RESULTS

### A. Global rotational dynamics of protein

The mean square angle deviations for protein inertia axes as a function of time are shown in Fig. 1 for PsTonB simulation with OPC4 water model, which the longest data set in this work (1.2  $\mu$ s). Linear behaviour of root mean square angle deviations are observed for the lag times up to 12 ns, suggesting that the protein follows linear diffusion model with a good approximation. Plot with log-log scale in Fig. 9 B) reveals a weakly subdiffusive region only below very short timescales of approximately 0.12 ns. Deviations from linear behaviour are also seen with longer lag times than the one hundredth of the simulation length as demonstrated for shorter simulations in Figs. ???. This is expected to be the maximum lag time for a good statistics for rotational dynamics analyzed from a single molecule in MD simulations [22]. Thus, we conclude that the rotational diffusion coefficients of protein can be calculated from the slope of mean square angle deviations according to Eq. 9 by using lag times less than one hundredth of the total MD simulation length.

The resulting rotational diffusion constants from different simulations are shown in Table I. The values are in line with previously reported experimental and simulation results for different proteins with similar size [23, 24]. As expected, rotational diffusion coefficients increase with the temperature and decrease with the size of a protein. Exceptionally large diffusion constants from simulations with tip3p water model are also previously observed and explained with overestimated water self diffusion [23].

The calculated rotational diffusion constants can be used to determine the timescales in Eq. 8 for the overall rotational dynamics of a protein, while prefactors in the equation depend on the orientations of the each N-H bond respect to inertia axes in a complicated way [?]. Here we determine the prefactors by fitting Eq. 8 with the timescales given by diffusion coefficients to the overall rotational correlation function calculated from MD simulations. The prefactors and timescales for overall rotational correlation functions can be then combined with internal correlation function calculated from oriented MD simulations to determine the new correlation functions from Eq. ???. The new correlation function has less fluctuations with longer time scales, thus giving more robust spin relaxation times.

The rotational correlation function analysis is exemplified

TABLE I. Simulated systems and rotational diffusion coefficients ( $\text{rad}^2 \cdot 10^7/\text{s}$ ) calculated from simulations.

Protein	Water model	T (K)	$t_{\text{sim}}$ (ns)	$t_{\text{anal}}$ (ns)	$D_{xx}$	$D_{yy}$	$D_{zz}$	$D_{  }/D_{+}$	$D_{av}$	files
PaTonB	tip4p	298	400	390	$1.81 \pm 0.01$	$2.06 \pm 0.03$	$4.55 \pm 0.03$	$2.35 \pm 0.04$	$2.80 \pm 0.02$	[? ]
PaTonB	tip4p	310	400	390	$2.60 \pm 0.02$	$2.22 \pm 0.05$	$5.0 \pm 0.1$	$2.07 \pm 0.09$	$3.26 \pm 0.07$	[? ]
PaTonB	OPC4	310	1200	1190	$2.01 \pm 0.01$	$2.19 \pm 0.01$	$5.01 \pm 0.03$	$2.39 \pm 0.02$	$3.07 \pm 0.01$	[? ]
HpTonB-92	tip3p	310	570	370	$8.25 \pm 0.05$	$7.67 \pm 0.06$	$15.9 \pm 0.3$	$1.99 \pm 0.06$	$10.6 \pm 0.2$	[? ]
HpTonB-92	tip3p	303	800	790	$6.24 \pm 0.02$	$7.04 \pm 0.03$	$11.9 \pm 0.2$	$1.80 \pm 0.03$	$8.40 \pm 0.07$	[? ]
HpTonB-92	tip4p	310	470	370	$3.6 \pm 0.1$	$3.24 \pm 0.01$	$6.3 \pm 0.3$	$1.8 \pm 0.1$	$4.4 \pm 0.2$	[? ]
HpTonB-92	tip4p	303	400	200	$2.7 \pm 0.1$	$2.71 \pm 0.02$	$5.6 \pm 0.5$	$2.1 \pm 0.2$	$3.7 \pm 0.2$	[? ]
HpTonB-92	OPC4	310	800	790	$2.85 \pm 0.01$	$2.70 \pm 0.01$	$5.56 \pm 0.01$	$2.00 \pm 0.01$	$3.70 \pm 0.01$	[? ]

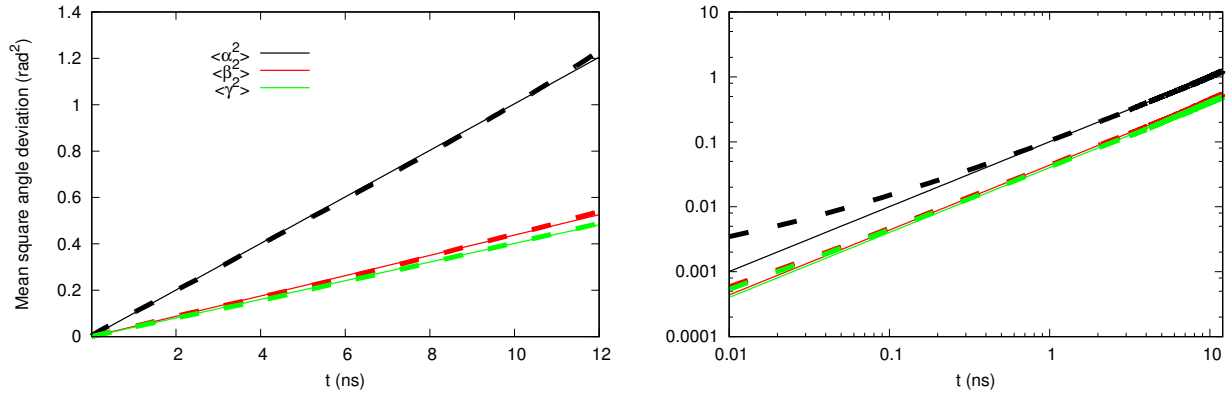


FIG. 1. The inertia tensor angles as a function of time and mean square angular deviations for PsTonB simulation with OPC water model.

in Fig. 2 for residues in different parts of PsTonB having different features in rotational dynamics. Flexible C-terminus is represented by residue 341, more rigid  $\beta$ -sheet by residue 331 and a flexible loop between two sheets by residue 322. The total correlation functions  $C(t)$  calculated from original MD trajectories, shown in Fig. 2 A) (solid lines), decay toward zero within  $\sim 10$ -50 ns. Internal correlation functions  $C_I(t)$  calculated from trajectory with overall protein rotation removed, shown in Fig. 2 B), decay to a plateau value, which defines the square of the order parameter  $S^2$ . As expected, the  $\beta$ -sheet residue 331 has the largest order parameter value and fastest decay to it, while order parameters for loop and C-terminus are significantly smaller and decay is slower due to larger conformational ensemble sampled by these regions. The overall rotational correlation functions  $C_O(t)$  for a protein, determined as  $C_O(t) = C(t)/C_I(t)$ , are shown in Fig. 2 C) (solid lines).

The overall rotational correlation functions from Eq. 8 are also shown in Fig. 2 C) (dashed lines). The timescales  $\tau_i$  were determined from rotational diffusion constants [6] in Table I and prefactors  $A_j$  were determined by fitting to the MD simulation data. The determined parameters were combined with internal correlation function from MD simulations by using Eq. 10 to determine new total correlation functions, which are also shown in Fig. 2 A) (dashed lines). The fitted correlation functions are indistinguishable from MD simulation results

with lag times shorter than one hundredth of total simulation time (approximately 4-12 ns for the studied systems), which is the expected limit for a good statistics in single molecule MD simulations [22]. Exceptionally large statistical fluctuations for overall rotational correlation function of flexible C-terminus (residue 341) with longer lag times are observed, because small contribution of overall rotation dynamics is difficult to detect with high accuracy for segments with small order parameters.

Good agreement between correlation functions from Eq. 10 with overall rotational timescales from diffusion coefficients and MD simulations, suggest that the anisotropic diffusion model (Eq. 8) and separation of internal and global motions (Eq. 6) are good approximations for the studied system. Thus, we conclude that the determined new correlation functions can be then used to reduce the influence of statistical fluctuations arising from long time scale dynamics on spin relaxation time calculations from MD simulations. In addition, incorrect overall diffusion can be corrected post-simulationally in this calculation by scaling the diffusion coefficients when calculating new correlation functions. This analysis is exemplified for PsTonB and HpTonB-92 proteins in sections below.

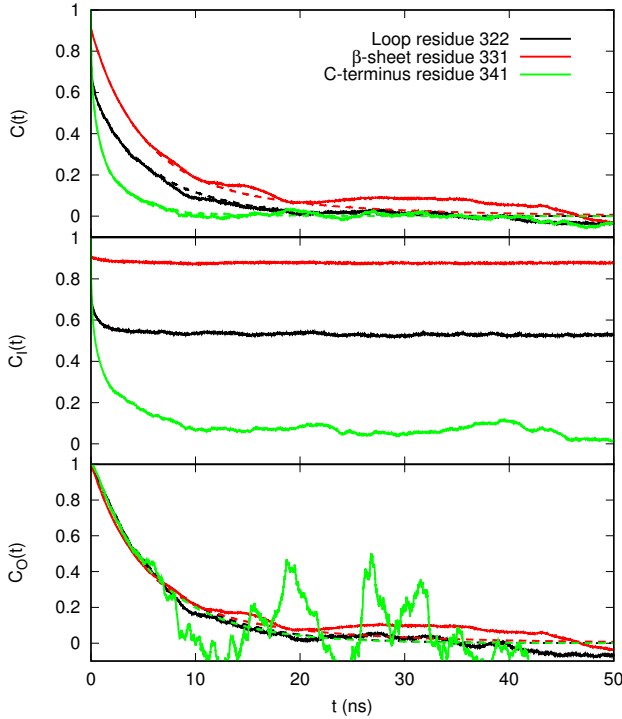


FIG. 2. Rotational correlation functions calculated from MD simulations of PsTonB with tip4p water model at 298K for residues at different regions. A) total correlation functions  $C(t)$  calculated from MD simulation (solid lines) and new correlation functions determined from Eqs. 6 and 8 by using rotational diffusion constants and fitted prefactors (dashed lines), B) correlation functions for internal motions calculated from simulation with removed overall protein rotation C) correlation function for overall motions determined as  $C_O(t) = C(t)/C_I(t)$  (solid lines) and by fitting to Eq. 8 with timescales from rotational diffusion coefficients in Table I (dashed lines).

### B. Global rotational dynamics in simulations and experiments

Rotational correlation functions from simulations were determined by using Eq. 10, where internal correlation functions were taken from MD simulations, timescales were determined from protein overall rotation diffusion constants and prefactors from the fit to the MD simulation data, as described in previous section. The correlation functions were used to calculate spectral densities from Eqs. 11-12 and spin relaxation times were then calculated from Eqs. 1-3.

Spin relaxation time results for HpTonB-92 in Fig. 3 show a good agreement between simulation with tip4p water model and experiments. However, the results from simulation with tip3p water model are significantly off from experiments. The underestimated  $T_1/T_2$  ratio suggests too fast overall rotational diffusion dynamics [25] with tip3p water, in agreement with previous studies [23]. Indeed, division of diffusion constants by a constant factor of 2.9 before applying Eq. 10 brings spin relaxation times in good agreement with experiments, as shown in Fig. 4.

Simulations of PsTonB with tip4p or OPC4 water models,

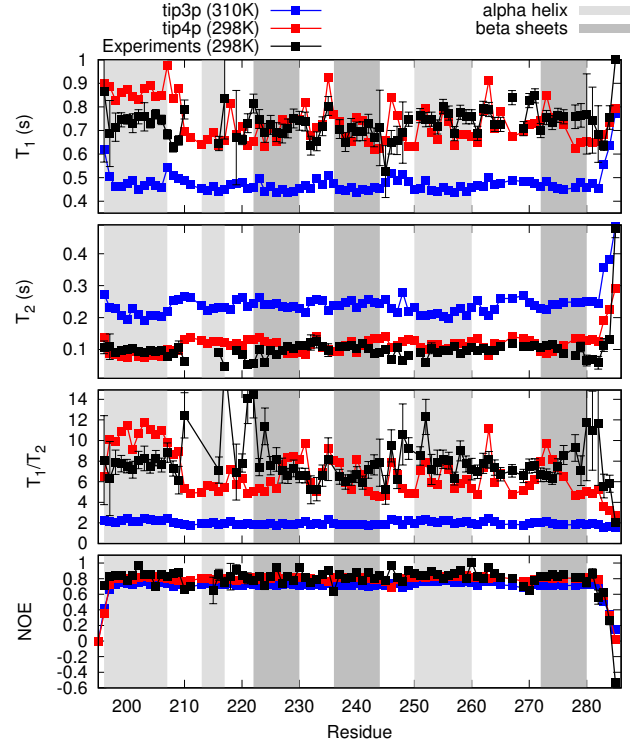


FIG. 3. Spin relaxation times for HpTonB-92 from experiments [?] and simulations with different water models.

shown in Fig. 5, give systemically smaller  $T_1$  values and  $T_1/T_2$  ratio than experiments. The effect of temperature difference of 12 degrees is significantly smaller than the different with experiments. Dividing the diffusion coefficients with a constant factor of 1.2 gives a good agreement for spin relaxation rates between tip4p simulation and experiments at 298K, as seen in Fig. 6.

The good agreement between spin relaxation data from experiments and simulations with scaled diffusion coefficients suggests that the values can be used to interpret the anisotropic diffusion of these proteins from the measured NMR data. The scaled rotational diffusion coefficients from MD models giving the best agreement with experimental spin relaxation data are shown in Table I. The scaled diffusion coefficients from tip3p simulations were chosen for HpTonB-92, because they give slightly better overall agreement with experiments than simulations with tip4p water model. The results are in line with values for more isotropic proteins, which are previously determined by different methods to interpret the spin relaxation data [24].

### C. Interpretation of protein internal relaxation from MD simulations

Spin relaxation rates from tip4p simulations for PsTonB are in good agreement with experiments (see Figs ??), thus the simulations can be used to give interpretation for rotational relaxation processes in proteins, which correspond the NMR

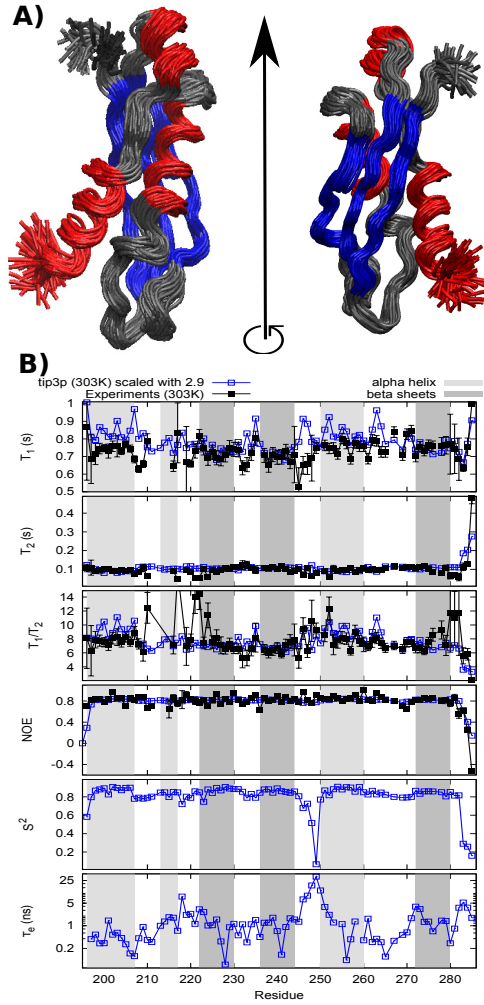


FIG. 4. A) Structures sampled by HpTonB-92 from MD simulations with tip3p at 303 K (100 structures from 400ns long trajectory). Secondary structures are colour labelled with Visual Molecular dynamics [26, 27];  $\alpha$ -helices are red and  $\beta$ -sheets are blue. B) Spin relaxation times from experiments and tip3p simulations with rotational diffusion coefficients divided by a constant factor of 2.9 at 303 K. Order parameters and effective internal correlation times calculated from simulations

relaxation results. Spin relaxation rate deviations from baseline are observed for residues 246-251 in N-terminus, residues 320-326 and residues 338-342 in C-terminus. These segments are coloured in yellow in Fig. ?? A), which already reveals enhanced conformational sampling in these regions. Also order parameters are low and effective internal relaxation times long for these segments as seen in Fig. ?? B).

More detailed interpretation of different relaxation processes experienced by different residues can be done by analysing timescales which leads in simulation model to the spin relaxation rates in agreement with experiments. Prefac-

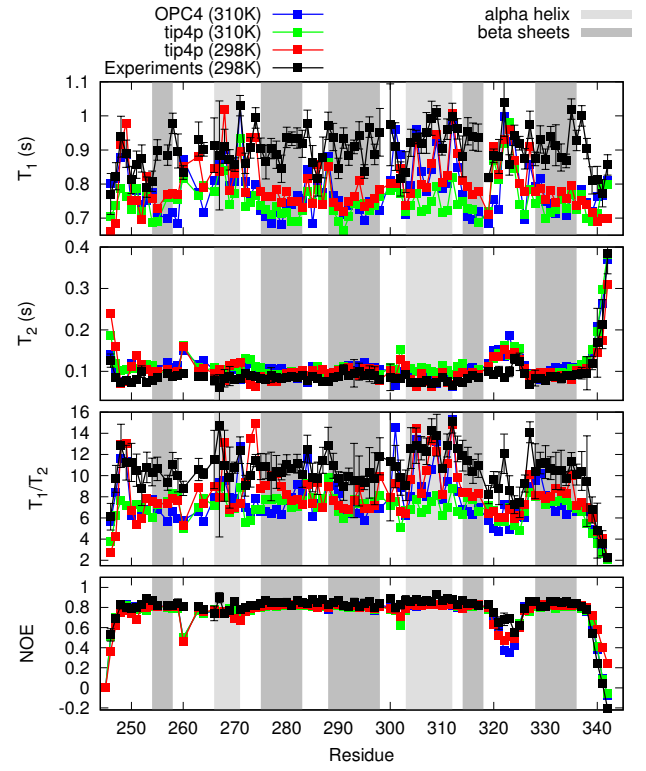


FIG. 5. Spin relaxation times for PsTonB from experiments [?] and simulations with different water models.

TABLE II. Rotational diffusion coefficients scaled with constant factor which gives a good agreement for spin relaxation data, 2.9 for tip3p simulation of HpTonB and by 1.2 for tip4p simulation of PsTonB.

	HpTonB-92	PsTonB
$D_{xx}$	$2.15 \pm 0.01$	$1.51 \pm 0.01$
$D_{yy}$	$2.43 \pm 0.01$	$1.72 \pm 0.03$
$D_{zz}$	$4.10 \pm 0.01$	$3.79 \pm 0.03$
$D_{av}$	$2.90 \pm 0.03$	$2.3 \pm 0.02$
$\tau_c$ (ns)	$5.7 \pm 0.1$	$7.2 \pm 0.1$

tors from Eq. ?? fitted in rotational correlation functions in agreement with spin relaxation data are shown in Fig 7 for different residues in PsTonB. Residue 331 represents the last alpha helix before C terminus and its rotational relaxation is mostly dominated by relaxations with timescales  $\sim 5.5$  ns and  $\sim 8$  ns, which arise from global rotation of protein and only small fraction of relaxation arises from fast internal motions, in accordance with large order parameter value (??). Relaxation of residue 322 is also dominated by relaxation processes with timescale around  $\sim 8$  ns, but fast motions related to internal protein dynamics are more significant than for alpha helix residue 331. This explains the low order parameter (??) measured small NOE and large  $T_2$  relaxation times values shown in Fig. ?. Rotational dynamics of residue 341 located in N terminus is dominated by the fast motions related to the internal protein relaxation, as expected from the low order pa-



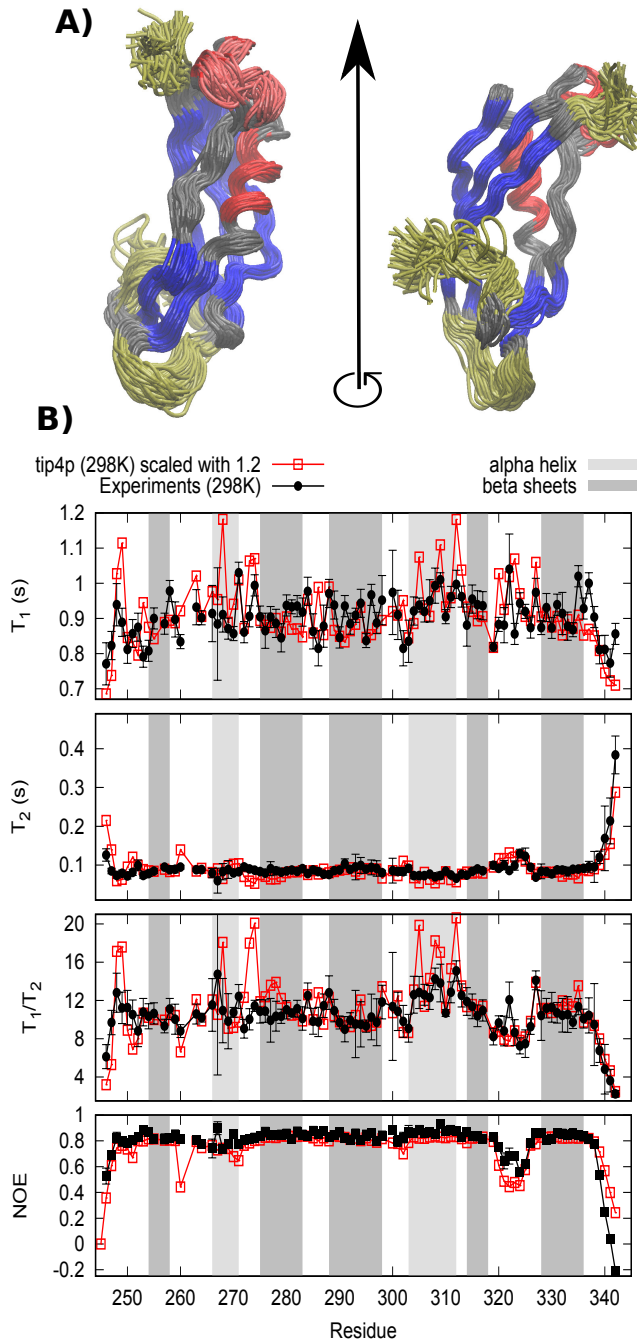


FIG. 6. A) Structures sampled by PsTonB from MD simulations with tip4p at 298 K (100 structures from 400ns long trajectory). Secondary structures are colour labelled with Visual Molecular dynamics [26, 27];  $\alpha$ -helices are red and  $\beta$ -sheets are blue. Residues 246-251, 320-326 and 338-342 with increased internal dynamics are yellow and  $\alpha$ -helix sampling between two orientations (residues 266-270) is pink in the left column. B) Spin relaxation times from experiments and tip4p simulations with rotational diffusion coefficients divided by a constant factor of 1.2 at 298 K. Order parameters and effective internal correlation times calculated from simulations.

parameter. The contribution from timescales close to  $\sim 13$  ns are probably related to slower conformational sampling of the N

terminus, which is also seen in sampled conformations and large effective correlation times in Fig. ??.

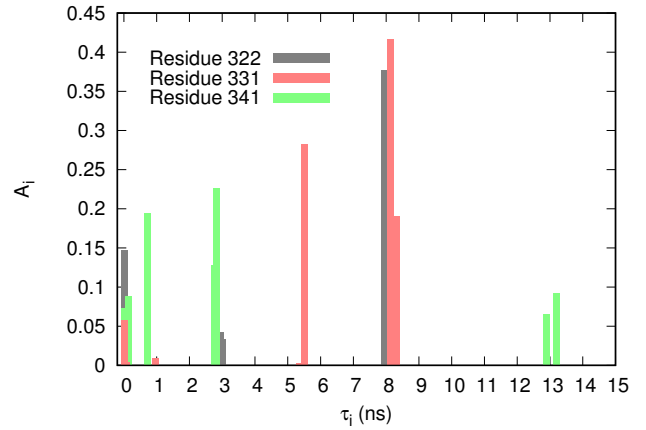


FIG. 7. Prefactors  $A_i$  corresponding different timescales  $\tau_i$  in Eq.?? resulting from a fit in correlation functions giving good agreement with spin relaxation rates in experiments for PsTonB.

In addition, low order parameters and larger effective correlation times are observed for PsTonB residues 260-274 and 300-303 in Fig. 5 B). For residues 260-274 this can be explained by two different orientations sampled by alpha helix on that region, as highlighted with pink in Fig. 5 A). This explains also the lower resolution in NMR spectra observed for this and similar region in HpTonB structures [? ]. Low order parameters and large effective correlation times between residues 300-303 are not seen in spin relaxation data, thus it is not clear if these arise from simulation artefact.

Spin relaxation rate deviations from baseline are observed only few residues in terminal ends for HpTonB-92 as seen in Fig. ?? . This is the case also in simulations, except that the N terminus flexibility seems to be somewhat overestimated and low order parameters and long effective correlation times are also observed for residues 245-250. The latter observation probably arises (at least partly) from simulation artefacts because deviation from experimental relaxation data is relatively large for these residues. More detailed discussion together with longer HpTonB construct is presented elsewhere [? ].

#### IV. DISCUSSION AND CONCLUSIONS

and it is typically overcome by introducing isotropic rotational diffusion term in the correlation functions [? ] or correcting overall rotation by using quaternions [? ].

The scaling factors were adjusted to bring  $T_1/T_2$  ratio in optimal agreement with experiments, which resulted scaling factor of 1.2 for PsTonB simulation with tip4p and factor 2.9 for HpTonB with tip3p.

The advantage of this approach is that the statistical fluctuations with longer lag times do not interfere the spin relaxation rate calculations. The higher precision is essentially achieved

by assuming the anisotropic diffusion behaviour, which allows the determination of global rotational correlation functions from diffusion constants. Diffusion constants can be determined by fitting a single parameter to a linear slope which is more robust than fitting exponential function with several parameters to a correlation function calculated from MD simulation. In addition, diffusion constants can be scaled with a constant factor before calculating the new correlation functions to compensate the overestimated rotational diffusion due to inaccuracies in some water models.

Rotation of protein inertia axes is observed to experience linear diffusion behaviour and overall diffusion component of rotational correlation functions of individual N-H bonds can be successfully fitted to the model assuming anisotropic diffusion for whole molecule.

Rotational diffusion of whole molecules is overestimated

by a factor of  $\sim 3$  in simulations with tip3p water, in agreement with previous studies [? ]. The simulations with tip4p and opc4 water models give more realistic diffusion coefficients, which overestimate diffusion only with factors  $\sim 1$ -1.2.

The overestimated overall diffusion coefficient can be corrected post-simulationally to compare internal dynamics and order with experiments.

The presented methodology can be used to interpret spin relaxation experiments by using MD simulations [? ] and assess the quality of protein force fields against NMR experiments. The presented scaling of overall anisotropic diffusion allows this also for simulations with incorrect rotational diffusion due to water models, which is the case in simulations with tip3p.

## ACKNOWLEDGMENTS

- 
- [1] V. A. Jarymowycz and M. J. Stone, *Chemical Reviews* **106**, 1624 (2006).
  - [2] D. Korzhnev, M. Billeter, A. Arseniev, and V. Orekhov, *Progress in Nuclear Magnetic Resonance Spectroscopy* **38**, 197 (2001).
  - [3] .
  - [4] A. Abragam, *The Principles of Nuclear Magnetism* (Oxford University Press, 1961).
  - [5] L. E. Kay, D. A. Torchia, and A. Bax, *Biochemistry* **28**, 8972 (1989).
  - [6]  $\tau_1 = (4D_{xx} + D_{yy} + D_{zz})^{-1}$ ,  $\tau_2 = (D_{xx} + 4D_{yy} + D_{zz})^{-1}$ ,  $\tau_3 = (D_{xx} + D_{yy} + 4D_{zz})^{-1}$ ,  $\tau_4 = [6(D + (D^2 - L^2)^{-1/2})^{-1}]^{-1}$ ,  $\tau_5 = [6(D - (D^2 - L^2)^{-1/2})^{-1}]^{-1}$ ,  $D = \frac{1}{3}(D_{xx} + D_{yy} + D_{zz})$  and  $L^2 = \frac{1}{3}(D_{xx}D_{yy} + D_{xx}D_{zz} + D_{yy}D_{zz})$ .
  - [7] M. J. Abraham, T. Murtola, R. Schulz, S. Pii, J. C. Smith, B. Hess, and E. Lindahl, *SoftwareX* **12**, 19 (2015).
  - [8] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, *Proteins: Structure, Function, and Bioinformatics* **78**, 1950 (2010).
  - [9] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
  - [10] S. Izadi, R. Anandakrishnan, and A. V. Onufriev, *The Journal of Physical Chemistry Letters* **5**, 3863 (2014).
  - [11] A. Ciragan, A. S. Aranko, I. Tascon, and H. Iwa, *Journal of Molecular Biology* **428**, 4573 (2016).
  - [12] G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126** (2007).
  - [13] M. Parrinello and A. Rahman, *J. Appl. Phys.* **52**, 7182 (1981).
  - [14] T. Darden, D. York, and L. Pedersen, *J. Chem. Phys.* **98**, 10089 (1993).
  - [15] U. L. Essman, M. L. Perera, M. L. Berkowitz, T. Larden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).
  - [16] B. Hess, *J. Chem. Theory Comput.* **4**, 116 (2008).
  - [17] M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS user manual version 5.0.7* (2015).
  - [18] 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, and V. S. Pande, *Biophysical Journal* **109**, 1528 (2015).
  - [19] .
  - [20] A. Nowacka, N. Bongartz, O. Ollila, T. Nylander, and D. Topgaard, *J. Magn. Res.* **230**, 165 (2013).
  - [21] T. M. Ferreira, O. H. S. Ollila, R. Pigliapochi, A. P. Dabkowska, and D. Topgaard, *J. Chem. Phys.* **142**, 044905 (2015).
  - [22] C.-Y. Lu and D. A. V. Bout, *The Journal of Chemical Physics* **125**, 124701 (2006).
  - [23] V. Wong and D. A. Case, *The Journal of Physical Chemistry B* **112**, 6013 (2008).
  - [24] V. Krishnan and M. Cosman, *Journal of Biomolecular NMR* **12**, 177 (1998).
  - [25] W. R. Carper and C. E. Keller, *The Journal of Physical Chemistry A* **101**, 3246 (1997).
  - [26] D. Frishman and P. Argos, *Proteins: Structure, Function, and Bioinformatics* **23**, 566 (1995).
  - [27] W. Humphrey, A. Dalke, and K. Schulten, *Journal of Molecular Graphics* **14**, 33 (1996).

## SUPPLEMENTARY INFORMATION



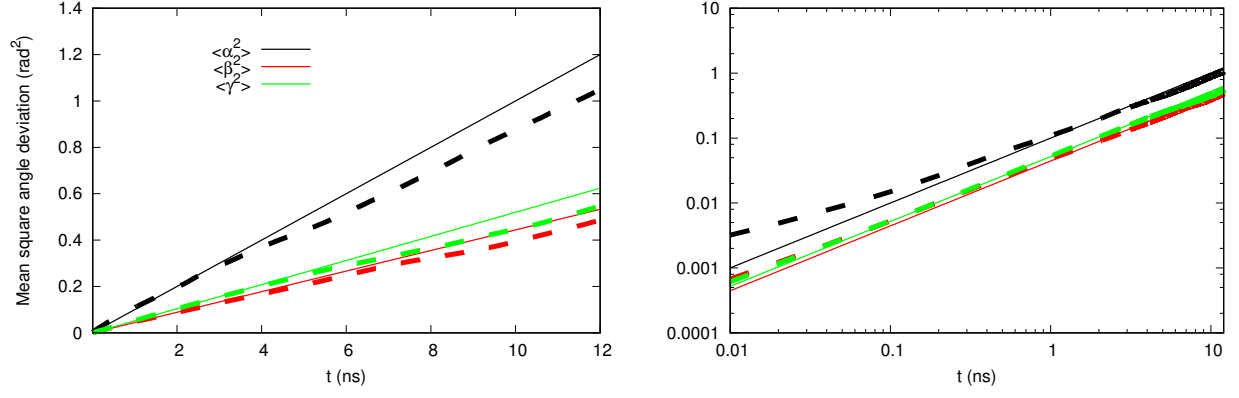


FIG. 8. The inertia tensor angles as a function of time and mean square angular deviations for PsTonB simulation with OPC water model.

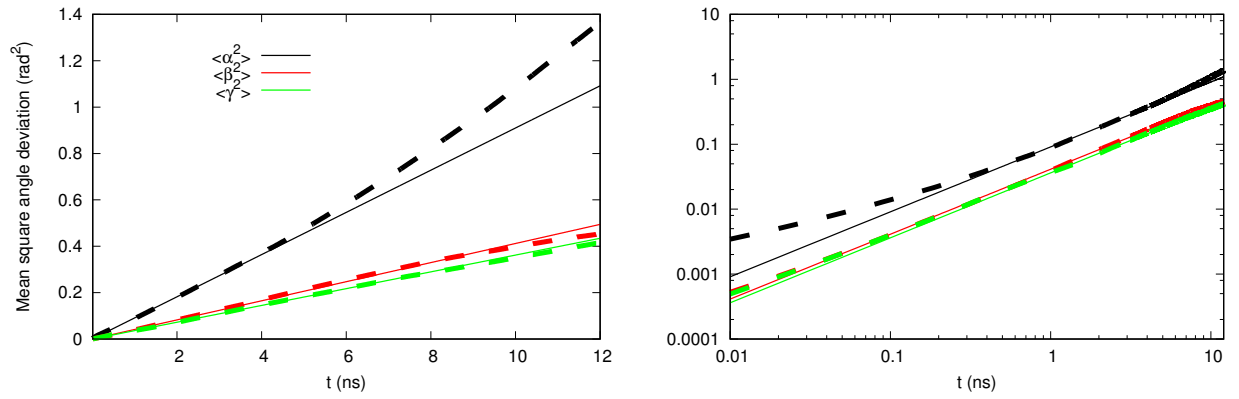


FIG. 9. The inertia tensor angles as a function of time and mean square angular deviations for PsTonB simulation with OPC water model.