Removing Thermostat Distortions of Protein Dynamics in Constant-Temperature Molecular Dynamics Simulations

Alan Hicks¹, Matthew MacAinsh¹, and Huan-Xiang Zhou^{1,2,*}

¹Department of Chemistry and ²Department of Physics, University of Illinois at Chicago, Chicago, IL 60607, USA

ABSTRACT Molecular dynamics simulations are widely used to determine equilibrium and dynamic properties of proteins. Nearly all simulations nowadays are carried out at constant temperature, with a Langevin thermostat among the most widely used. Thermostats distort protein dynamics, but whether or how such distortions can be corrected has long been an open question. Here we show that constant-temperature simulations with a Langevin thermostat dilate protein dynamics and present a correction scheme to remove the dynamic distortions. Specifically, ns-scale time constants for overall rotation are dilated significantly but sub-ns time constants for internal motions are dilated modestly, while all motional amplitudes are unaffected. The correction scheme involves contraction of the time constants, with the contraction factor a linear function of the time constant to be corrected. The corrected dynamics of eight proteins are validated by NMR data for rotational diffusion and for backbone amide and side-chain methyl relaxation. The present work demonstrates that, even for complex systems like proteins with dynamics spanning multiple timescales, one can predict how thermostats distort protein dynamics and remove such distortions.

The correction scheme will have wide applications, facilitating force-field parameterization and propelling simulations to be on par with NMR and other experimental techniques in determining dynamic properties of proteins.

TOC graphic

1. INTRODUCTION

Molecular dynamics (MD) simulations are now an indispensable tool for characterizing the thermodynamic and dynamics properties of biomolecular systems. To achieve realism and facilitate comparison with experiments, modern simulations are all done at constant temperature. A number of thermostat algorithms have been developed¹⁻⁸. One of the most popular choices is Langevin dynamics⁵, where two terms are added to Newton's equation of motion for each degree of freedom of the system:

$$m\ddot{x} = F - \zeta m\dot{x} + R(t) \quad [1]$$

One of the added terms, $R(t)$, is a stochastic force that thermalizes the system whereas the other, $-\zeta m\dot{x}$, is a frictional force that damps the velocity when the temperature is too high. In the limit where the damping constant $\zeta \rightarrow 0$, both of the added terms vanish and one recovers the system-only equation of motion, which leads to constant energy (constant NVE) rather than constant

temperature (constant NVT). Over the years, a number of artifacts of thermostats have been discovered and addressed, leading some to call thermostats a “necessary evil”⁹. These include the “flying ice cube” problem (hot solvent-cold solute)¹⁰; non-ergodicity⁶; and distortion of configuration-space distributions¹¹. It is important to recognize that all thermostats are only designed to produce the correct equilibrium thermodynamic properties at a given temperature. All thermostats inevitably perturb the dynamic (i.e., time-dependent) properties of the biomolecular systems¹²⁻¹⁴. There is only sporadic attention paid to this crucial issue^{7, 12, 15-18}. As MD simulations more and more routinely reach microseconds and beyond and direct comparison with experimental data on macromolecular dynamics become more and more practical, it is becoming pressing to assess and correct the dynamic distortions of thermostats. Here we address this problem for the Langevin thermostat.

Early studies identified the influences of a Langevin thermostat on the dynamics of small molecules. Loncharich et al.¹⁵ showed that the isomerization rate of N-acetyl-methylaldehyde was dictated by the damping constant ζ . In line with Kramers’ theory¹⁹, the isomerization rate reached a maximum at an intermediate ζ (around 2 ps⁻¹) and became inversely proportional to ζ in high-friction ($\zeta > 10$ ps⁻¹) Langevin simulations. Similarly, the translational diffusion constants of neon and water decreased with increasing ζ ¹⁶. Tobias et al.¹² compared NVE and NVT simulations and found that, at a small ζ (= 2 ps⁻¹), Langevin and Nose-Hoover thermostats did not alter the sub-ps underdamped time correlation function of the Tyr41 χ_2 torsion angle in bovine pancreatic trypsin inhibitor (BPTI). However, this correlation function became overdamped in high-friction ($\zeta = 20$ or 50 ps⁻¹) Langevin simulations. Bussi et al.⁷ also presented evidence of dynamic distortions when their thermostat, involving stochastic velocity rescaling, was at a high damping constant. Basconi and Shirts¹⁷ investigated the effects of

several thermostats on the transport dynamics of water and conformational dynamics of a coarse-grained homopolymer. They stated that, with stronger coupling to the thermal bath (e.g., by increasing the damping constant ζ of a Langevin thermostat), “the dynamics will begin to diverge from the NVE dynamics.” In particular, water diffusion and homopolymer chain decorrelation were both slowed down by several-fold when a Langevin thermostat at $\zeta = 10 \text{ ps}^{-1}$ was used. Debiec et al.¹⁸ reported ~50% increases in the overall rotational correlation times of three proteins (the third immunoglobulin-binding domain of protein G (GB3), ubiquitin, and binase) in Langevin simulations (at $\zeta = 1 \text{ ps}^{-1}$) compared with NVE simulations. Recognizing the conflicting effects of thermostats in achieving thermal equilibration and producing dynamic distortions, Leimkuhler et al.¹⁴ proposed using the ratio of these two effects as a measure for selecting an optimal thermostat.

Many efforts in force-field parameterization have included experimental data of dynamic properties for parameter tuning. Translational and rotational diffusion constants are often used in parameterizing water models, as illustrated by the re-parameterization of the SPC/E water model into SPC/Eb²⁰. NMR spin relaxation properties have long been used to validate force fields²¹⁻²⁴. In particular, Hoffmann et al.²³ tuned side-chain methyl rotation energy barriers in the AMBER ff99SB*-ILDN force field²⁵⁻²⁷, and used NMR relaxation of methyls in ubiquitin²⁸⁻²⁹ to validate calculations in NVE and NPT simulations (with Bussi et al.’s stochastic velocity rescaling⁷ as the thermostat). This study did not reveal any significant effects of the thermostat and barostat on methyl axis order parameters and correlation times of fast internal motions. However, in theory all thermostats distort protein dynamics, and hence force-field parameters might be tuned in part to compensate for dynamic distortions of thermostats. Moreover, the extent of dynamic distortion

may depend on the particular protein and on the thermostat used, therefore raising serious issues on force-field transferability.

For globular proteins, the model-free approach of Lipari and Szabo³⁰ and its extension³¹ have been highly successful in analyzing NMR spin relaxation, for data from experiments as well as from MD simulations. This approach assumes separation between timescales of overall rotation and internal motions. For intrinsically disordered proteins, we and others carried out NVT simulations using a Langevin thermostat and borrowed the mathematical form of the model-free approach to analyze NMR relaxation³²⁻³⁴. Specifically, the time correlation function of a backbone NH bond vector, $C(t) = \langle P_2(\mathbf{n}(t') \cdot \mathbf{n}(t' + t)) \rangle_{t'}$, was fit to a sum of exponential functions of time. Here P_2 is the second-order Legendre polynomial and $\mathbf{n}$ denotes the unit vector along the NH bond. In our study of ChiZ³⁴, we used three exponentials, with time constants (τ_i ; $i = 1$ to 3) at ~ 10 ns, ~ 2 ns, and ~ 300 ps, as well as an ultrafast (~ 10 ps) component. The observed patterns of the longitudinal (R_1) and transverse (R_2) relaxation rates along the amino-acid sequence were predicted well, but the MD simulations systematically underestimated R_1 and overestimated R_2 , suggesting that the time constants, especially the longest one, were too long. By contracting, i.e., scaling down, the time constants by a factor $1 + b\tau_i$, we significantly improved the agreement between the relaxation properties calculated from the MD simulations and the experimental data. Dynamic distortion by the Langevin thermostat was suggested as one of the reasons for the need of the contraction in τ_i .

This study was designed to fully assess the dynamic distortions of the Langevin thermostat and develop a scheme to correct such distortions. We chose to study a set of eight globular proteins of various sizes (with the number of amino acids ranging from 18 to 129; Fig. 1). Globular proteins were chosen because their dynamics involve motions on different

timescales and the nature of these motions (i.e., overall rotation and internal motions) is well-known. A wide range of protein sizes was chosen because we suspected that the extent of Langevin distortions is dependent on protein size. This suspicion originated from the time contraction factor introduced for ChiZ, which as a linear function of τ_i was greater for a mode with a longer time constant than for a mode with a shorter time constant; the former mode presumably involved the correlated motion of a larger group of atoms. Here dynamic distortions of the Langevin thermostat was identified by comparing against NVE simulations. The Langevin thermostat is found to dilate the time constants, but leave the motional amplitudes unaffected. Moreover, applying a contraction factor, $a + b\tau_i$, to the time constants completely removes the dynamic distortions. The corrected NVT results for NMR relaxation properties of backbone amides and side-chain methyls agree well with experimental data.

2. COMPUTATIONAL METHODS

2.1. MD Simulations. Starting structures of the eight proteins were downloaded from the Protein Data Bank using the entry names listed in Fig. 1. Each structure was prepared for AMBER simulations³⁵ in *tleap* using the ff14SB protein force field³⁶ and TIP4P-D water model³⁷. The unit cell for each system was an orthorhombic box with a 15 Å solvent layer of solvent around the protein. Lastly protein-neutralizing ions (Na^+ or Cl^-) and NaCl (at the experimental salt concentrations) were added.

For each system, energy minimization (2000 steps of steepest descent and 3000 steps of conjugate gradient) was followed by short equilibration (100 ps of NVT and 3 ns of NPT). In the NVT portion, the temperature was ramped from 0 to the target value (Table S1) in the first 40 ps and the protein was restrained with a 1 kcal/mol/Å² force constant. Temperature was regulated by

the Langevin thermostat (damping constant at 1 ps^{-1}) and pressure was regulated by the Berendsen barostat ² at 1 atm. All MD simulations were carried out in AMBER18 using *pmemd.cuda* ³⁸.

Figure 1. The eight proteins studied in the present work. The Protein Data Bank entry and the total number of residues of each protein are given.

Four replicate simulations of each equilibrated system were run for 500 ns at constant NVE and 1000 ns at constant NVT. Snapshots were saved every 0.5 ps. Bond lengths involving hydrogens were constrained using the SHAKE algorithm ³⁹. Long-range electrostatic interactions were treated by the particle mesh Ewald (PME) method ⁴⁰. The cutoff distance for nonbonded

interactions were 10 Å. The timesteps were 1 fs in the NVE simulations and 2 fs in the NVT simulations. In the latter simulations, temperature was regulated by the Langevin thermostat, with the damping constant set to each of 3 values: 0.2 ps⁻¹, 2 ps⁻¹, and 20 ps⁻¹.

Replicates for NVE production simulations were generated by independent short NPT simulations following the equilibration stage. To control energy drift, very strict convergence levels (at 0.5×10^{-6}) were chosen for PME and SHAKE. If the energy and temperature still drifted due to known cutoff and numerical integration artefacts in NVE simulations¹³, then more replicate simulations were run. In most cases, at least eight NVE simulations were run and the four that had the least energy and temperature drifts were selected for analysis.

2.2. Secondary Structure and Solvent Accessible Surface Area (SASA). Secondary structures and SASAs were calculated using the *dssp* and *surf* commands in *cpptraj*⁴¹. The SASA for each residue was converted to the percent exposure as $(1 - \text{SASA}/\text{SASA}_{\text{max}}) \cdot 100\%$, where SASA_{max} was the nominal maximum SASA for a residue of that type (e.g., alanine).

2.3. Rotational Correlation Time (RCT). RCT was calculated using the *rotdif* command in *cpptraj* as described in Wong et al.¹³. In brief, 500 random unit vectors attached to the protein molecular frame were used to calculate time correlation functions. The latter entailed determining the dot product between each random vector, $\mathbf{v}(t_1)$, at time t_1 along a simulation and the same vector, $\mathbf{v}(t_2)$, at time t_2 , using the rotation matrices required to align the protein structures at t_1 and t_2 to a reference (chosen as the first saved snapshot). The second-order Legendre polynomial (i.e., P_2) of this dot product as a function of the lag time $t = t_2 - t_1$ was averaged (over the simulation and the random vector) and integrated to yield the RCT for isotropic tumbling. Finally, an average of RCTs was taken over the four replicate simulations.

2.4. NMR Spin Relaxation Properties. The backbone ^1H - ^{15}N NMR relaxation

properties were calculated from the NH bond correlation function, $C(t) = \langle P_2(\mathbf{n}(t') \cdot \mathbf{n}(t' + t)) \rangle_{t'}$. $C(t)$ was obtained using the *timecorr* function in *cpptraj*. After averaging over the four replicate simulations, it was fit to a bi-exponential function,

$$C(t) = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2} \quad [2]$$

The longer time constant τ_1 is for overall rotation whereas the shorter time constant τ_2 is for internal motion. We did not treat τ_1 as a global parameter but made it residue-specific, just as τ_2 was residue-specific. Nor did we constrain the sum of the amplitudes, $A_1 + A_2$, to 1. The fitting was performed using the trust region reflexive algorithm in the *scipy.optimize.curve_fit* module of python, with τ_2 bounded between 0.5 ps and 150 ps. The time constants τ_1 and τ_2 (as do the amplitudes A_1 and A_2) have large disparities. To ensure a proper balance between the portion of the data that determined the first exponential and the portion of the data that determined the second exponential, in the fitting we picked data points that were uniformly distributed on a log scale of time instead of a linear scale of time. Specifically, we used only data points with indices given by $\lceil 1.02^i \rceil$, where $\lceil x \rceil$ denotes the ceiling function (i.e., the smallest integer greater than or equal to x) and $i = 0, 1, 2, \dots$

From the bi-exponential fit, the spectral density was found as

$$J(\omega) = \frac{A_1 \tau_1}{1 + (\tau_1 \omega)^2} + \frac{A_2 \tau_2}{1 + (\tau_2 \omega)^2} \quad [3]$$

Finally R_1 , R_2 , and NOE were given by

$$R_1 = f_{\text{DD}}[J(\omega_{\text{H}} - \omega_{\text{N}}) + 3J(\omega_{\text{N}}) + 6J(\omega_{\text{H}} + \omega_{\text{N}})] + f_{\text{CSA}}J(\omega_{\text{N}}) \quad [4]$$

$$R_2 = \frac{R_1}{2} + f_{\text{DD}}[2J(0) + 3J(\omega_{\text{H}})] + \frac{2}{3}f_{\text{CSA}}J(0) \quad [5]$$

$$\text{NOE} = 1 + \frac{f_{\text{DD}}}{R_1} \frac{\gamma_{\text{H}}}{\gamma_{\text{N}}} [6J(\omega_{\text{H}} + \omega_{\text{N}}) - J(\omega_{\text{H}} - \omega_{\text{N}})] \quad [6]$$

where $f_{\text{DD}} = \frac{1}{10} \left(\frac{\mu_0 \hbar \gamma_{\text{N}} \gamma_{\text{H}}}{4\pi r_{\text{NH}}^3} \right)^2$ and $f_{\text{CSA}} = \frac{2}{15} \omega_{\text{N}}^2 \Delta_{\text{CSA}}^2$. The meanings of the symbols are: μ_0 , permittivity of free space; $\hbar$, reduced Plank constant; γ_{N} and γ_{H} , gyromagnetic ratios of nitrogen and hydrogen; $\omega_{\text{H}} = \gamma_{\text{H}} B_0$, Larmor frequency of hydrogen; ω_{N} , counterpart of nitrogen; r_{NH} , NH bond length (set at 1.02 Å); and Δ_{CSA} ($= -170$ ppm), chemical shift anisotropy of nitrogen.

The A_1 amplitude of the bi-exponential fit can be directly interpreted as the order parameter S^2 (the A1 method). Order parameters can also be determined after removing overall rotation by structural alignment. Alignment was done with each block of an MD trajectory. The block length should be between τ_1 and τ_2 , and was chosen to be 1 ns for GB3 and ubiquitin. Longer block lengths were also tested. After alignment with the first snapshot of the block, the order parameter was calculated as an average second-order Legendre polynomial ⁴²:

$$S^2 = \frac{3}{2} (< x^2 >^2 + < y^2 >^2 + < z^2 >^2 + < xy^2 >^2 + < xz^2 >^2 + < yz^2 >^2) - \frac{1}{2} \quad [7]$$

where (x, y, z) denote the Cartesian components of the unit vector along an NH bond, and $< \dots >$ signifies averaging over the block. A final average of the S^2 values from the different blocks was taken. We call the foregoing procedure the P2 method.

Relaxation properties for side-chain methyls were calculated in much the same way. The CH correlation functions were averaged over the three hydrogens in the methyl, and fit to a bi-exponential function (with τ_2 bounded between 0.5 ps and 350 ps) to produce the spectral densities. The in-phase transverse and longitudinal relaxation rates were finally calculated as

$$R(D_y) = \frac{1}{80} \left(\frac{e^2 q Q}{\hbar} \right)^2 [9J(0) + 15J(\omega_D) + 6J(2\omega_D)] \quad [8]$$

$$R(D_z) = \frac{3}{40} \left(\frac{e^2 q Q}{\hbar} \right)^2 [J(\omega_D) + 4J(2\omega_D)] \quad [9]$$

with the quadrupolar coupling constant $\frac{qQe^2}{\hbar}$ set to $167 \cdot 2\pi$ kHz⁴³ and ω_D denoting the Larmor frequency of deuterium. Methyl axis order parameters, S_{axis}^2 , were calculated from CC correlation functions (A1 method).

2.5 Time-contraction scheme. The time-contraction factor was obtained as the ratio between the RCT of a set of replicate NVT simulations and the NVE counterpart (Eq [10]). A linear correlation analysis (Eq [11]) was then carried out in Excel between this factor and RCT_{NVT} , using pairs of such values for the eight proteins studied. Values of the a and b coefficients were obtained for the Langevin thermostat at three damping constants (0.2 ps^{-1} , 2 ps^{-1} , and 20 ps^{-1}), and then used to parameterize a and b as functions of the damping constant ζ (Eqs [13] and [14]).

3. RESULTS

For each of the eight proteins displayed in Fig. 1, we carried out four replicate NVE simulations and four replicate NVT simulations each with the Langevin thermostat at $\zeta = 0.2, 2, 20 \text{ ps}^{-1}$. The latter simulations are referred to as NVT – 0.2, NVT – 2, and NVT – 20, respectively. We analyzed three types of dynamic properties: rotational correlation times; backbone NH time correlation functions; and side-chain methyl time correlation functions. We used the NVE simulations as benchmarks to identify dynamic distortions of the Langevin thermostat and to find a correction scheme. The corrected NVT results are validated by experimental data.

3.1. Increase in RCTs by Langevin Thermostat, and a Tentative Correction. In Fig.

2A, we compare the RCTs from the NVE and NVT – 2 simulations against the experimental data for the eight proteins⁴⁴⁻⁵¹ (see Supporting Table S1). As expected, RCTs increase with protein size, for both the experimental data and for the NVE and NVT results. For example, the NVE results increase from ~1 ns for the two smallest proteins (θ -defensin and Trp-cage) to ~6 ns for the largest protein (hen egg-white lysozyme (HEWL)). These NVE results closely match the experimental values; a linear correlation analysis yielded a slope of 1.01 and an R^2 of 0.988 (not shown), both very close to an ideal value of 1. In contrast, the NVT – 2 simulations consistently overestimate the RCTs, with the extent of overestimating growing with protein size. Indeed, the NVT – 2 results also show a strong correlation with the experimental values ($R^2 = 0.989$), but with an exaggerated slope of 2.08 (Fig. 2A, green line).

Figure 2. Rotational correlation times (RCTs) and scaling coefficients. (A) RCTs from NVE simulations show excellent agreement with experimental values, whereas those from NVT simulations (at damping constant $\zeta = 2 \text{ ps}^{-1}$) are overestimated but, among the eight proteins, show strong correlation with the experimental values. The overestimation is completely removed after contraction by a factor $\alpha = a + b \cdot RCT_{NVT}$. The coefficients a and b were

obtained taking RCT_{NVT}/RCT_{NVE} as the contraction factor and doing a correlation analysis between it and RCT_{NVT} among the eight proteins. (B) Dependences on a and b on the damping constant ζ of the Langevin thermostat. Symbols are from correlation analysis; curves are functions, given by Eqs [13] and [14], designed to pass through the symbols. The values of the constants are $\alpha_1 = 0.09$; $\beta_1 = 1.11$; $\gamma_1 = 2.01$; $\alpha_2 = 0.059$; $\beta_2 = 0.19$; and $\gamma_2 = 0.62$.

It is clear that the Langevin thermostat dilates RCTs. A simple way to correct the dynamic distortion is to contract RCTs (and other time constants) calculated from Langevin simulations. We hypothesized that the ratio between the RCT calculated from NVT simulations and the NVE counterpart:

$$\alpha = \frac{RCT_{NVT}}{RCT_{NVE}} \quad [10]$$

could serve as the time-contraction factor. Interestingly, as we suspected, this correction factor shows a growing trend with increasing RCT_{NVT} (or protein size; Fig. S1). Analysis based on a linear correlation,

$$\alpha = a + b \cdot RCT_{NVT} \quad [11]$$

yielded $a = 1.31$ and $b = 0.078 \text{ ns}^{-1}$ for a Langevin thermostat at $\zeta = 2 \text{ ps}^{-1}$. For any time constant $\tau(\text{NVT})$ (including RCTs) calculated from NVT simulations, we apply a time-contraction factor $a + b\tau(\text{NVT})$, leading to the corrected NVT (cNVT for short) time constant

$$\tau(\text{cNVT}) = \frac{\tau(\text{NVT})}{a + b\tau(\text{NVT})} \quad [12]$$

Applying this correction to the RCTs of the NVT – 2 simulations produces results that are in excellent agreement with the experimental values. A linear correlation analysis yielded a slope of 1.01 and an R^2 of 0.996 (Fig. 2A, star symbols and red line).

To develop Eq [12] into a universal correction scheme for an arbitrary damping constant, we carried out additional Langevin simulations at $\zeta = 0.2$ and 20 ps^{-1} . The coefficients of the linear correlation in Eq [11] were found to be $a = 1.001$ and $b = 0.019 \text{ ns}^{-1}$ at $\zeta = 0.2 \text{ ps}^{-1}$ and $a = 4.33$ and $b = 0.113 \text{ ns}^{-1}$ at $\zeta = 20 \text{ ps}^{-1}$. The a and b values at the three ζ values were used to parameterize the following functional representations (Fig. 2B):

$$a = 1 + \alpha_1 \cdot \zeta^{\beta_1 + \frac{\gamma_1}{1+\zeta}} \quad [13]$$

$$b = \alpha_2 \cdot \zeta^{\beta_2 + \frac{\gamma_2}{1+\zeta}} \quad [14]$$

By design, these functions have the appropriate values, 1 for a and 0 for b , at $\zeta = 0$ (equivalent to constant NVE). Both the a and b functions then gradually increase with increasing ζ . We expect that these functions, when used in Eq [12], provide the necessary time contraction for the Langevin thermostat with ζ in the range from 0.1 to 50 ps^{-1} , which covers almost all the values of the damping constant found in simulations.

3.2 Langevin Thermostat Affects the Time Constants but Not Amplitudes of Backbone NH Correlation Functions. The time correlation function of a backbone NH bond involves protein dynamics on timescales ranging from ps to at least ns, and can be detected by NMR relaxation experiments. From the NVE and NVT simulations, we calculated the correlation functions of all the NH bonds in each of the eight proteins, and fit them to a sum of two exponentials (Eq [2]). NH correlation functions of three residues from NVT – 2 simulations of ubiquitin, along with the bi-exponential fits, are illustrated in Fig. S2. The amplitudes (A_1 and A_2) and time constants (τ_1 and τ_2 , with $\tau_1 > \tau_2$) of the fits, along with secondary structure and solvent percent exposure of each residue, are displayed in Fig. 3 for GB3 and ubiquitin and in Fig. S3 for the other six proteins.

Figure 3. Parameters of bi-exponential fits for NH time correlation functions. (A-E) are for GB3 and (F-J) are for ubiquitin. Results for NVE and NVT (at $\zeta = 2 \text{ ps}^{-1}$) simulations are shown. (A) & (F) Secondary structure (SS), according to the following color scheme: magenta, helix; green, β -sheet, yellow, turn; and white, coil. SS was assigned to the type with the highest propensity. (B) & (H) Percent of solvent exposure of each residue presented on a gray scale (black, buried; white, exposed). (C) & (H) Amplitudes A_1 and A_2 of the bi-exponential fits. Insets display the correlation functions for residue 28 in GB3 and residue 33 in ubiquitin. (D) & (I) Time constant τ_1 . The rotational correlation times calculated from these simulations are displayed as dashed horizontal lines. (E) & (J) Time constant τ_2 .

The NVE results can be summarized as follows. Most A_1 values are in the 0.8 to 0.9 range, whereas most A_2 values are in the 0 to 0.1 range. The sum $A_1 + A_2$ is fairly constant at

around 0.9, with the “missing” amplitude (~ 0.1) attributable to an ultrafast (sub-ps) component. For most residues, τ_1 values are close to the RCTs (1 to 6 ns) of the respective proteins, but can be much lower for flexible loops, e.g., residues 9-14 in GB3 (Figs. 3D and 1), or flexible termini, e.g., the C-terminus of ubiquitin (Figs. 3I and 1). These flexible residues, also distinguished by lower-than-average A_1 (and higher-than-average A_2), are outside α -helices and β -sheets and also solvent-exposed. Most τ_2 values fall in the 10-100 ps range, and do not seem to be correlated with secondary structure or solvent exposure.

Comparing these NVE fitting parameters with the NVT – 2 counterparts, a striking observation is that the Langevin thermostat does not affect the amplitudes. The NVE and NVT – 2 simulations are totally independent and yet, residue by residue, A_1 (A_2) values of the two types of simulations track each other extremely well. The most noticeable effect of the Langevin thermostat is on τ_1 values, which suffer a dilation similar to that described above for RCTs. Still, the Langevin thermostat preserves the profile of the τ_1 trace over residue number, as if suggesting a constant ratio between NVT – 2 and NVE τ_1 values among all the residues of a protein. For most residues, the Langevin thermostat also increases τ_2 values, though to a much smaller extent than for τ_1 .

The forgoing effects of the Langevin thermostat – great on τ_1 , moderate on τ_2 , and null on A_1 and A_2 – can also be detected by directly comparing NH correlation functions from NVE and NVT – 2 simulations. Residue 28 in GB3 represents a rigid residue, located in middle of the central helix (residues 23-36; Fig. 1). As shown in Fig. 3C inset, the NH correlation functions of this residue are essentially a single exponential in both NVE and NVT – 2 simulations, but the time constant, on the ns scale, is clearly increased in the latter simulations, as indicated by a significant rightward shift of the rapidly decaying part of the correlation function (on a log-linear

plot). Consistent with this qualitative difference, the bi-exponential fits yielded $A_1 = 0.91$ and a very small $A_2 = 0.01$ for both the NVE and NVT – 2 simulations, but with τ_1 values, at 3.45 ns and 6.78 ns, differing by nearly a factor of 2. These τ_1 values track closely the corresponding RCTs (3.22 ns and 6.09 ns) of GB3. On the other hand, residue 33 in ubiquitin represents a more flexible residue, located at the end of the major helix (Fig. 1). The NH correlation functions of this residue (Fig. 3H inset) show a clear bi-exponential decay. The first decays (on the 10 ps timescale) of the NVE and NVT – 2 correlation functions track each other closely. However, the second decays (on the ns timescale) deviate from each other, with the NVT – 2 time constant again about 2-fold longer than the NVE counterpart.

3.3 Dilation of τ_1 and τ_2 by Langevin Thermostat Grows with Increasing ζ . The benchmarking in backbone dynamics against NVE simulations has so far demonstrated that the Langevin thermostat at $\zeta = 2 \text{ ps}^{-1}$ dilates τ_1 significantly and τ_2 moderately, but leaves the amplitudes of the global and local motions intact. To attain a complete picture on the effects of the Langevin thermostat, we now report on the results of NVT – 0.2 and NVT – 20 simulations.

As illustrated by GB3 and ubiquitin (Fig. S4), as the damping constant expands to a wide range (from 0.2 to 20 ps^{-1}), the A_1 and A_2 amplitudes remain unaffected by the Langevin thermostat. Mirroring the results for RCTs (Fig. 2B and Table S1), the dilation in τ_1 is smaller at $\zeta = 0.2 \text{ ps}^{-1}$ than at $\zeta = 2 \text{ ps}^{-1}$, but reaches 10-fold at $\zeta = 20 \text{ ps}^{-1}$. As already suggested by results at $\zeta = 2 \text{ ps}^{-1}$, the Langevin thermostat preserves the profile of the τ_1 trace over residue number. As for τ_2 , at $\zeta = 20 \text{ ps}^{-1}$ there is clearly dilation for all but one or two residues in each protein, with the τ_2 profile over residue number also largely preserved.

In Figs. 4 and S5, we plot τ_1 and τ_2 values of each protein as violin plots, one each for NVE, NVT – 0.2, NVT – 2, and NVT – 20 simulations. Each violin plot displays the distribution

of τ_1 or τ_2 values for all the residues of a protein, as well as the quartiles and the mean value.

The relatively short heights of the τ_1 violins, spanning a small fraction of a decade, reflect the fact that most τ_1 values are close to the RCTs of the respective proteins. An exception is the outlying tail in Fig. 4B, due to the very flexible C-terminal residue in ubiquitin. For each protein, the mean τ_1 values increase monotonically from NVE to NVT – 20.

Figure 4. Dependences of the τ_1 and τ_2 time constants on the damping constant of the Langevin thermostat. NVE simulations correspond to $\zeta = 0$. (A) & (C) are for GB3 and (B) & (D) are for ubiquitin. (A) & (B) τ_1 values for each protein at a given ζ are displayed as a violin plot. Quartiles are indicated by two horizontal dashed lines in each violin plot. The mean τ_1 of

the protein is shown as a black diamond; these mean values at different ζ are connected by dashed lines between violin plots. The outlying tail in (B) arises from the excessively low τ_1 of the very C-terminal residue in ubiquitin. (C) & (D) Corresponding data for τ_2 . In addition, τ_2 values for individual residues are shown as a scatter plot, colored according to secondary structure.

The τ_2 violins, on the other hand, cover the entire decade from 10 to 100 ps and often go beyond in both directions. As already noted, τ_2 values do not seem to be correlated with secondary structure or solvent exposure. Here we further explore this matter by displaying the individual τ_2 values as a scatter plot, colored according to secondary structure. For the most part, secondary structures indeed appear randomly distributed in the scatter plots. An exception occurs in a few τ_2 violin plots with outlying tails, for Trp-cage, ribonuclease (RNase) T1, and HEWL in NVE and NVT – 0.2 simulations. These outlying tails originate from rigid residues located in α -helices or β -sheets. The fact that these τ_2 outliers occur in NVE and NVT – 0.2 simulations but not when the Langevin thermostat is applied with the damping constant at $\geq 2 \text{ ps}^{-1}$ suggest that rigid residues located in α -helices or β -sheets residues may be prone to being trapped in local minima in NVE and NVT – 0.2 simulations (see below).

Given the wide range of τ_2 values of each protein in a given set of simulations, the trend of mean τ_2 values on going from NVE simulation to NVT simulations with increasing damping constants is less clear-cut compared with the mean- τ_1 trend. The change in mean τ_2 could even be influenced by a few outliers (e.g., for ShK on going from NVE to NVT – 0.2). Still, there definitely is an overall increasing trend of mean τ_2 going from NVE to NVT – 20.

3.4. Correction for τ_1 and τ_2 Dilated by Langevin Thermostat. As concluded above, τ_1 traces from NVE and NVT simulations have similar profiles over residue number, and typical

τ_1 values track the corresponding RCTs. It thus seems obvious that we should apply Eq [12] to correct the dilation of τ_1 by the Langevin thermostat, where the a and b coefficients in the time contraction factor are determined by comparing RCTs of NVE and NVT simulations. In Fig. 5A we compare the thus corrected τ_1 for NVT – 2 (referred to as cNVT – 2) against the NVE benchmark. They show a strong correlation, with a near ideal slope of 1.01 and a very high R^2 of 0.996.

Figure 5. Comparison of corrected NVT time constants against the NVE results. (A) τ_1 values for all the residues in the eight proteins from NVT simulations (at $\zeta = 2 \text{ ps}^{-1}$), after correction by time contraction, and the NVE results are compared on a scatter plot. The correlation line and R^2 value are also shown. (B) Corresponding data for τ_2 , except the correlation analysis was done on a log instead of linear scale.

Encouraged by the excellent outcome of the τ_1 correction, we applied Eq [12], with the same a and b coefficients, to contract τ_2 . The resulting τ_2 is compared with the NVE benchmark in Fig. 5B. The agreement is also good, with a linear correlation showing a slope of 0.964 and R^2

of 0.982. Because the Langevin thermostat does not affect the A_1 and A_2 amplitudes, we expect NVT simulations after the time-constant contraction to represent the actual protein dynamics. Next we turn to experimental NMR relaxation data for validating cNVT.

3.5 cNVT Accurately Predicts Backbone NMR Relaxation Properties. The NH time correlation function determines the longitudinal (R_1) and transverse (R_2) relaxation rates of the ^{15}N spin and the ^1H - ^{15}N heteronuclear Overhauser effect (het. NOE). R_1 and R_2 are sensitive to ns dynamics and hence to A_1 and τ_1 , whereas NOE is sensitive to both ns dynamics and sub-ns dynamics (i.e., A_2 and τ_2).

As illustrated in Fig. S6A for GB3 and Fig. S6B for ubiquitin, R_1 , R_2 , and NOE calculated from NVE simulations are in very good agreement with experimental values^{47, 49}. The calculated R_2 values for GB3 have a root-mean-square-error (RMSE) of 0.33 s^{-1} , compared to the mean R_2 value of approximately 5 s^{-1} . Both the low R_2 values of the flexible 9-14 loop and the high R_2 values of the rigid central helix (residues 23-36) are correctly predicted. NOE is also well predicted, with an RMSE of only 0.04 (compared to the mean NOE that is between 0.7 and 0.8 and typical of rigid globular proteins). The RMSE for R_1 is 0.18 s^{-1} , due to small but systematic overestimation. For ubiquitin, the RMSEs are 0.55 s^{-1} for R_2 , 0.17 s^{-1} for R_1 , and 0.10 for NOE, with each property showing slight overestimation.

However, R_2 is significantly overestimated while R_1 is significantly underestimated in NVT – 2 simulations (see Fig. S6C for GB3 and Fig. S6D for ubiquitin). The over- and underestimation get much worse in NVT – 20 simulations (Fig. S6E&F). These are clear signs that τ_1 is too long. NOE overestimation is also observed in NVT – 2 simulations (Fig. S6C&D), and can largely be attributed to τ_1 dilation by the Langevin thermostat. In NVT – 20 simulations, NOE overestimation does not get worse but rather is moderated; for the more flexible residues

(e.g., the 9-14 loop in GB3) NOE actually becomes too low (Fig. S6E&F). The change in NOE behavior from NVT – 2 to NVT – 20 simulations is due to the fact that NOE has a non-monotonic dependence on τ_1 and on τ_2 ³¹. NOE increases when τ_1 is dilated by 2-fold as in NVT – 2 simulations, but starts to decrease upon further τ_1 dilation. The decrease is exacerbated by τ_2 dilation, and becomes excessive for flexible residues with higher A_2 , leading to the changed NOE behavior in NVT – 20 simulations.

These problems caused by the Langevin thermostat are eliminated when the time contraction of Eq [12] is applied to τ_1 and τ_2 . In Figs. 6 and S7 we compare cNVT – 2 calculations of R_1 , R_2 , and NOE for the eight proteins against the experimental data. These corrected NVT results have essentially the same high accuracy as the NVE results. Moreover, the correction works well for NVT simulations at the high damping constant of 20 ps⁻¹ (Fig. S6G&H). For the eight proteins, NVE, cNVT – 0.2, cNVT – 2, and cNVT – 20 all have relatively small and comparable RMSEs in R_1 , R_2 , and NOE (Table S2). All the calculations miss high R_2 values attributable to exchange between different chemical environments, including residues 4, 5, 7, and 11 in Trp-cage⁴⁵, residues 9, 18, 19, 21, 25, 31, 33, and 34 in ShK⁴⁶, and residues 14, 15, and 39 in BPTI⁴⁸. Also, at 300 K, Trp-cage undergoes partial unfolding⁴⁵, which probably explains why the experimental R_2 values are overall higher than the calculated ones (Fig. S7B). These misses are expected, since exchange and partial unfolding processes occur on timescales longer than our MD simulations.

Figure 6. Comparison of backbone NH NMR relaxation properties predicted by corrected NVT simulations against experimental data. (A) & (C) are for GB3 and (B) & (D) are for ubiquitin. (A) & (B) Longitudinal and transverse relaxation rates from experiment (black) and from corrected NVT (at $\zeta = 2$ ps⁻¹; color). The root-mean-square-error (RMSE) calculated among all the residues of each protein is shown in the legend. (C) & (D) Corresponding comparison for ¹H-¹⁵N heteronuclear Overhauser effect.

The spectral density of our bi-exponential fit (Eq [3]) has the same mathematical form as that in the extended model-free approach (given by Eq [4] of Clore et al.³¹), with our A_1 playing the role of the order parameter S^2 . In both formulations, the amplitude sum ($A_1 + A_2$) is not constrained to 1, with the deficit from 1 implying an ultrafast component. One distinction is that we treated τ_1 as residue-specific but Clore et al. treated τ_1 as a global parameter for overall rotation. Traditionally, order parameters have been calculated as an average second-order

Legendre polynomial (P_2) after removing overall rotation by structural alignment⁴². We found that A_1 matches with the average P_2 if structural alignment is done within blocks of a length that is shorter than τ_1 but longer than τ_2 , such that the average in P_2 captures conformational sampling due to sub-ns internal motions (Fig. S8A&B for GB3 and ubiquitin, with 1-ns blocks for structural alignment). Indeed, A_1 represents the residual order after the sub-ns decay (modeled by time constant τ_2). In Fig. S8C&D, we compare order parameters calculated using the P_2 method with 1-ns blocks from NVE and NVT simulations, against each other and also against experimental data^{47,49}. This comparison confirms once again that the Langevin thermostat does not affect motional amplitudes. Compared with the experimental data, the calculated order parameters show slight overestimation overall and larger overestimation for a few of the more flexible residues (e.g., residue 12 in GB3). It is possible that the experimental S^2 has contributions from internal motions on timescales longer than 1 ns. When we increased the length of the blocks for averaging, the calculated order parameters did get reduced, but the reduction could be excessive for flexible residues (not shown).

3.6. cNVT also Predicts Well Side-Chain NMR Relaxation Properties. To verify that our findings on the distortions of the Langevin thermostat and our correction scheme are not limited to backbone dynamics, we also calculated side-chain methyl CH and CC correlation functions from the NVE and NVT simulations. These calculated results can be directly tested by experimental data on methyl relaxation.

As illustration, in Fig. S2 we compare the CH correlation functions of three methyl-containing residues against the backbone NH correlation functions of the same residues, all from NVT – 2 simulations of ubiquitin. A clear distinction is that NH bonds undergo limited wobbling so that the sub-ns decay has a very small amplitude (small A_2 and large A_1 as noted above),

while CH wobbling spans nearly an entire hemisphere on top of the preceding CC bonds, leading to a large amplitude for the sub-ns decay and a small amplitude for the ns decay. Accordingly, the value ranges, 0 to 0.1 for A_1 and 0.6 to 0.8 for A_2 from bi-exponential fits for CH bonds are the reverse of those for NH bonds (Fig. S9A). This contrast makes side-chain methyls a good test ground for our understanding of the Langevin thermostat. Despite the reversal in magnitudes for the two amplitudes (A_1 and A_2), the Langevin thermostat still does not perturb either amplitude. Moreover, the effects of the Langevin thermostat on the two time constants (τ_1 and τ_2) of CH bonds are also similar to those observed for backbone NH bonds, i.e., significant dilation on τ_1 and slight change in τ_2 (Fig. S9B&C). The τ_1 values of CH bonds in both NVE and NVT – 2 simulations are still comparable to the respective RCTs, but have much greater site-to-site variations than found for backbone NH bonds (compare Figs. 3I and S9B). The greater variations likely reflect the fact that methyls are less tightly coupled to the protein matrix than backbone amides.

After applying the same correction scheme (Eq [12]) to τ_1 and τ_2 of CH bonds from NVT simulations, we calculated the CH in-phase transverse and longitudinal relaxation rates, denoted by $R(D_y)$ and $R(D_z)$, respectively. Figure 7A&B compares the cNVT – 2 results for 50 methyls in ubiquitin against experimental data²⁸⁻²⁹. Overall good agreement is seen, but with overestimation in $R(D_z)$ for a few residues (e.g., I3 C γ H).

As a final validation that the Langevin thermostat does not perturb motional amplitudes, we calculated methyl axis order parameters, S_{axis}^2 , from NVE and NVT simulations. Figure 7C displays S_{axis}^2 results for 50 methyls in ubiquitin from NVE, NVT – 0.2, NVT – 2, and NVT – 20 as well as from experiments²⁸⁻²⁹. Overall, S_{axis}^2 values agree well among the different simulations. However, there is a gap between NVE and NVT – 0.2 values and NVT – 2 and

NVT – 20 values for some of the most flexible methyls (indicated by low S_{axis}^2), including those on L8, L43, L71, and L73. The NVT – 2 and NVT – 20 values are in closer agreement with experiments, and thus the higher S_{axis}^2 from NVE and NVT – 0.2 simulations is a sign that they suffer from insufficient conformational sampling. The RMSEs of the NVE and NVT – 0.2 simulations from Lee et al.²⁸ are 0.15 and 0.16, respectively. These are already relatively low errors for methyl order parameters. The RMSEs further decrease to 0.11 and 0.12, respectively, for NVT – 2 and NVT – 20 simulations.

Figure 7. Comparison of side-chain methyl NMR relaxation properties predicted by corrected NVT simulations against experimental data for ubiquitin. (A) & (B) Side-chain methyl CH in-phase transverse and longitudinal relaxation rates, from corrected NVT (at $\zeta = 2 \text{ ps}^{-1}$) and from experimental studies of Lee et al. (28) and Liao et al. (29). (C) Methyl axis order parameters, S_{axis}^2 , predicted by the A_1 amplitudes of the CC correlation functions in NVE and NVT simulations (at $\zeta = 0.2, 2$, and 20 ps^{-1}), and measured by Lee et al. and by Liao et al.

4. DISCUSSION

We have thoroughly assessed the distortions of the dynamics of globular proteins by the Langevin thermostat, and developed a correction scheme to remove the distortions. The Langevin thermostat dilates the time constants of overall and internal motions, but does not affect the amplitudes of these motions. The extents of dilation are significant for the slower (ns) overall rotation but modest for the fast (sub-ns) internal motions. The correction scheme involves contraction of the time constants, with the contraction factor a linear function of the time constant to be corrected. The corrected dynamics of eight proteins are validated by NMR data for rotational diffusion and for backbone amide and side chain methyl relaxation.

Thermostat distortions observed in limited past studies can now be understood in light of our findings. For example, the sub-ps underdamped time correlation function of a side-chain torsion angle was preserved by a Langevin thermostat at $\zeta = 2 \text{ ps}^{-1}$, but became overdamped at $\zeta = 20$ and 50 ps^{-1} ¹². This observation is in line with slight τ_2 dilation at $\zeta = 2 \text{ ps}^{-1}$ but clear τ_2 dilation at $\zeta = 20 \text{ ps}^{-1}$ in our simulations. Basconi and Shirts¹⁷ observed several-fold slowdown in water diffusion and homopolymer chain decorrelation by a Langevin thermostat at $\zeta = 10 \text{ ps}^{-1}$. For this damping constant, our Eq [13] predicts a value of 2.8 for the a coefficient, meaning that

all time constants are increased by at least 2.8-fold by the Langevin thermostat. Debiec et al.¹⁸ reported RCTs for GB3, ubiquitin, and binase from NVE and NVT simulations at $\zeta = 1 \text{ ps}^{-1}$. For this damping constant, our Eqs [13] and [14] yield $a = 1.09$ and $b = 0.059 \text{ ns}^{-1}$. Applying the resulting time-contraction correction to their NVT results, we predict RCTs of 3.58, 4.49, and 6.34 ns for the three proteins in NVE simulations. These predictions agree well with the reported values, 3.47, 4.62, and 7.26 ns. Lastly Hoffmann et al.²³ reported that NPT and NVE simulations produced “very similar” results for methyl axis order parameters and CH in-phase transverse and longitudinal relaxation rates. Their observation on S_{axis}^2 supports our finding that motional amplitudes are not perturbed by thermostats. Also, $R(D_y)$ and $R(D_z)$ are dominated by fast (sub-ns) local motions (as $A_2 \gg A_1$ for methyl CH bonds) and hence should not be perturbed much either.

We anticipate that our correction scheme for thermostat distortions will have wide applications. Although isotropic overall rotation is modeled in the present study, the correction scheme can be easily extended to model anisotropic rotation of folded proteins or complexes that deviate significantly from a globular shape. Another tantalizing application is to remove dynamic distortions of intrinsically disordered proteins (IDPs) in NVT simulations. As illustration, we apply the correction of Eq [12] to our previous NVT simulations of ChiZ at $\zeta = 3 \text{ ps}^{-1}$ ³⁴. The correction indeed brings the calculated backbone NH relaxation properties into closer agreement with experimental data (Fig. S10). Most significantly, since the extent of time dilation by a Langevin thermostat grows with increasing timescales of motions, we expect that time dilation for functional important conformational switches⁵² and protein folding/unfolding that occur on μs or longer timescales will exceed the 2-fold observed here for 5-ns motions. As NVE simulations suffer from temperature drifts¹³ (Fig. S11) as well as insufficient conformational

sampling, it will be impractical to use them for studying IDP dynamics and other slow processes. A thermostat and the necessary dynamic correction provide the only way to propel MD simulations to be on par with NMR and other experimental techniques in determining dynamic properties of proteins.

That the Langevin thermostat does not affect motional amplitudes can be rationalized if we recognize that motional amplitudes are thermodynamic, not dynamic properties, and the Langevin thermostat, by design, preserves thermodynamic properties. That the Langevin thermostat dilates time constants of overall rotation to a much greater extent than those of internal motions can also be rationalized as follows. In overall rotation, a protein experiences only the solvent friction in NVE simulations. A Langevin thermostat not only increases the solvent friction (due to Langevin dynamics of the solvent) but also directly adds friction to each protein atom (due to Langevin dynamics of the protein). The total increase in friction may be comparable in magnitude to the original solvent friction. In contrast, for internal motions, the dominant contribution is internal friction arising from a rugged energy surface, and hence the increase in friction by the Langevin thermostat is less important. It is less straightforward to explain why time dilation for overall rotation grows with RCT and hence protein size (Fig. S1). The RCT from an NVE simulation is proportional to the original solvent viscosity η_0 and the protein volume V : $RCT_{NVE} = \beta\eta_0V$, where β is a constant. Correspondingly we have $RCT_{NVT} = \beta(\eta_0 + \eta_1)V$, where η_1 is the total increase in viscosity by the Langevin thermostat. If η_1 depends on protein size, then the time dilation, RCT_{NVT}/RCT_{NVE} , will also depend on protein size. These foregoing qualitative analyses are also instructive for other types of thermostats.

Just as water models affect the RCT of a protein^{13, 20}, we should expect that the dynamic distortions of a Langevin thermostat depend on the water model. As already noted, the dynamic

distortions also depend the timescales of motions and on protein size. Without corrections, variations in distortions due to these inter-related factors can greatly complicate force-field parameterization and validation and affect the transferability of force fields. Our correction scheme will be valuable for both force-field parameterization and transferability.

ASSOCIATED CONTENT

Supporting Information. The Supporting Information is available free of charge.

Two tables listing rotational correlation times and root-mean-square-errors and 11 figures displaying time dilation by a Langevin thermostat; backbone NH and side-chain methyl CH time correlation functions; parameters of bi-exponential fits for NH time correlation functions; dependences of τ_1 and τ_2 on the damping constant of the Langevin thermostat; comparison of backbone NH NMR relaxation properties predicted by NVE and NVT simulations against experimental data; backbone NH order parameters; parameters of bi-exponential fits for ubiquitin methyl CH time correlation functions; comparison of backbone NH NMR relaxation properties calculated from NVT simulations of ChiZ against experimental data; and temperature drift in NVE simulations and temperature regulation by a Langevin thermostat (PDF)

AUTHOR INFORMATION

Corresponding Author

Huan-Xiang Zhou – Department of Chemistry and Department of Physics, University of Illinois at Chicago, Chicago, IL 60607, USA

orcid.org/0000-0001-9020-0302; E-mail: hzhou43@uic.edu

Author Contributions

A.H. and H.-X.Z. designed the research. A.H. and M.M. performed the research and analyzed the data. A.H. and H.-X.Z. wrote the manuscript.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

This work was supported by National Institutes of Health Grant R35 GM118091.

ABBREVIATIONS

BPTI, bovine pancreatic trypsin inhibitor; GB3, the third immunoglobulin-binding domain of protein G; HEWL, hen egg white lysozyme; MD, molecular dynamics; NVE, particle number / volume / energy; NVT, particle number / volume / temperature; PME, particle mesh Ewald; RCT, rotational correlation time; RMSE, root-mean-square-error; RNase, ribonuclease; SASA, solvent accessible surface area.

REFERENCES

1. Andersen, H. C. Molecular-Dynamics Simulations at Constant Pressure and-or Temperature. *J Chem Phys* **1980**, 72, 2384-2393.
2. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J Chem Phys* **1984**, 81, 3684-3690.
3. Nose, S. A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods. *J Chem Phys* **1984**, 81, 511-519.
4. Hoover, W. G. Canonical Dynamics - Equilibrium Phase-Space Distributions. *Phys Rev A* **1985**, 31, 1695-1697.

5. Pastor, R. W.; Brooks, B. R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol Phys* **1988**, *65*, 1409-1419.
6. Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nose-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics. *J Chem Phys* **1992**, *97*, 2635-2643.
7. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **2007**, *126*, 014101.
8. Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations. *Living J Comput Mol Sci* **2019**, *1*, 5957.
9. Wong-Ekkabut, J.; Karttunen, M. The good, the bad and the user in soft matter simulations. *Biochim Biophys Acta* **2016**, *1858*, 2529-2538.
10. Harvey, S. C.; Tan, R. K. Z.; Cheatham, T. E. The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J Comput Chem* **1998**, *19*, 726-740.
11. Rosta, E.; Buchete, N. V.; Hummer, G. Thermostat artifacts in replica exchange molecular dynamics simulations. *J Chem Theory Comput* **2009**, *5*, 1393-1399.
12. Tobias, D. J.; Martyna, G. J.; Klein, M. L. Molecular dynamics simulations of a protein in the canonical ensemble. *J Phys Chem* **1993**, *97*, 12959-12966.
13. Wong, V.; Case, D. A. Evaluating rotational diffusion from protein MD simulations. *J Phys Chem B* **2008**, *112*, 6013-24.
14. Leimkuhler, B.; Noorizadeh, E.; Penrose, O. Comparing the Efficiencies of Stochastic Isothermal Molecular Dynamics Methods. *J Stat Phys* **2011**, *143*, 921-942.

15. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylethylamide. *Biopolymers* **1992**, *32*, 523-35.
16. Paterlini, M. G.; Ferguson, D. M. Constant temperature simulations using the Langevin equation with velocity Verlet integration. *Chem Phys* **1998**, *236*, 243-252.
17. Basconi, J. E.; Shirts, M. R. Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations. *J Chem Theory Comput* **2013**, *9*, 2887-99.
18. Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J Chem Theory Comput* **2016**, *12*, 3926-47.
19. Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940**, *7*, 284-304.
20. Takemura, K.; Kitao, A. Water model tuning for improved reproduction of rotational diffusion and NMR spectral density. *J Phys Chem B* **2012**, *116*, 6279-87.
21. Showalter, S. A.; Bruschweiler, R. Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB force field. *J Chem Theory Comput* **2007**, *3*, 961-975.
22. Chen, P. C.; Hologne, M.; Walker, O.; Hennig, J. Ab Initio Prediction of NMR Spin Relaxation Parameters from Molecular Dynamics Simulations. *J Chem Theory Comput* **2018**, *14*, 1009-1019.
23. Hoffmann, F.; Mulder, F. A. A.; Schafer, L. V. Accurate Methyl Group Dynamics in Protein Simulations with AMBER Force Fields. *J Phys Chem B* **2018**, *122*, 5038-5048.

24. Marcellini, M.; Nguyen, M. H.; Martin, M.; Hologne, M.; Walker, O. Accurate Prediction of Protein NMR Spin Relaxation by Means of Polarizable Force Fields. Application to Strongly Anisotropic Rotational Diffusion. *J Phys Chem B* **2020**, *124*, 5103-5112.
25. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712-25.
26. Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* **2009**, *113*, 9004-15.
27. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950-8.
28. Lee, A. L.; Flynn, P. F.; Wand, A. J. Comparison of H-2 and C-13 NMR relaxation techniques for the study of protein methyl group dynamics in solution. *J Am Chem Soc* **1999**, *121*, 2891-2902.
29. Liao, X.; Long, D.; Li, D. W.; Bruschweiler, R.; Tugarinov, V. Probing side-chain dynamics in proteins by the measurement of nine deuterium relaxation rates per methyl group. *J Phys Chem B* **2012**, *116*, 606-20.
30. Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* **1982**, *104*, 4546-4559.
31. Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of N-15 Nuclear Magnetic Relaxation of Proteins. *J Am Chem Soc* **1990**, *112*, 4989-4991.

32. Salvi, N.; Abyzov, A.; Blackledge, M. Analytical Description of NMR Relaxation Highlights Correlated Dynamics in Intrinsically Disordered Proteins. *Angew Chem Int Ed Engl* **2017**, *56*, 14020-14024.
33. Kampf, K.; Izmailov, S. A.; Rabdano, S. O.; Groves, A. T.; Podkorytov, I. S.; Skrynnikov, N. R. What Drives (15)N Spin Relaxation in Disordered Proteins? Combined NMR/MD Study of the H4 Histone Tail. *Biophys J* **2018**, *115*, 2348-2367.
34. Hicks, A.; Escobar, C. A.; Cross, T. A.; Zhou, H. X. Sequence-Dependent Correlated Segments in the Intrinsically Disordered Region of ChiZ. *Biomolecules* **2020**, *10*, 1-23.
35. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham T.E., I. I. I.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K., et al. AMBER 2018. **2018**.
36. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **2015**, *11*, 3696-3713.
37. Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J Phys Chem B* **2015**, *119*, 5113-5123.
38. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **2013**, *9*, 3878-3888.
39. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* **1977**, *23*, 327-341.

40. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J Chem Phys* **1995**, *103*, 8577-8593.
41. Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013**, *9*, 3084-3095.
42. Chatfield, D. C.; Szabo, A.; Brooks, B. R. Molecular dynamics of staphylococcal nuclease: Comparison of simulation with ¹⁵N and ¹³C NMR relaxation data. *J Am Chem Soc* **1998**, *120*, 5301-5311.
43. Mittermaier, A.; Kay, L. E. Measurement of methyl ²H quadrupolar couplings in oriented proteins. How uniform is the quadrupolar coupling constant? *J Am Chem Soc* **1999**, *121*, 10608-10613.
44. Conibear, A. C.; Wang, C. K.; Bi, T.; Rosengren, K. J.; Camarero, J. A.; Craik, D. J. Insights into the Molecular Flexibility of θ -Defensins by NMR Relaxation Analysis. *J Phys Chem B* **2014**, *118*, 14257-14266.
45. Rovõ, P.; Stráner, P.; Láng, A.; Bartha, I.; Huszár, K.; Nyitray, L.; Perczel, A. Structural insights into the Trp-cage folding intermediate formation. *Chem Eur J* **2013**, *19*, 2628-2640.
46. Sher, I.; Chang, S. C.; Li, Y.; Chhabra, S.; Palmer, A. G.; Norton, R. S.; Chill, J. H. Conformational flexibility in the binding surface of the potassium channel blocker ShK. *ChemBioChem* **2014**, *15*, 2402-2410.
47. Hall, J. B.; Fushman, D. Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *J Biomol NMR* **2003**, *27*, 261-75.
48. Beeser, S. A.; Goldenberg, D. P.; Oas, T. G. Enhanced protein flexibility caused by a destabilizing amino acid replacement in BPTI. *J Mol Biol* **1997**, *269*, 154-164.

49. Tjandra, N.; Bax, A.; Feller, S. E.; Pastor, R. W. Rotational Diffusion Anisotropy of Human Ubiquitin from ^{15}N NMR Relaxation. *J Am Chem Soc* **1995**, *117*, 12562-12566.
50. Fushman, D.; Weisemann, R.; Thüring, H.; Rüterjans, H. Backbone dynamics of ribonuclease T1 and its complex with 2'GMP studied by two-dimensional heteronuclear NMR spectroscopy. *J Biomol NMR* **1994**, *4*, 61-78.
51. Buck, M.; Boyd, J.; Redfield, C.; Dobson, C. M.; MacKenzie, D. A.; Jeenes, D. J.; Archer, D. B. Structural Determinants of Protein Dynamics: Analysis of ^{15}N NMR Relaxation Measurements for Main-Chain and Side-Chain Nuclei of Hen Egg White Lysozyme. *Biochemistry* **1995**, *34*, 4041-4055.
52. Hempel, T.; Plattner, N.; Noe, F. Coupling of Conformational Switches in Calcium Sensor Unraveled with Local Markov Models and Transfer Entropy. *J Chem Theory Comput* **2020**, *16*, 2584-2593.