Rotational dynamics of proteins from spin relaxation times and molecular dynamics simulations

O. H. Samuli Ollila*
Insititute of Biotechnology, University of Helsinki and
Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague 6, Czech Republic

Hideo Iwaï
Insititute of Biotechnology, University of Helsinki
(Dated: August 17, 2017)

I. INTRODUCTION

Conformational sampling and entropy of proteins play a significant role in functionality and interactions with other biomolecules. Conformation sampling and overall brownian tumbling of proteins are experimentally accessible through spin relaxation times of ^{15}N and ^{13}C nucleai measured with nuclaer magnetic resonance (NMR) techniques [1–7]. Spin relaxation rate experiments have been used to, for example, analyze conformational entropies [1, 8–10], binding entropies [1, 11], resolve sampled structures [3–5, 12] and validate molecular dynamics simulations [13–17]. These analyzes are almost exclusively based on the separation of internal conformational sampling and overall rotational tumbling [18, 19] and isotropic overall diffusion is often assumed, while analysis of anisotropic molecules is significantly more complicated [1, 2, 20–23]. Thus, new approaches are needed to intepret spin relaxation times measured from anisotropic or intrinsically disordered molecules.

Classical molecular dynamics simulation methods are promising tools to interpretate spin relaxation experiments for molecules with significant anisotropy or correlations between internal and overall rotational motions. Practical applications are, however, limited by inaccuracies in the force field descriptions and available time scales in the simulations [16, 17, 24–26]. The main issues have been the overestimated overall rotational diffusion of proteins due to inaccuracies in water models [25] and insufficient accuracy of correlation functions calculated from single molecules in MD simulations [26, 27].

In this work we overcome these issues by assuming that the overall rotational dynamics of protein follows anisotropic rigid body diffusion. Diffusion coefficients around inertia axes are directly calculated from angular displacements. The diffusion coefficients are then used to determine the contribution of overall rotational tumbling to the N-H bond rotational correlation functions. This reduces the required simulation length for accurate determination of rotational correlation functions. Furthermore, the overestimated overall brownian tumbling rates due to the inaccurate water model can be corrected during correlation function calculation by scaling diffusion coefficients in all directions with a constant factor. The corrected correlation functions can be used to interpret spin relaxation experiments for proteins with arbitrarily anisotropic shape.

The developed approach is demonstrated by interpreting spin relaxation experiments for C-terminal domains of TonB proteins from *H. pyroli* (*Hp*TonB) [28] and from *Pseudomonas aeruginosa* (*Pa*TonB). Both proteins having significantly anisotropic shape, which would significantly complicate the standard spin relaxation data analysis [1, 2, 20–23].

II. METHODS

A. Spin relaxation experiments and rotational dynamics of molecules

Molecular dynamics of protein backbone residues and spin relaxation experiments can be connected by using the spectral density $J(\omega)$

$$J(\omega) = 2\int_0^\infty C(t)\cos(\omega t)\mathrm{d}t, \tag{1}$$

which is the Fourier transformation of the second order rotational correlation function for N-H bond vector

$$C(t) = \langle \frac{3}{2}\cos^2\theta_{t'+t} - \frac{1}{2}\rangle_{t'}, \tag{2}$$

where $\theta_{t'+t}$ is the N-H bond angle between times t' and $t'+t$ and angular brackets refer to the ensemble average. Connection to experimentally measured spin relaxation times T_1, T_2 and T_{NOE} is given by Redfield equations [29, 30]

$$\frac{1}{T_1} = \frac{d_{\mathrm{NH}}^2 N_{\mathrm{H}}}{20}\Big[J(\omega_{\mathrm{H}}-\omega_{\mathrm{N}}) + 3J(\omega_{\mathrm{N}}) + 6J(\omega_{\mathrm{N}}+\omega_{\mathrm{H}})\Big] + \frac{(\sigma\omega_{\mathrm{N}})^2}{15}j(\omega_{\mathrm{N}}), \tag{3}$$

$$\frac{1}{T_2} = \frac{1}{2}\frac{d_{\mathrm{NH}}^2 N_{\mathrm{H}}}{20}\Big[4J(0) + 3j(\omega_{\mathrm{N}}) + J(\omega_{\mathrm{H}}-\omega_{\mathrm{N}}) + 6J(\omega_{\mathrm{H}}) + 6J(\omega_{\mathrm{N}}+\omega_{\mathrm{H}})\Big] + \frac{(\sigma\omega_{\mathrm{N}})^2}{90}[4J(0)+3J(\omega_{\mathrm{N}})], \tag{4}$$

* samuli.ollila@helsinki.fi

$$\frac{1}{T_{\mathrm{NOE}}} = 1 + \frac{d_{\mathrm{NH}}^2 N_{\mathrm{H}}}{20}\left[6J(\omega_{\mathrm{N}}+\omega_{\mathrm{H}}) + J(\omega_{\mathrm{H}}-\omega_{\mathrm{N}}))\right]\frac{\gamma_{\mathrm{H}}}{\gamma_{\mathrm{N}} R_1}, \quad (5)$$

where ω_{N} and ω_{H} are the Larmor angular frequencies of ^{15}N and ^{1}H respectively, and the number of bound protons is $N_{\mathrm{H}} = 1$ for N-H bonds. The dipolar coupling constant is given by

$$d_{\mathrm{NH}} = -\frac{\mu_0 \hbar \gamma_{\mathrm{H}} \gamma_{\mathrm{N}}}{4\pi \langle r_{\mathrm{CN}}^3 \rangle},$$

where μ_0 is the magnetic constant or vacuum permeability, $\hbar$ is the reduced Planck constant, γ_{N} and γ_{H} are the gyromagnetic constants of ^{15}N and ^{1}H, respectively. Average cubic length is calculated as $\langle r_{\mathrm{CN}}^3 \rangle = (0.101\mathrm{nm})^3$ and the value of $\Delta\sigma = -160$ ppm is used for the chemical shift anisotropy of N-H bonds in proteins [30, 31].

Spin relaxation experiments are typically interpreted for proteins by assuming that motions related to overall brownian tumbling and conformational sampling are independent. The rotational correlation function for chemical bonds can be then written as [1, 2, 18, 19, 32]

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

where $C_I(t)$ and $C_O(t)$ are correlation functions for internal and overall rotations, respectively. Conformational sampling can be described in this approximation by using the square of order parameter respect to molecular axes S^2, which is given by the plateau of the internal rotational correlation function, and the effective correlation time

$$\tau_{\mathrm{eff}} = \int_0^\infty C_I'(t)\mathrm{d}t, \quad (7)$$

where $C_I'(t) = \frac{C_I - S^2}{1-S^2}$ is the reduced correlation function [19].

The overall rotational correlation function is often described by approximating a protein as a rigid body, which gives a sum of five exponentials fully describing a anisotropic molecule [2, 20]

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

where time constants τ_j are related to the diffusion constants around three principal axes of a molecule (D_{xx}, D_{yy} and D_{zz}) [33] and prefactors A_j to the directions of chemical bonds respect to the molecular axes [20, 22].

The simplest approach to extract molecular dynamics from experimental data is the original "model free analysis" [19]. Assumption of isotropic diffusion for overall rotational reduces Eq. 8 to monoexponential and the overall rotational dynamics can be described with a single time constant τ_c in the original model free analysis. Also internal correlation function for each residue are assumed to decay exponentially with single time constant τ_{eff} towards square of the order parameter S^2. The parameters can be successfully fit to experimental data for isotropic molecules, however, the fitting becomes significantly more difficult for proteins with anisotropic overall diffusion or several internal timescales [1, 22, 34]. Alternative approaches to describe overall protein diffusion with hydrodynamical calculations are less problematic for anisotropic systems, but they are sentitive for the assumptions about protein hydration shell [35].

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. Rotational correlation functions for each bond can be directly calculated from the trajectories by using Eq. 2 and then used to calculate the spin relaxation times through Eqs. 1-5. The resulting values can be compared to experimental data in order to assess simulation model quality [13–17, 36] or interpret experiments [36].

The direct comparison with experiments is often complicated by insufficient statistics for calculated correlation functions and overestimated rotational diffusion due to inaccuracies in water models [25, 26]. Here we show that the statisitical accuracy of overall tumbling contribution to correlation functions, $C_0(t)$ in Eq. 6, can be increased for rigid proteins by directly calculating diffusion constants for intertia axes. The rotational diffusion coefficients can be related to the timescales in correlation function for anisotropic rigid body rotation in Eq. 8 [20, 33].

The diffusion coefficients are calculated by fitting a slope on linear mean square angle deviation of intertia axes (see below). This requires less simulation data for good statistics than a direct fit of multiexponential sum in Eq. 8 to the rotational correlation function calculated from MD simulation. In addition, overestimated rotional diffusion due to water model [25] can be corrected by scaling the diffusion coefficients around all intertia axes by a constant factor. This approach takes into account the anistoropy of the molecule. This is a signficant advancement to the studies, which assume isotropic rotational diffusion with single exponential rotational correlation function [10, 14–16, 37] or use order parameters to compare simulations with experiments [13, 16, 17, 37].

The practical analysis can be divided in essentially six steps:
1) Total rotational correlation functions $C(t)$ for protein N-H bonds are directly calculated from MD simulation trajectory by applying Eq. 2.
2) Rotational correlation functions for internal dynamics $C_I(t)$ are calculated from MD simulation trajectory with removed overall rotation of the protein.
3) The overall and internal motions are assumed to be independent and overall rotational correlation function is calculated from Eq. 6 as $C_O(t) = C(t)/C_I(t)$.
4) Mean square angle deviations of rotation around protein intertia axes 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 the mean square angle devia-

tions

$$\begin{aligned}\langle\Delta\alpha^2_{t'+t}\rangle_{t'} &= 2D_x t\\ \langle\Delta\beta^2_{t'+t}\rangle_{t'} &= 2D_y t\\ \langle\Delta\gamma^2_{t'+t}\rangle_{t'} &= 2D_z t,\end{aligned} \tag{9}$$

where $\langle\Delta\alpha^2_{t'+t}\rangle_{t'}$, $\langle\Delta\beta^2_{t'+t}\rangle_{t'}$ and $\langle\Delta\gamma^2_{t'+t}\rangle_{t'}$ are the mean square angle deviations from the shortest protein intertia axis to the longest, respectively.

6) Contribution of overall rotational tumbling to all correlation functions is assumed to follow Eq. 8 with timescales τ_j calculated from rotational diffusion constants [33]. Weighting factors A_j are determined by fitting the equation with new timescales to the overall rotational correlation functions calculated from MD simulations in step 3.

7) New correlation functions are calculated by substituting internal correlation functions, $C_I(t)$, from step 2 and anisotropic rigid body rotational correlation functions, $C_O(t)$, from step 6 to Eq. 6 giving

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

These correlation functions are then used to calculate spin relaxation times from Eqs. 1-5. The incorrect overall rotational diffusion due to water model can be corrected at this point by scaling the rotational diffusion coefficients, i.e. timescales τ_j, with a constant factor before calculating new correlation functions from Eq. 10.

C. Simulation and analysis details

All simulations were ran using Gromacs 5 [38] and Amber ff99SB- ILDN [39] force field for proteins . The proteins were solvated to tip3p[40], tip4p [40] or OPC4 [41] water models. Initial structures were taken from NMR structures for *Hp*TonB [28] and *Pa*TonB [**?**]. Temperature was coupled to desired value with v-rescale thermistat [42] and pressure was isotropically set to 1 bar using Parrinello-Rahman barostat [43]. Timestep was 2 fs, Lennart-Jones interactions were cut-off at 1.0 nm, PME [44, 45] was used for electrostatics and LINCS was used to constraint all bond lengths [46]. Simulation trajectory and related files are available at [**?**]. The simulated systems are listed in Table I

Rotational correlation functions are calculated with *gmx rotacf* from Gromacs package [47]. Overall rotation was removed for $C_I(t)$ calculation by using fit option in *gmx trjconv* from Gromacs package [47]. The order parameters S^2 were determined by averaging rotational correlation functions from oriented trajectory, $C_I(t)$, over lag times above 50 ns. Effective correlation times were then calculated by using Eq. 7. Inertia axes of protein were calculated with *compute_inertia_tensor* function from MDTraj python library [48].

Spectral density was calculated by fitting a sum of 471 exponentials with timescales from 1 ps to 50 ns with logarithmic spacing

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

to the new correlation function from Eq. 10 by using the *lsqnonneg* routine in MATLAB [49]. The Fourier transform was then calculated by using analytical function for the sum of exponentials

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

Similar approach is used previously for lamellar systems in combination with solid state NMR experiements [50, 51].

D. Spin relaxation experiments

III. RESULTS

A. Global rotational dynamics of protein

Mean square angle deviations for rotation around inertia axes of *Pa*TonB protein in simulation with OPC4 water model are shown in Fig. 1. This is the longest simulation data set in this work (1.2μs) and linear behaviour of mean square angle deviations are observed for the lag times up to one hundreth of the total simulation length (12 ns), which is expected to be the maximum lag time for a good statistics for rotational dynamics analyzed from a single molecule in MD simulations [27]. Deviations from linear behaviour are only seen with lag times longer than this limit, as also demonstrated for shorter simulations in Figs. 8 and 9 for simulations with tip4p water at two different temperatures. The plots with log-log scale in Figs. 1, 8 and 9 B) reveals a weakly subdiffusive region only below very short timescales of approximately 0.12 ns. Thus, we conclude that the protein experiences brownian rotational tumbling with a good approximation and that the diffusion coefficients can be calculated from the slope of mean square angle deviations according to Eq. 9 by using lag times less than one hundreth of the total MD simulation length. The error bars were calculated varying the lag time with 1 ns to both directions. The data from *Hp*TonB protein (not shown) led to the similar conclusions.

The resulting rotational diffusion constants from different simulations are shown in Table I. The values are larger than would be expected from experiments for different proteins with similar size [52], especially when tip3p water model is used. This is in line with previously reported simulation results, where the results were explained by overestimated water self diffusion [25]. As expected, rotational diffusion coefficients increase with the temperature and decreasing size of a protein.

The analysis leading to the new correlation functions in Eq. 10 (see section II B) is exemplified in Fig. 2 for residues in different parts of *Pa*TonB protein having different features

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

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

[a] Simulation temperature
[b] Total simulation time
[c] Analyzed simulation time
[d] $D_{||} = D_{zz}$, $D_\perp = \frac{1}{2}(D_{xx} + D_{yy})$
[e] $D = \frac{1}{3}(D_{xx} + D_{yy} + D_{zz})$

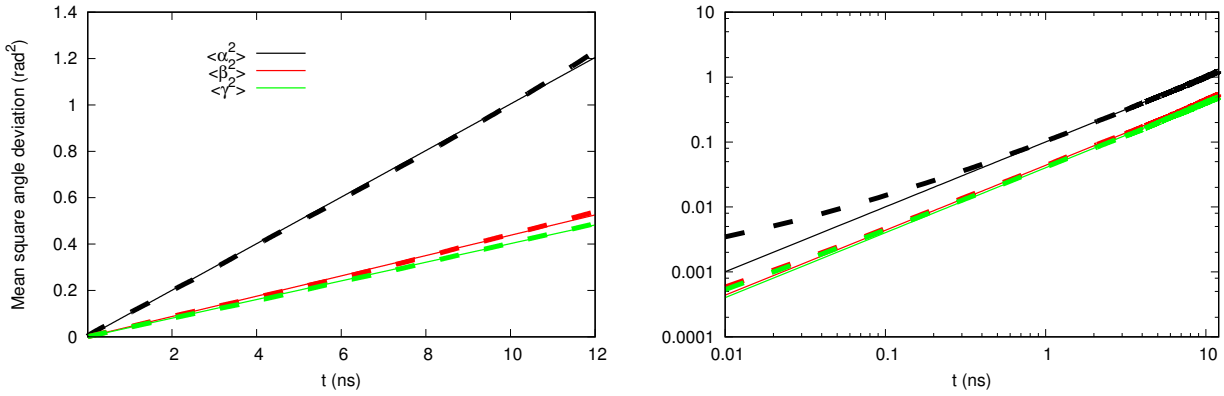

FIG. 1. The intertia tensor angles as a function of time and mean square angular deviations for *Pa*TonB simulation with OPC water model.

in rotational dynamics. Flexible C-terminus is represented by residue 341, more rigid β-sheet by residue 331 and a flexible loop between two sheets by residue 322. The total correlation functions $C(t)$ in Fig. 2 A (solid lines), calculated from original MD trajectories, decay toward zero within ~10-50 ns for all residues. Internal correlation functions $C_I(t)$ in Fig. 2 B), calculated from trajectory with removed overall rotation of the protein, decay to a plateau value, which defines the square of the order parameter S^2. As expected, the rigid β-sheet residue 331 has the largest order parameter value and fastest decay to it, while order parameters for the loop and C-terminus residues are significantly smaller and deacay 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).

Anisotropic rigid body rotational correlation functions from Eq. 8 with the timescales τ_i from rotational diffusion constants [33] and prefactors A_j from the fit to MD simulation data are shown in Fig. 2 C (dashed lines). New correlation functions from Eq. 10 are shown in Fig. 2 A (dashed lines). The new correlation functions are indistinguishable from the original MD simulation results with lag times shorter than one hundredth of total simulation time (approximately 4-12ns for the studied systems), which is the expected limit for a good statistics in single molecule MD simulations [27]. This suggests that the anisotropic rigid body diffusion model (Eq. 8) and separation of internal and global motions (Eq. 6) are good approximations for the studied system. Statistical fluctuations with long lag times are clearly reduced in the new correlation functions in Fig. 2, because overall rotation is analytically described with Eq. 8. This is noticeable especially in flexible C-terminus (residue 341), where contribution of overall rotation dynamics to correlation function is small due to the small order parameters values.

B. Global rotational dynamics in simulations and experiments

Spin relaxation times are compared between experiments and simulations from two different water models in Fig. 3 for *Hp*TonB construct. Simulations with tip3p water model are

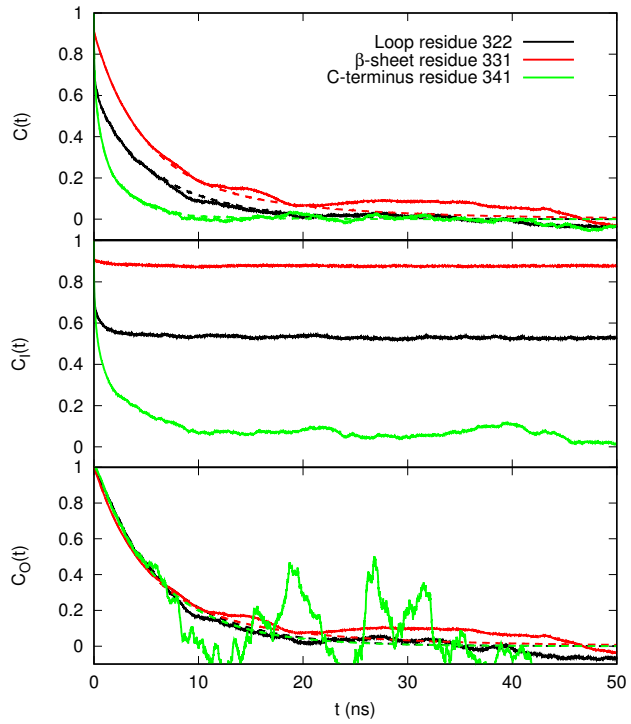

FIG. 2. Rotational correlation functions calculated from MD simulations of *Pa*TonB with tip4p water model at 298K for residues at different regions. A) total correlation functions $C(t)$ calculated from MD simulation (solid lines) and new correlation functions determined from Eqs. 6 and 8 by using rotational 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).

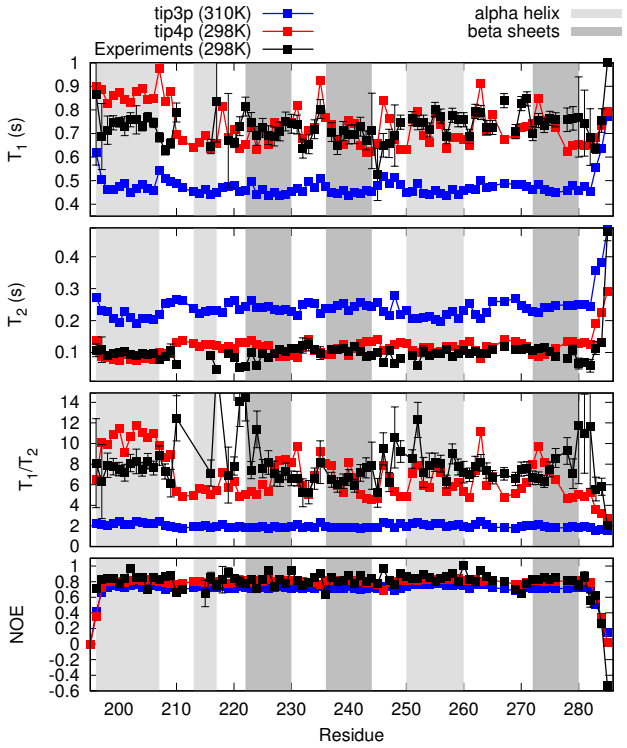

FIG. 3. Spin relaxation times for *Hp*TonB from experiments [**?**] and simulations with different water models.

significantly off from experiments and underestimate T_1/T_2 ratio, suggesting too fast overall rotational diffusion dynamics [53]. This is in agreement with previous study, where overestimation of rotational diffusion was attributed to the overestimated self diffusion of tip3p [25]. On the other hand, simulation results with tip4p water model show signifigantly better agreement with experiments in Fig. 3.

To see if the discrepancy in spin relaxation times for simulations with tip3p water model could be explained purely by overestimated protein overall diffusion, we scaled the diffusion constants by a constant factor of 2.9 before applying Eq. 10 to calculate the new correlation functions. The spin relaxation times calculated from correlation functions with the scaled rotational diffusion constants are shown in Fig. 4 together with experimental values. Indeed, the reduction of overall diffusion constants with a scaling factor brings the spin relaxation times in good agreement with experiments.

Spin relaxation times are compared between experiments and simulations with different water models and temperatures for *Pa*TonB construct in Fig. 5. Simulations with both tip4p or OPC4 water models systemically underestimate T_1 values and T_1/T_2 ratio when compared with experiments. The discrepancy is, however, less severe than for *Hp*TonB simulated with tip3p in Fig. 3. Spin relaxation times calculated from the new correlation function in Eq. 10 after scaling the diffusion coefficients with a constant factor of 1.2 are shown for *Pa*TonB simulation with tip4p at 298K in Fig. 6. The results are in good agreement with experiments, suggesting that the the scaling of overall diffusion constants with a constant factor removes the discprepancy with experimental spin relaxation data also for this system. Notably, the effect of temperature difference of 12 degrees on spin relaxation times is significanlty smaller than the difference between simulations and experiments or the effect of the diffusion constant scaling.

The scaled diffusion coefficients together with the simulation data lead to a good agreement between experimental spin relaxation data as seen in Figs. 4 and 6. This suggests that the scaled diffusion constant values can be used to interpret the anisotropic diffusion observed in NMR experiments. The rotational diffusion coefficients giving the best agreement with experimental data are collected in Table II. In contrast to the diffusion constants in Table I, these results are in line with previously reported values for more isotropic proteins [52].

C. Interpretation of protein internal relaxation from MD simulations

Good agreement between experimental spin relaxation times and MD simulation results with scaled overall rotational

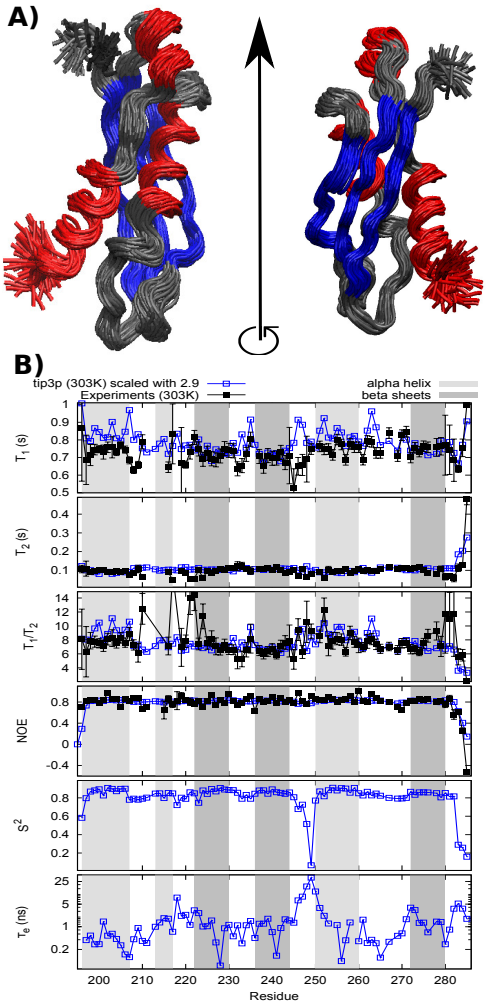

FIG. 4. A) Structures sampled by *Hp*TonB from MD simulations with tip3p at 303 K (100 structures from 400ns long trajectory). Secondary structures are colour labelled with Visual Molecular dynamics [54, 55]; α-helices are red and β-sheets are blue. B) Spin relaxation times from experiments and tip3p 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

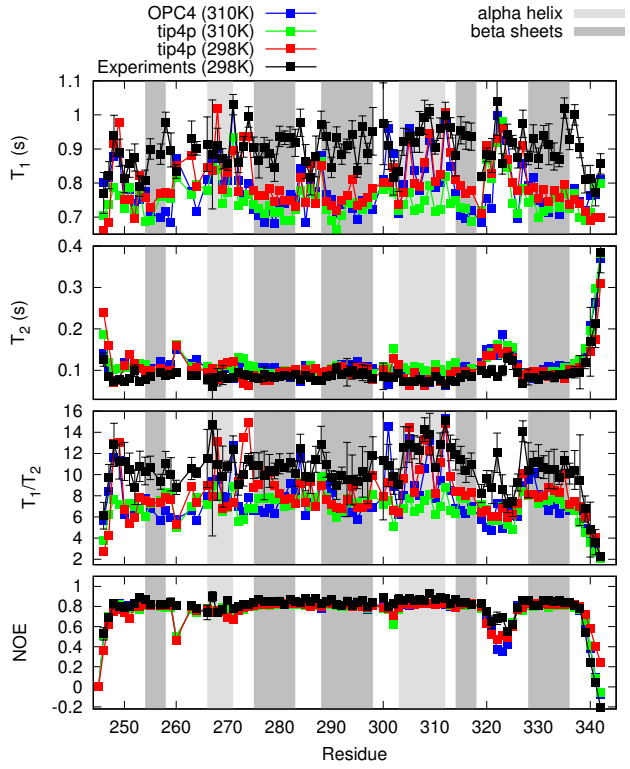

FIG. 5. Spin relaxation times for *Pa*TonB from experiments [**?**] and simulations with different water models.

TABLE II. Rotational diffusion coefficients giving the best agreement with experimental spin relaxation data. For *Hp*TonB construct the values calculated from simulation with tip3p were scaled with 2.9 (spin relaxation data in Fig. 4) and for *Pa*TonB the values from tip4p simulation at 298K were scaled by 1.2 (spin relaxation data in Fig. 6)

	*Hp*TonB	*Pa*TonB
D_{xx}	2.15 ± 0.01	1.51 ± 0.01
D_{yy}	2.43 ± 0.01	1.72 ± 0.03
D_{zz}	4.10 ± 0.01	3.79 ± 0.03
D_{av}	2.90 ± 0.03	2.3 ± 0.02
τ_c(ns)	5.7 ± 0.1	7.2 ± 0.1

diffusion constants in Figs. 4 and 6 suggest that the simulations can be used to interpret the internal dynamics of proteins observed in spin relaxation experiments.

Only small variations between different residues are observed for spin relaxation times of *Hp*TonB construct in Fig. 3. This indicates a rather rigid protein structure, which is also seen in MD simulation snapshots overlayed in Fig. 3 A). Only few residues in the terminal ends show slightly enhanced conformational sampling in MD simulations and in spin relaxation data. Deviations from average spin relaxation times are also observed in experiements close to residues 210-222, where also difficulties in peak assignment were observed [28]. Simulations for *Hp*TonB construct do not reveal explanation for this, however, similar region in *Pa*TonB construct shows sampling between two orientations of α-helix (see discussion below). Residues 245-250 in *Hp*TonB give exceptionally low order parameters and long effetive correlation times in simulations and small T_1 times in experiments. The interpretation of this is, however, not straightforward, because low T_1 values in this region are not reproduced by simulations.

More variety in internal dynamics between residues is observed for *Pa*TonB protein and segments with enhanced conformational sampling are labelled with yellow colour in Fig. 6. Larger number of sampled conformations in both terminal ends are characterized by low order parameters and long effective internal correlation times observed in simulations. En-

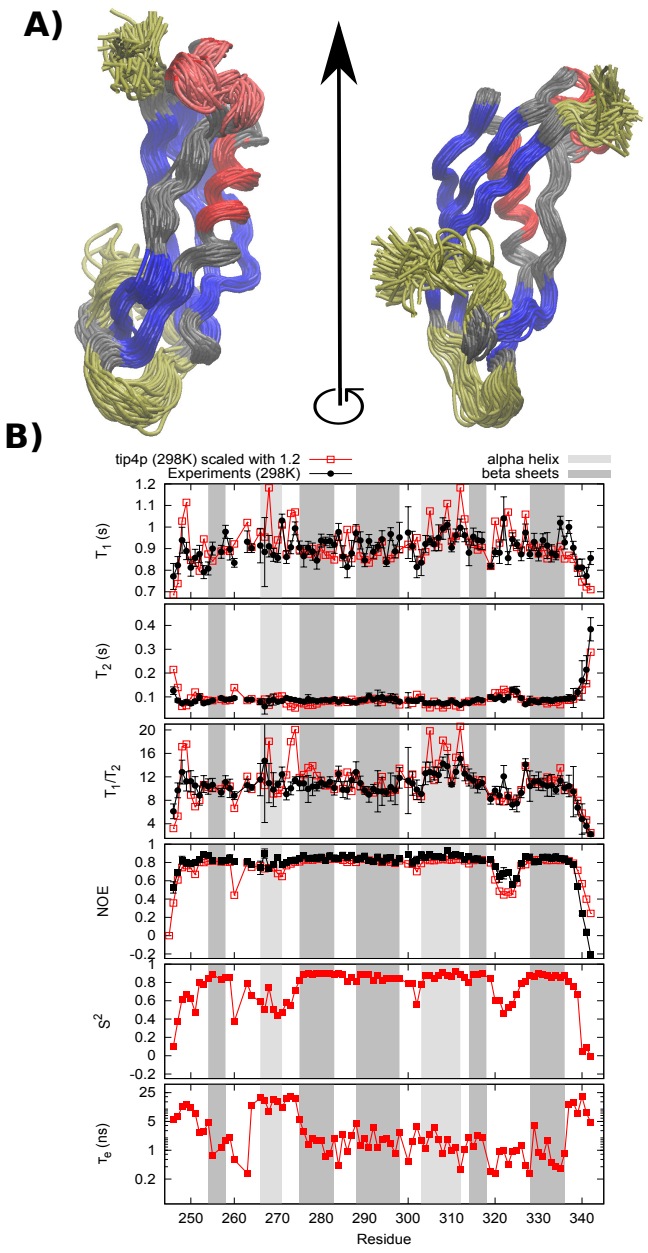

FIG. 6. A) Structures sampled by *Pa*TonB from MD simulations with tip4p at 298 K (100 structures from 400ns long trajectory). Secondary structures are colour labelled with Visual Molecular dynamics [54, 55]; α-helixces are red and β-sheets are blue. Residues 246-251, 320-326 and 338-342 with increased internal dynamics are yellow and α-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.

hanced conformational sampling is also observed for residues between 320-326, which is a loop between two β-sheets. MD simulations predict low order parameters and long internal effective correlation times are also for residues between 260-274. These can be explained by two different orientations sampled by the α-helix in this region (colour labelled with pink in Fig. 6 A). The orientational sampling of the similar short helix could also explain lower spectral resolution and deviations in spin relaxation times observed for residues close to 210-222 in *Hp*TonB [28].

MD simulations can be used to suggest how rotational dynamics of N-H bonds builds up from different components. Here we fit a sum of 471 different timescales to the correlation functions according to Eq. 11. Most prefactor values are zero in all correlation functions in this work, thus the remaining timescales with non-zero prefactors are considered as different components of the total relaxation process. The prefactors are shown in Fig. 7 for the same residues of *Pa*TonB, which were used to exemplify correlation functions in Fig. 2. As expected for residue 322 in rigid β-sheet with large order parameter value, the rotational relaxation is dominated by timescales of $\sim$5.5 ns and $\sim$8 ns, matching with protein overall rotation. Also the dynamics of the selected residue 322 in flexible loop is dominated by timescales around $\sim$8 ns corresponding protein overall rotation, however, fast motions from internal dynamics are more significant than for the rigid β-sheet residue. This is in agreement with lower order parameter value in the flexible loop residues. On the other hand, the rotational dynamics of the selected residue 341 in the flexible N-terminus is dominated by timescales below 3 ns, most likely related to the internal protein dynamics. Also the contribution from timescales close to $\sim$13 ns probably arises from slow conformational sampling of N-terminus, rather than overall rotational dynamics. This supports the conclusion that sampling of large amount of conformations leads to small order parameters and large effective correlation times observed in Fig. 5. While the results are very coherent, it should be kept in mind that the fractionalization of rotational dynamics to components is based on fitting of multiexponential sum to the simulation data and the solution of such fit is not unique.

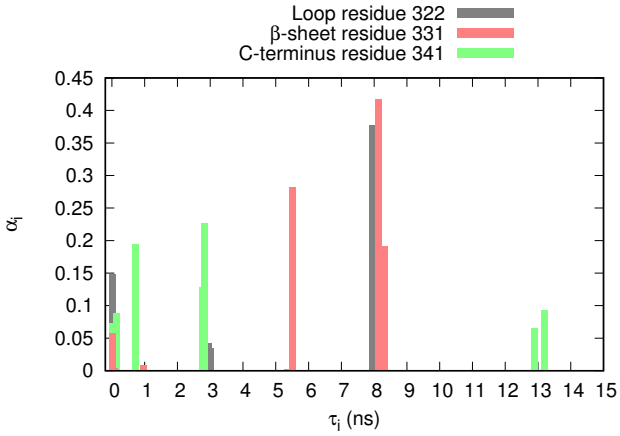

FIG. 7. Prefactors α_i corresponding different timescales τ_i resulting from a fit of Eq.11 to correlation functions from MD simulation of *Pa*TonB at 298K. The used correlation functions give a good agreement with experimental spin relaxation times as shown in Fig. 6.

IV. DISCUSSION AND CONCLUSIONS

Experimental spin relaxation data for protein backbone N-H bonds was successfully reproduced by using classical molecular dynamics simulation data for two different proteins with anisotropic molecular shape. Thus, the simulation trajectories give an atomistic resolution interpretation for the protein dynamics measured with NMR experiments. Interpretation of overall and internal dynamics was demonstrated for two relatively rigid proteins. The proposed approach is expected to be especially useful in the interpretation of spin relaxation data of proteins containinig some interesting flexible domains and having an anisotropic shape. Interpretation of relaxation data measured from such proteins has been very challenging with the previously available methods [56].

Direct analysis of classical molecular dynamics trajectories did not, however, reproduce the experimental data. Comparison of calculated diffusion constants and spin relaxation times with experimental data suggested that the overall brownian tumbling of protein is too rapid in simulations, in agreement with previous studys suggesting that the discrepance arises from the inaccuracise in water model [25]. Scaling down the diffusion coefficients in simulation data with the same factor in all directions led to good agreement with experiments and allowed the interpretation. The overall rotation of proteins was found to be brownian having only a small subdiffusive behaviour with short timescales below $\sim$0.12 ns. This can be constrasted with crowded environments, where anomalous diffusion is expected to be more significant [57].

Similarity between correlation functions from the original MD trajectory and the new correlation functions from Eq. 10 suggest that the usage of intertia axes and the separation of internal and overall rotational motions (Eq. 6) are good approximations for the studied proteins. However, rather rigid proteins are studied in this work and in previous publications with similar conclusion [10, 25]. These assumptions become less rigorous for proteins with larger fractions of intrinsically disoredered segments. In such cases the option might be to use isotropic reorientational eigenmode dynamics (iRED) [24] or quaternions [26]. However, the correction of overall rotational diffusion artefact due to water model would be complicated for intrinsically ordered molecules in all the discussed approaches. Thus, a water model giving correct overall rotational diffusion rates for biomolecules is probably necessary to compare simulations of intrinsically disordered molecules to experimental spin relaxation data.

Overall rotational diffusion coefficients were overestimated by a factor of $\sim$3 in *Hp*TonB simulations with tip3p water model, in agreement with previous studies [24–26]. Simulations with tip4p and opc4 water models gave spin relaxation times in reasonable agreement with experiments with scaling factors of $\sim$1-1.2, which is significantly less than for tip3p. It should be noted, however, that the correct scaling factor for tip4p seems to slightly different for *Hp*TonB and *Pa*TonB suggesting that the correction is not fully determined by bulk water properties. Thus, the role of hydration layer and water-protein interactions in rotational dynamics is also expected to be significant.

ACKNOWLEDGMENTS

We acknowledge CSC-IT center for science for computational resources

[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] F. A. Mulder, A. Mittermaier, B. Hon, F. W. Dahlquist, and L. E. Kay, Nat Struct Mol Biol **8**, 932 (2001).

[4] E. Z. Eisenmesser, O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, J. J. Bosco, Daryl A. Skalicky, L. E. Kay, and D. Kern, Nature **438**, 117 (2005).

[5] V. den Bedem H and F. JS., Nat. Methods **12**, 307 (2015).

[6] J. R. Lewandowski, M. E. Halse, M. Blackledge, and L. Emsley, Science **348**, 578 (2015).

[7] J. M. Lamley, M. J. Lougher, H. J. Sass, M. Rogowski, S. Grzesiek, and J. R. Lewandowski, Phys. Chem. Chem. Phys. **17**, 21997 (2015).

[8] D. Yang and L. E. Kay, Journal of Molecular Biology **263**, 369 (1996).

[9] V. Kasinath, K. A. Sharp, and A. J. Wand, Journal of the American Chemical Society **135**, 15092 (2013).

[10] O. Allnér, N. Foloppe, and L. Nilsson, The Journal of Physical Chemistry B **119**, 1114 (2015).

[11] M. Akke, R. Brüschweiler, and A. G. Palmer, Journal of the American Chemical Society **115**, 9832 (1993).

[12] C. Sanchez-Medina, A. Sekhar, P. Vallurupalli, M. Cerminara, V. Muoz, and L. E. Kay, Journal of the American Chemical Society **136**, 7444 (2014).

[13] R. B. Best and M. Vendruscolo, Journal of the American Chemical Society **126**, 8090 (2004).

[14] S. A. Showalter and B. Rafael, Journal of Chemical Theory and Computation **3**, 961 (2007).

[15] S. A. Showalter, E. Johnson, M. Rance, and R. Brüschweiler, Journal of the American Chemical Society **129**, 14146 (2007).

[16] P. Maragakis, K. Lindorff-Larsen, M. P. Eastwood, R. O. Dror, J. L. Klepeis, I. T. Arkin, M. . Jensen, H. Xu, N. Trbovic, R. A. Friesner, A. G. Palmer, and D. E. Shaw, The Journal of Physical Chemistry B **112**, 6155 (2008).

[17] N. Trbovic, B. Kim, R. A. Friesner, and A. G. Palmer, Proteins: Structure, Function, and Bioinformatics **71**, 684 (2008).

[18] H. Wennerstroem, B. Lindman, O. Soederman, T. Drakenberg, and J. B. Rosenholm, Journal of the American Chemical Society **101**, 6860 (1979).

[19] G. Lipari and A. Szabo, J. Am. Chem. Soc. **104**, 4546 (1982).

[20] D. E. Woessner, The Journal of Chemical Physics **37**, 647 (1962).

[21] H. Shimizu, The Journal of Chemical Physics **37**, 765 (1962).

[22] P. Luginbühl, K. V. Pervushin, H. Iwaï, and K. Wüthrich, Biochemistry **36**, 7305 (1997).

[23] J. Blake-Hall, O. Walker, and D. Fushman, "Characterization of the overall rotational diffusion of a protein from 15n relax-

ation measurements and hydrodynamic calculations," in *Protein NMR Techniques*, edited by A. K. Downing (Humana Press, Totowa, NJ, 2004) pp. 139–159.
[24] J. J. Prompers and R. Brüschweiler, Journal of the American Chemical Society **124**, 4522 (2002).
[25] V. Wong and D. A. Case, The Journal of Physical Chemistry B **112**, 6013 (2008).
[26] J. S. Anderson and D. M. LeMaster, Biophysical Chemistry **168**, 28 (2012).
[27] C.-Y. Lu and D. A. V. Bout, The Journal of Chemical Physics **125**, 124701 (2006).
[28] A. Ciragan, A. S. Aranko, I. Tascon, and H. Iwaï, Journal of Molecular Biology **428**, 4573 (2016).
[29] A. Abragam, *The Principles of Nuclear Magnetism* (Oxford University Press, 1961).
[30] L. E. Kay, D. A. Torchia, and A. Bax, Biochemistry **28**, 8972 (1989).
[31] Y. Hiyama, C. H. Niu, J. V. Silverton, A. Bavoso, and D. A. Torchia, Journal of the American Chemical Society **110**, 2378 (1988).
[32] B. Halle, The Journal of Chemical Physics **131**, 224507 (2009).
[33] $\tau_1 = (4D_{xx} + D_{yy} + D_{zz})^{-1}$, $\tau_2 = (D_{xx} + 4D_{yy} + D_{zz})^{-1}$, $\tau_3 = (D_{xx} + D_{yy} + 4D_{zz})^{-1}$, $\tau_4 = [6(D + (D^2 - L^2)^{-1/2}]^{-1}$, $\tau_5 = [6(D - (D^2 - L^2)^{-1/2}]^{-1}$, $D = \frac{1}{3}(D_{xx} + D_{yy} + D_{zz})$ and $L^2 = \frac{1}{3}(D_{xx}D_{yy} + D_{xx}D_{zz} + D_{yy}D_{zz})$.
[34] P. Dosset, J.-C. Hus, M. Blackledge, and D. Marion, Journal of Biomolecular NMR **16**, 23 (2000).
[35] J. G. de la Torre, M. Huertas, and B. Carrasco, Journal of Magnetic Resonance **147**, 138 (2000).
[36] O. Fisette, P. Lage, S. Gagn, and S. Morin, Journal of Biomedicine and Biotechnology **2012**, 254208 (2012).
[37] Y. Gu, D.-W. Li, and R. Brüschweiler, Journal of Chemical Theory and Computation **10**, 2599 (2014).
[38] M. J. Abraham, T. Murtola, R. Schulz, S. Pll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **12**, 19 (2015).
[39] 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).
[40] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[41] S. Izadi, R. Anandakrishnan, and A. V. Onufriev, The Journal of Physical Chemistry Letters **5**, 3863 (2014).
[42] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126** (2007).
[43] M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).
[44] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[45] U. L. Essman, M. L. Perera, M. L. Berkowitz, T. Larden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[46] B. Hess, J. Chem. Theory Comput. **4**, 116 (2008).
[47] M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS user manual version 5.0.7* (2015).
[48] 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).
[49] "Matlab, r2016a, the mathworks, inc., natick, massachusetts, united states.".
[50] A. Nowacka, N. Bongartz, O. Ollila, T. Nylander, and D. Topgaard, J. Magn. Res. **230**, 165 (2013).
[51] T. M. Ferreira, O. H. S. Ollila, R. Pigliapochi, A. P. Dabkowska, and D. Topgaard, J. Chem. Phys. **142**, 044905 (2015).
[52] V. Krishnan and M. Cosman, Journal of Biomolecular NMR **12**, 177 (1998).
[53] W. R. Carper and C. E. Keller, The Journal of Physical Chemistry A **101**, 3246 (1997).
[54] D. Frishman and P. Argos, Proteins: Structure, Function, and Bioinformatics **23**, 566 (1995).
[55] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics **14**, 33 (1996).
[56] G. Barbato, M. Ikura, L. E. Kay, R. W. Pastor, and A. Bax, Biochemistry **31**, 5269 (1992).
[57] F. Hfling and T. Franosch, Reports on Progress in Physics **76**, 046602 (2013).

SUPPLEMENTARY INFORMATION

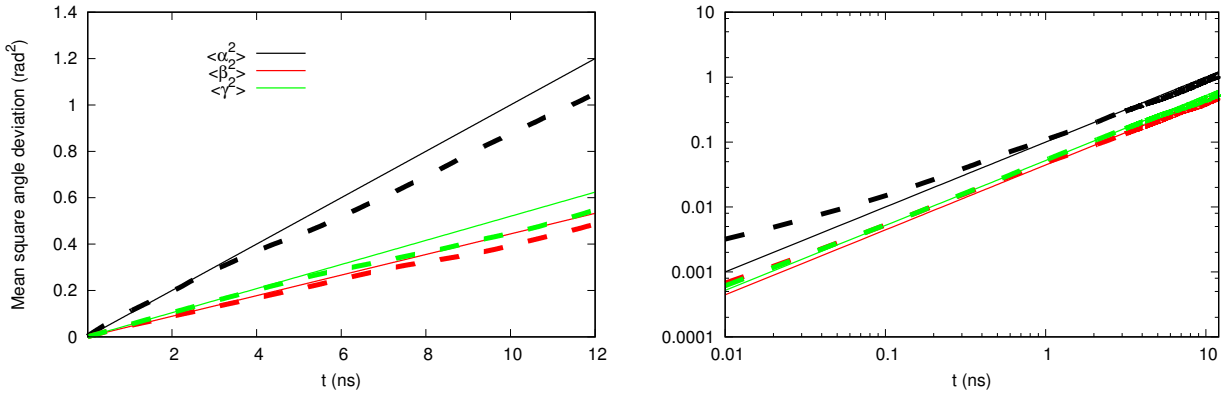

FIG. 8. The intertia tensor angles as a function of time and mean square angular deviations for *Pa*TonB simulation with tip4p water model at 310K.

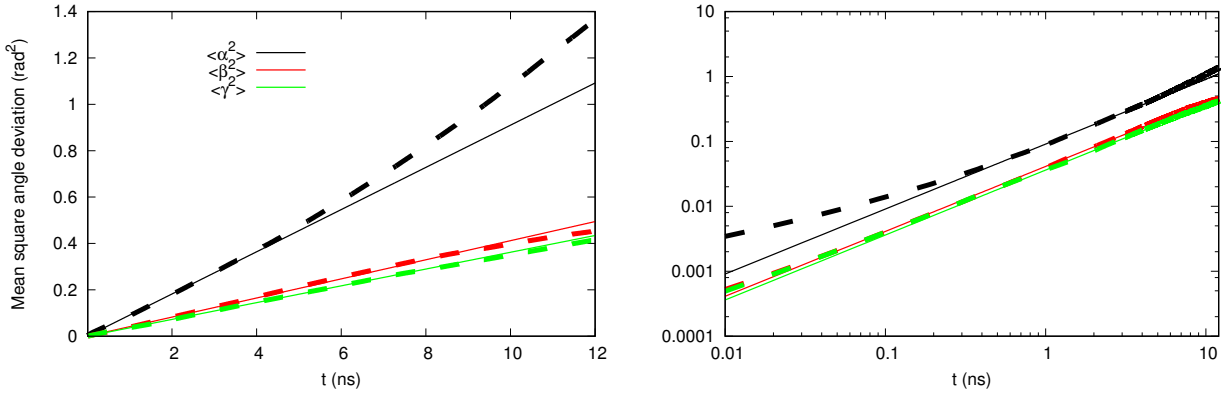

FIG. 9. The intertia tensor angles as a function of time and mean square angular deviations for *Pa*TonB simulation with tip4p water model at 298K.