JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

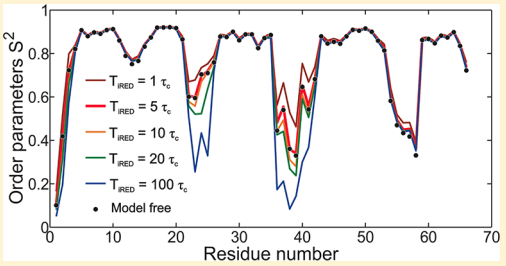
NMR Order Parameter Determination from Long Molecular Dynamics Trajectories for Objective Comparison with Experiment

Yina Gu,[†] Da-Wei Li,[‡] and Rafael Brüschweiler*[,†,‡]

[†]Department of Chemistry and Biochemistry and [‡]Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, United States

Supporting Information

ABSTRACT: Functional protein motions covering a wide range of time scales can be studied, among other techniques, by NMR and by molecular dynamics (MD) computer simulations. MD simulations of proteins now routinely extend into the hundreds of nanoseconds time scale range exceeding the overall tumbling correlation times of proteins in solution by several orders of magnitude. This provides a unique opportunity to rigorously validate these simulations by quantitative comparison with model-free order parameters derived from NMR relaxation experiments. However, presently there is no consensus on how such a comparison is best done. We address here how this can be accomplished in a way that is both efficient and objective. For this purpose, we analyze ^{15}N R_1 and R_2 and heteronuclear $\{^1H\}-^{15}N$ NOE NMR relaxation parameters computed from 500 ns MD trajectories of 10 different protein systems using the model-free analysis. The resulting model-free S^2 order parameters are then used as targets for S^2 values computed directly from the trajectories by the iRED method by either averaging over blocks of variable lengths or by using exponentially weighted snapshots (wiRED). We find that the iRED results are capable of reproducing the target S^2 values with high accuracy provided that the averaging window is chosen 5 times the length of the overall tumbling correlation time. These results provide useful guidelines for the derivation of NMR order parameters from MD for a meaningful comparison with their experimental counterparts.



1. INTRODUCTION

Molecular dynamics (MD) is a powerful tool to study biomolecular motions at atomic detail on time scales from picoseconds to nanoseconds.[1] For the purpose of an accurate and realistic *in silico* description of biophysically and biologically relevant molecular events, important strides have been made through the development of improved protein force fields[2−4] and by extending trajectory lengths into the microsecond range and beyond.[5,6] Proper benchmarking of these simulations against quantitative experimental data reporting on protein structure and dynamics at atomic resolution that have been measured under the same conditions is critical for the further improvement of the MD method as well as for the enhancement of the interpretation of experimental data.[7−9]

NMR spin relaxation parameters of proteins in solution are uniquely suited for this task as they allow site-specific observation of both the magnitudes and the time scales of internal protein dynamics.[10−14] NMR backbone ^{15}N longitudinal R_1 rates, transverse R_2 relaxation rates, and $\{^1H\}-^{15}N$ heteronuclear Overhauser effects (NOEs) are routinely translated into generalized Lipari−Szabo order parameters S^2 by the model-free approach and its extension.[15−17] Mathematically, S^2 is defined as the plateau value of the internal reorientational correlation function of the internuclear vector, e.g., $^{15}N-^1H$ bond vector, of the magnetic dipole−dipole interaction of interest. The magnitude of S^2 can vary between 0 and 1 whereby lower S^2 values generally correspond to larger amplitude internal reorientational motions, while high values indicate a higher degree of motional restriction of the bond vector relative to the molecular frame. These parameters sensitively reflect reorientational dynamics on time scales of pico- to nanoseconds, which are of the same order of magnitude as the global tumbling correlation time, τ_c, or faster.[15]

S^2 values can also be computed from MD trajectories.[8,9,16,18−20] Since order parameters report about internal motions, internal motions need to be separated from overall tumbling.[21] This can be done explicitly by orientationally aligning each snapshot relative to a reference structure or implicitly as is done in the isotropic reorientational eigenmode dynamics (iRED) analysis method.[22] When the iRED method was introduced over a decade ago, trajectory lengths were naturally limited to the low nanosecond range, i.e., lengths that are comparable to or shorter than τ_c, and hence it was natural to average the iRED matrix over the longest available trajectories at that time. With the advent of trajectories into the hundreds of nanoseconds range and microseconds range, a choice for the window size (block size) of the MD trajectory needs to be made over which S^2 values are computed and averaged. Selected iRED window lengths reported in the literature vary widely ranging from 0.5 ns to 2 μs,[23−29] indicating that there is presently no consensus on how such comparisons are best done. While S^2

Received: February 28, 2014
Published: April 25, 2014

values in well-defined secondary structures are typically high and will converge on the low nanoseconds time scale, the choice of the window size is critical for flexible loop regions and termini as they can display substantially slower time scale motions.[4,25] An iRED window corresponding to a length of approximately τ_c is often chosen, because (i) spin relaxation parameters are most sensitive to motions with correlation times τ that fulfill $\omega_0\tau \approx 1$ where ω_0 is the ^{15}N Larmor frequency and (ii) overall tumbling motion forces the total correlation function toward zero for times much longer than τ_c, suppressing the effect of internal motions with longer correlation times on spin relaxation.

These arguments are, however, qualitative in nature and have not been tested quantitatively. In particular, it is unknown whether a universal relationship between the optimal iRED window size and τ_c exists, which permits the accurate reproduction of model-free order parameters S^2 fitted from R_1, R_2, and NOE values computed from trajectories. This question is addressed here following the strategy depicted in the flowchart of Figure 1. S^2 order parameters are determined in three different

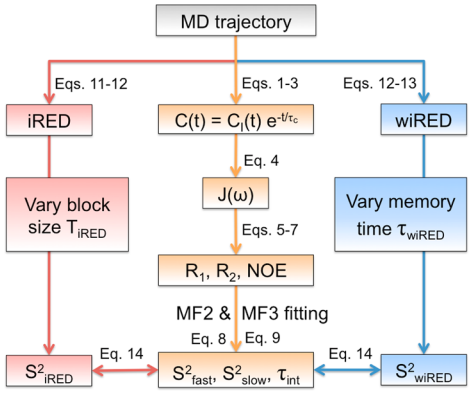


Figure 1. Flowchart for the optimization of S^2 order parameters extracted from long MD trajectories using iRED (left path) and wiRED (right path) for optimal agreement with S^2 values extracted by the model-free approach (MF2 or MF3) from back-calculated R_1, R_2, and NOE relaxation data.

ways and compared with each other. The first method (in orange) mimics the experimental approach. It is based on the model-free fitting of S^2 from R_1, R_2, and NOE parameters directly computed from the MD trajectory. For this purpose, the total reorientational correlation function $C(t)$ is computed as the product of the overall tumbling correlation function, assumed to be monoexponential consistent with isotropic overall tumbling with a predefined correlation time τ_c, and the internal correlation function computed explicitly for each N−H vector from the trajectory. R_1, R_2, and NOE parameters are obtained after converting $C(t)$ to the spectral density function by Fourier transformation. Second, the S^2 values are determined directly using the iRED method (in red) whereby the window size is systematically varied to optimally reproduce the S^2 values obtained by the first approach. Third, a variant of iRED is introduced here, referred to as wiRED, in which the iRED matrix is computed by altering the relative weights of snapshots in an exponentially decaying fashion according to their time separation (in blue). We also investigate the effect of the selection of the number of internuclear vectors on the iRED order parameters and propose a vector set that produces accurate and stable S^2 values for proteins of both small and large size.

It is not the purpose of this study to test how well MD-derived S^2 values agree with their experimental counterparts for different proteins and force fields. This has been the subject of several studies in the past.[2,25,30,31] Our objective is to determine how S^2 values should be computed so that they fully agree with the experimentally derived S^2 values in case that the protein dynamics displayed during the MD trajectory is identical with the dynamics displayed by the proteins in the NMR tube.

2. MATERIALS AND METHODS

2.1. Molecular Dynamics Simulations. Ten globular proteins of variable length and topology have been used in this work with PDB codes 1MJC, 1EKG, 1GNU, 1GST, 1QZM, 1VYK, 2B02, 2END, 1IGD, and 1UBI (for additional information see the Supporting Information, Table S1). All MD simulations were performed using the Gromacs 4 package[32−35] with the ff99SB_$\varphi\psi$(g24;CS) force field.[36] Proteins were solvated in a cubic box with explicit TIP3P water model[37] extending at least 8 Å from the protein to the edges. The internal water geometry was fixed using the SETTLE algorithm.[38] The protonation state of the protein was adjusted to neutral pH. The cutoff distance of van der Waals interaction was set to 10 Å, and the long-range electrostatic interactions were computed using the PME algorithm with 1.2 Å spacing at a cutoff of 8 Å. The system was run with a 2 fs time step and using the LINCS algorithm[39] to constrain all bond lengths involving hydrogen atoms. After standard energy minimization and equilibration, a 500 ns production MD was performed for each protein under constant 300 K temperature and 1 atm pressure (NPT) condition. Snapshots were sampled every 100 ps, resulting in 5000 snapshots per trajectory for subsequent analysis.

2.2. Back-Calculation of Spin Relaxation Parameters from Time-Correlation Functions. For globular proteins that tumble in good approximation isotropically and independently of the internal dynamics, the normalized time-autocorrelation function $C(t)$ of the lattice part of the spin-relaxation active magnetic dipole−dipole interaction can be factored into an overall tumbling part $C_O(t) = e^{-t/\tau_c}$ and an internal part $C_I(t)$:[15,16]

$$C(t) = C_O(t)\, C_I(t) \tag{1}$$

where τ_c is the isotropic tumbling correlation time of rank 2, which is related to the rotational diffusion D_r constant by $\tau_c = 1/(6D_r)$. The internal correlation function $C_I(t)$ describes internal reorientation, and it can be computed directly from an MD trajectory using the second-order Legendre polynomial:

$$C_I(t) = \frac{1}{2}\langle 3[\mathbf{e}(\tau)\,\mathbf{e}(\tau + t)]^2 - 1\rangle \tag{2}$$

where $\mathbf{e}(t)$ is the unit vector defining the ^{15}N−^{1}H bond orientation after overall tumbling has been eliminated by aligning each snapshot with respect to a reference snapshot (snapshot at time 0). The angular brackets indicate averaging from time $\tau = 0$ to $T_{MD} - t$, where T_{MD} is the total trajectory length. In practice, $C_I(t)$ can be efficiently computed by the fast Fourier transform (FFT) method. We limit the analysis of $C_I(t)$ to its initial part from $t = 0$ to $0.3T_{MD}$, which is statistically better converged than at longer times. Next, we fit a multiexponential decaying function to $C_I(t)$ with an offset A_0:[40]

$$C_I(t) = A_0 + \sum_{i=1}^{5} A_i e^{-t/\tau_i} \tag{3}$$

dx.doi.org/10.1021/ct500181v | J. Chem. Theory Comput. 2014, 10, 2599−2607

where A_i and τ_i are the best fitting parameters subject to the conditions:

$$\sum_{i=0}^{5} A_i = 1 \qquad A_i \geq 0,\ \tau_i \geq 0$$

This permits analytical Fourier transformation of $C(t)$, after multiplying $C_I(t)$ with the overall tumbling term e^{-t/τ_c} for a chosen τ_c, to obtain the spectral density function $J(\omega)$:

$$J(\omega) = 2\int_0^\infty C(t)\cos(\omega t)\,dt = \frac{A_0 2\tau_c}{1+(\omega\tau_c)^2} + \sum_{i=1}^{5}\frac{A_i 2\tau_{ieff}}{1+(\omega\tau_{ieff})^2} \quad (4)$$

where $\tau_{ieff} = \tau_c\tau_i/(\tau_c+\tau_i)$. NMR spin relaxation parameters R_1, R_2 and the heteronuclear $\{^1H\}-^{15}N$ NOE are then computed using the standard expressions:[41−44]

$$R_1 = 1/T_1 = d_{oo}[3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)] + c_{oo}\omega_N{}^2 J(\omega_N) \quad (5)$$

$$R_2 = 1/T_2 = \frac{1}{2}d_{oo}[4J(0) + 3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_N + \omega_H)] + \frac{1}{6}c_{oo}\omega_N{}^2[4J(0) + 3J(\omega_N)] \quad (6)$$

$$\mathrm{NOE} = 1 + \frac{\gamma_H}{\gamma_N}d_{oo}T_1[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad (7)$$

where $d_{oo} = (1/20)(\mu_0/4\pi)^2(h/2\pi)^2\gamma_H{}^2\gamma_N{}^2\langle r_{NH}{}^{-3}\rangle^2$; $c_{oo} = (1/15)\Delta\sigma^2$. μ_0 is the permeability of vacuum, h is Planck's constant, γ_H and γ_N are the gyromagnetic ratios of ^1H and ^{15}N nuclei, and the backbone N−H bond length is r_{NH} = 1.02 Å. The ^{15}N chemical shielding anisotropy (CSA) has been set to either $\Delta\sigma$ = 0 or −170 ppm, although the precise CSA value has no noticeable effect on the results of this study.

2.3. Fitting of Model-Free Order Parameters from R_1, R_2, NOE. The model-free approach and its extended version were applied to the computed R_1, R_2, and NOE parameters using either two or three fitting parameters with τ_c being kept fixed at its original value.[45,46] The model-free model with two fitting parameters (MF2) contains the order parameter S^2 and internal motional correlation time τ_{int}:

$$C_I(t) = S^2 + (1-S^2)e^{-t/\tau_{int}}$$

$$J(\omega) = S^2\frac{2\tau_c}{1+(\omega\tau_c)^2} + (1-S^2)\frac{2\tau_{eff}}{1+(\omega\tau_{eff})^2} \quad (8)$$

The extended model-free analysis (MF3)[17] describes the internal motions occurring on both fast and slower time scales by the three parameters S^2_{fast}, S^2_{slow}, and τ_{int}:

$$C_I^{MF3}(t) = S^2_{fast}C_I(t) = S^2_{fast}S^2_{slow} + S^2_{fast}(1-S^2_{slow})e^{-t/\tau_{int}}$$

$$J^{MF3}(\omega) = S^2_{fast}S^2_{slow}\frac{2\tau_c}{1+(\omega\tau_c)^2} + S^2_{fast}(1-S^2_{slow})\frac{2\tau_{eff}}{1+(\omega\tau_{eff})^2} \quad (9)$$

where $\tau_{eff} = \tau_c\tau_{int}/(\tau_c+\tau_{int})$. For MF3, τ_{int} is the effective correlation time that belongs to S^2_{slow}. The generalized order parameter S^2_{total} is the product of fast and slow time scale order parameters; $S^2_{total} = S^2_{fast}S^2_{slow}$. The model-free parameters are obtained by a nonlinear-least-squares fit.

In the original work by Lipari and Szabo, the order parameter S^2 was defined as the plateau value of $C_I(t)$, which can be estimated from the normalized spherical harmonics of rank 2, $Y_{2m}(\theta(t),\varphi(t))$, averaged over the full trajectory:

$$C_I(\infty) = S^2 = \frac{4\pi}{5}\sum_{m=-2}^{2}\langle Y_{2m}\rangle\langle Y_{2m}^*\rangle = \frac{3}{2}(\langle x^2\rangle^2 + \langle y^2\rangle^2 + \langle z^2\rangle^2 + 2\langle xy\rangle^2 + 2\langle xz\rangle^2 + 2\langle yz\rangle^2) - \frac{1}{2} \quad (10)$$

The bottom part of eq 10 is an alternative, computationally more convenient form, which is equivalent to the one in ref 47, where the normalized internuclear vectors are represented in Cartesian coordinates: $\mathbf{r} = (x(t),y(t),z(t))$.

It should be stressed that S^2 of eq 10 depends on the trajectory length and is generally not equivalent to S^2 fitted from R_1, R_2, and NOE (eqs 8 and 9). The latter is insensitive to internal motions on correlation times much longer than τ_c because of the suppression of long time scale internal motional effects by the product form of eq 1, whereas the former is affected by all motions that occur during the MD trajectory independent of their time scales.

2.4. Block-Averaged iRED and wiRED Analyses. The iRED method is an alternative approach for the computation of order parameters from MD ensembles, which is based on a principal component analysis of the isotropically averaged covariance matrix $\mathbf{M}^{iRED}$ of the spatial functions of the dipolar interaction:[22,48]

$$M_{ij}^{iRED} = \frac{1}{2}\langle 3(\mathbf{u}_{LF,i}\mathbf{u}_{LF,j})^2 - 1\rangle_{T_{iRED}} \quad (11)$$

where $\mathbf{u}_{LF,i}$ represents the N normalized bond vectors of interest in an arbitrary global frame, e.g., the laboratory frame. In most iRED studies concerned with the analysis of N−H order parameters, eq 11 only includes backbone N−H vectors. Here, we include a larger number of bond vectors, namely N−H, N−Cα, Cα−Hα, Cα−C′, and Cα−Cβ bonds for all amino acids, except for glycine residues where only the first four vectors were used. The angular brackets indicate averaging over the snapshots of a trajectory block (window) of length T_{iRED}. The iRED matrix (eq 11) is determined for each block, and order parameters are determined for each matrix by an eigenvalue decomposition:

$$S_k^2 = 1 - \sum_{m=6}^{N}\lambda_m |\langle m|_k|^2 \quad (12)$$

where λ_m is the mth eigenvalue belonging to eigenvector $|m\rangle$ ordered according to decreasing size and k denotes the interaction vector whose order parameter is of interest. The

dx.doi.org/10.1021/ct500181v | J. Chem. Theory Comput. 2014, 10, 2599−2607

final order parameters are obtained by averaging the S^2_k values determined by eq 12 over all (nonoverlapping) blocks of length T_{iRED}.

We also tested a variant of iRED, termed weighted iRED or wiRED, where the iRED matrix is constructed by differential weighting of sequential snapshots:

$$M_{ij}^{t_0} = \sum_{t_k=t_0}^{t_0+5\tau_{wiRED}} W_{t_k} \frac{3(\mathbf{u}_{LF,i}(t_k)\ \mathbf{u}_{LF,j}(t_k))^2 - 1}{2} \quad (13)$$

with normalized weights $W_{t_k} = e^{-t_k/\tau_{wiRED}} / \sum_{t_k=0}^{5\tau_{wiRED}} e^{-t_k/\tau_{wiRED}}$, where τ_{wiRED} represents a memory time. A wiRED matrix is computed for each t_0, which is shifted along the trajectory in steps of τ_{wiRED}. S^2_{wiRED} is calculated by averaging over the S^2_k values determined by eq 12 for each t_0.

2.5. Comparison of S^2 from iRED/wiRED Methods and Model-Free Analysis. Monte Carlo (MC) error analysis is performed by fitting MF2 or MF3 parameters after adding 5% random errors to the original R_1, R_2, and NOE data. The difference between S^2 values from MF2 or MF3 fits and iRED/wiRED is expressed by the χ^2 value:

$$\chi^2 = \sum_i \frac{(S^{2(i)}_{MF} - S^{2(i)}_{iRED/wiRED})^2}{\sigma^{2(i)}} \quad (14)$$

where i stands for the residue number and σ is the standard deviation of S^2 (from MF2 or MF3) determined over 30 MC simulations. A series of different iRED block windows (T_{iRED} = 5, 10, 25, 50, 100, 125, 250, and 500 ns) and wiRED memory times (τ_{wiRED} = 1, 2, 5, 10, 20, 30, 40, 60, 80, and 100 ns) are tested to identify those parameters that give a minimal χ^2 value compared to the model-free approach.

3. RESULTS

3.1. Order Parameters by iRED Using Five Vectors. The dependence of S^2 values determined by iRED on the choice of internuclear vectors included in eqs 11 and 12 was tested first. A comparison of order parameters determined by either using N−H vectors only or using five local bond vectors per residue (except for glycine) is given in Figure 2. Differences in S^2 are mostly observed in flexible regions (S^2 < 0.8) that sample picosecond to nanosecond time scale dynamics. In such regions, S^2 values computed from the covariance matrix of only the N−H vectors are slightly, but consistently, higher than the S^2 values obtained when using all five types of internuclear vectors. The correlation between Lipari−Szabo plateau values S^2 and iRED S^2 is slightly but statistically significantly better when using five vectors (r = 0.9997) compared with using N−H vectors only (r = 0.9975). Also, a better separability between internal and global motions is present when using five vectors as is reflected by the substantially larger amplitude gap between the largest five and the smallest (N-5) eigenvalues of the iRED covariance matrix.[22] Due to the limited number of local interaction vectors in small proteins, the ability to sufficiently sample an approximately isotropic distribution of vectors is diminished when using N−H vectors solely. As a consequence, the largest five eigenmodes that describe the average geometry can contain small but observable amounts of internal dynamics effects and, hence, this yields increased order parameters for mobile regions as seen in Figure 2. The use of five vectors guarantees a stable separation between global and internal motional effects, which is especially important for small proteins. For larger proteins, the effect is generally smaller, but the use of multiple vectors per amino acid is still recommended for the accurate calculation of order parameters by the iRED method, although for proteins with many hundreds of amino acids the increase in computational cost can become noticeable.

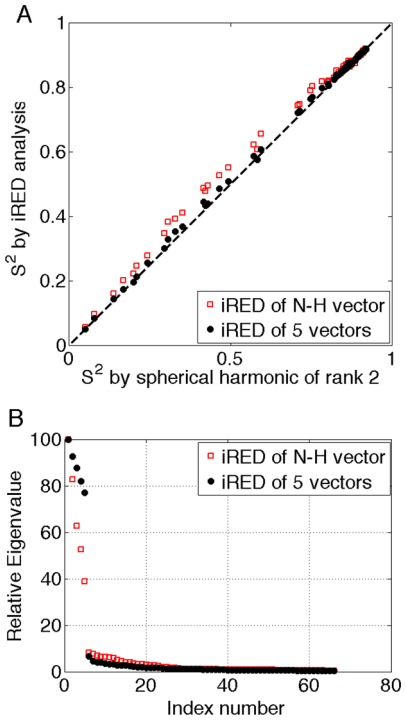


Figure 2. Dependence of iRED analysis on bond vector selection for protein Csp A (PDB 1MJC). (A) N−H S^2_{iRED} values calculated using N−H vectors only (red squares) and from iRED analysis using five sets of vectors (black dots). (B) Comparison of relative eigenvalues of iRED matrix constructed by one vector per amino acid (red squares) and by five vectors per (non-Gly) amino acid (black dots).

3.2. Comparison of S^2 Extracted by Different Methods. Four different methods for the determination of order parameters are compared for protein Csp A (PDB code 1MJC[49]). This protein serves as a good test case as its trajectory is converged displaying variable amounts of dynamics on multiple time scales for the loops and the termini. Figure 3 shows a comparison of S^2 values determined as the plateau values of the autocorrelation functions, by iRED and wiRED, and by the extended model-free analysis (MF3). For the iRED analyses (using five bond vectors per (non-Gly) amino acid), eight different block lengths of 5, 10, 25, 50, 100, 125, 250, and 500 ns were chosen (Figure 3A). The iRED order parameters that vary most as a function of window length belong to the two most flexible loops of Csp A (residues 22−26 and 36−42) undergoing nanosecond time scale dynamics.

Because the model-free fitting results are significantly better for MF3 than MF2, we focus our discussion on the MF3 results. The MF3 order parameters for a global tumbling correlation time τ_c of 5 ns (black dots) correlate extremely well with the iRED S^2 when an averaging block size of 25 ns is used with a correlation coefficient R of 0.9991 and χ^2 of 0.095. When iRED is used with a shorter block size (5 and 10 ns), the order parameters of flexible regions are overestimated, whereas a longer block size (>25 ns) yields order parameters that are too low when compared with MF3-derived S^2 values. The iRED results with longer window size merely reflect internal reorientational motions present at slower time scales. If these time scales substantially exceed τ_c,

dx.doi.org/10.1021/ct500181v | J. Chem. Theory Comput. 2014, 10, 2599−2607

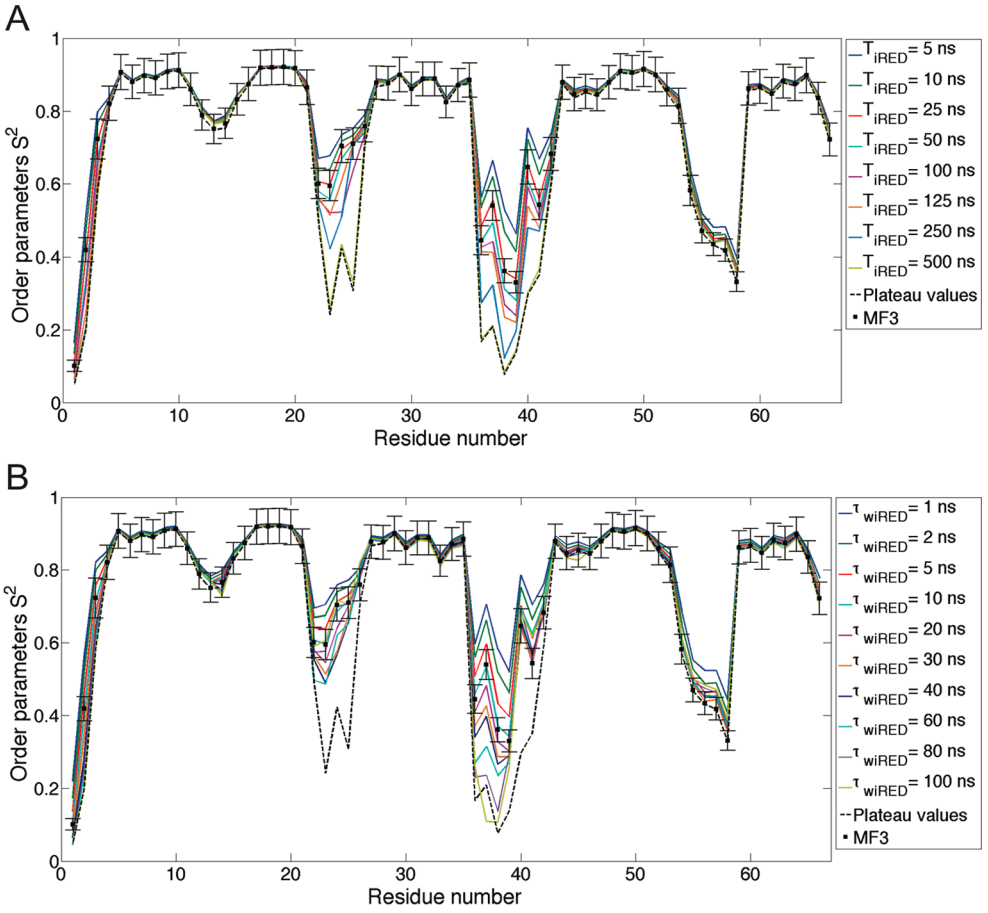


Figure 3. Comparison of S^2 values calculated by different methods for protein Csp A (PDB 1MJC). (A) S^2_{iRED} averaged over different window sizes (solid lines in different colors, see legend on right-hand side of figure), S^2_{MF3} from extended model-free analysis with τ_c = 5 ns (black squares), plateau values of autocorrelation functions from MD trajectory (dashed line). (B) Comparison of S^2_{wiRED} with different window sizes (solid lines in different colors), S^2_{MF3} by extended model-free analysis at τ_c = 5 ns (black squares), plateau values of autocorrelation functions from MD trajectory (dashed line).

MF3 becomes insensitive to these motions resulting in the observed discrepancy.[50] Figure 3 demonstrates the importance of the proper selection of the iRED window.

The analogous analysis with wiRED (also with five bond vectors per (non-Gly) amino acid) uses 10 memory times τ_{wiRED} of 1, 2, 5, 10, 20, 30, 40, 60, 80, and 100 ns (Figure 3B). For τ_c = 5 ns, the best agreement between wiRED and MF3 is obtained for τ_{wiRED} = 10 ns with R = 0.9990 and χ^2 = 0.096. When τ_c is increased, τ_{wiRED} needs to be increased accordingly following the relationships described below.

The iRED and wiRED methods show a similarly good performance. The iRED approach treats each snapshot within the T_{iRED} window with equal weight, whereas wiRED uses differential weights depending on the time difference between snapshots. The former averaging corresponds to a "hard" threshold, while the latter averaging can be viewed as "soft" thresholding as visualized in Figure S1 in the Supporting Information.

3.3. Dependence of Optimal iRED and wiRED Windows on τ_c. Next, we modeled the effect of protein size and solvent viscosity by varying the reorientational tumbling correlation time, τ_c. Three different τ_c values of 5, 10, and 20 ns were tested. Spectral densities and relaxation parameters R_1, R_2, and NOE were computed according to eqs 4−7, followed by MF3 model fitting to determine the dependence of S^2_{total} and τ_{int} on τ_c. The comparison of extended model-free results and iRED results for the three different τ_c values is shown in Figure 4. The χ^2 plots between S^2 from MF3 versus iRED with variable window lengths for different τ_c values all display a single, well-defined minimum. As can be clearly seen, when τ_c is increased from 5 to 20 ns, the optimal iRED window length with the lowest χ^2 shifts from 25 to 100 ns. The corresponding correlation coefficients, R, reach their optimal value very close to 1.0 at very similar, occasionally slightly shorter, iRED window lengths than the χ^2. Minimal χ^2 values for τ_c of 5, 10, or 20 ns are found for iRED window lengths of 25, 50, and 100 ns, respectively. Scatter plots of S^2 from iRED with the shortest (5 ns), longest (500 ns), and optimal window lengths against the corresponding S^2 from MF3 are depicted in Figure 4D−F. Errors introduced when using iRED windows that are either too short or too long are most evident for S^2 < 0.8.

Analysis of the dependence of iRED of the optimal window size on τ_c values suggests the approximate linear relationship

$$T_{iRED} \cong 5\tau_c \quad (15)$$

This relationship serves as a general guide for choosing the optimal iRED window length for an objective comparison between simulated and spin-relaxation-derived S^2 order parameters.

An analogous relationship holds between wiRED memory times τ_{wiRED} and τ_c as can be seen in Figure 5 yielding optimal τ_{wiRED} times of 10, 20, and 30 ns for τ_c of 5, 10, and 20 ns, respectively.

$$\tau_{wiRED} \cong 2\tau_c \quad (16)$$

dx.doi.org/10.1021/ct500181v | J. Chem. Theory Comput. 2014, 10, 2599−2607

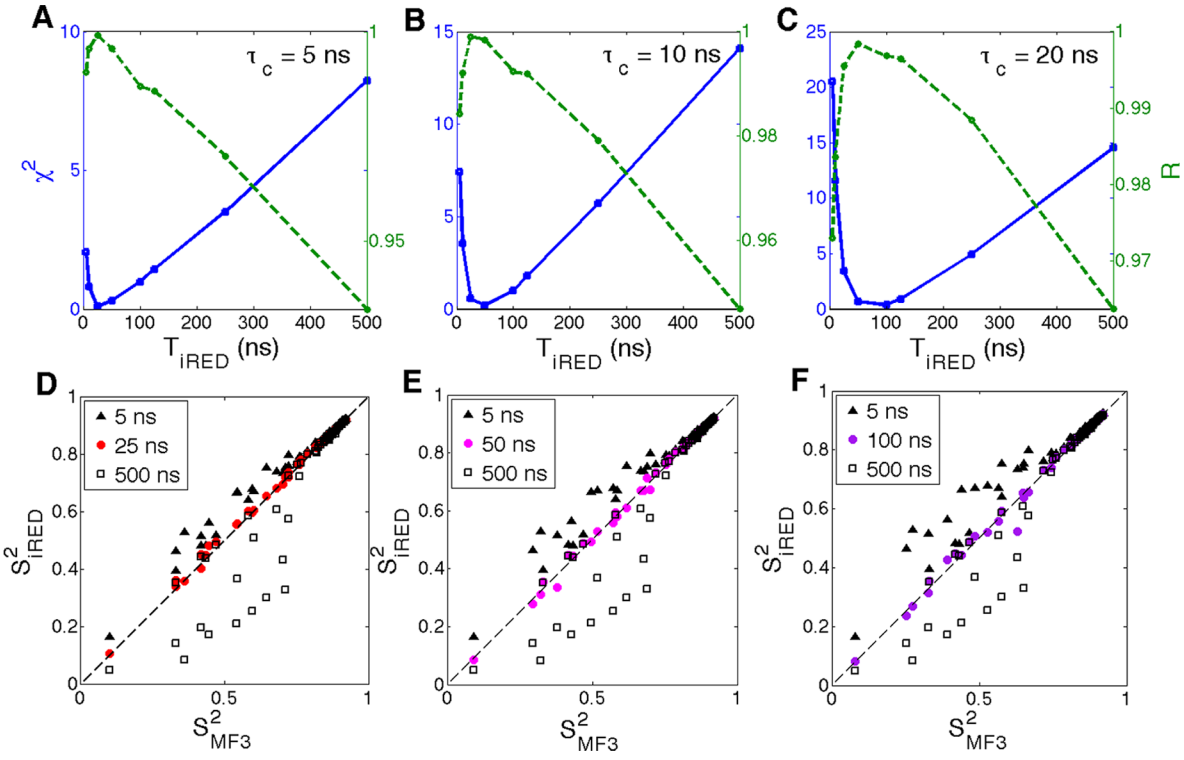


Figure 4. Plot of χ^2 error (blue solid line) and Pearson's correlation coefficient R (green dashed line) between S^2_{iRED} for different averaging window sizes and S^2_{MF3} at (A) $\tau_c = 5$ ns, (B) $\tau_c = 10$ ns, and (C) $\tau_c = 20$ ns for protein Csp A (PDB 1MJC). Scatter plots of S^2_{MF3} and S^2_{iRED} for averaging window sizes of 5 ns (triangles) and 500 ns (squares), and the optimal window size (filled circles): (D) $T_{iRED} = 25$ ns for $\tau_c = 5$ ns, (E) $T_{iRED} = 50$ ns for $\tau_c = 10$ ns, and (F) $T_{iRED} = 100$ ns for $\tau_c = 20$ ns.

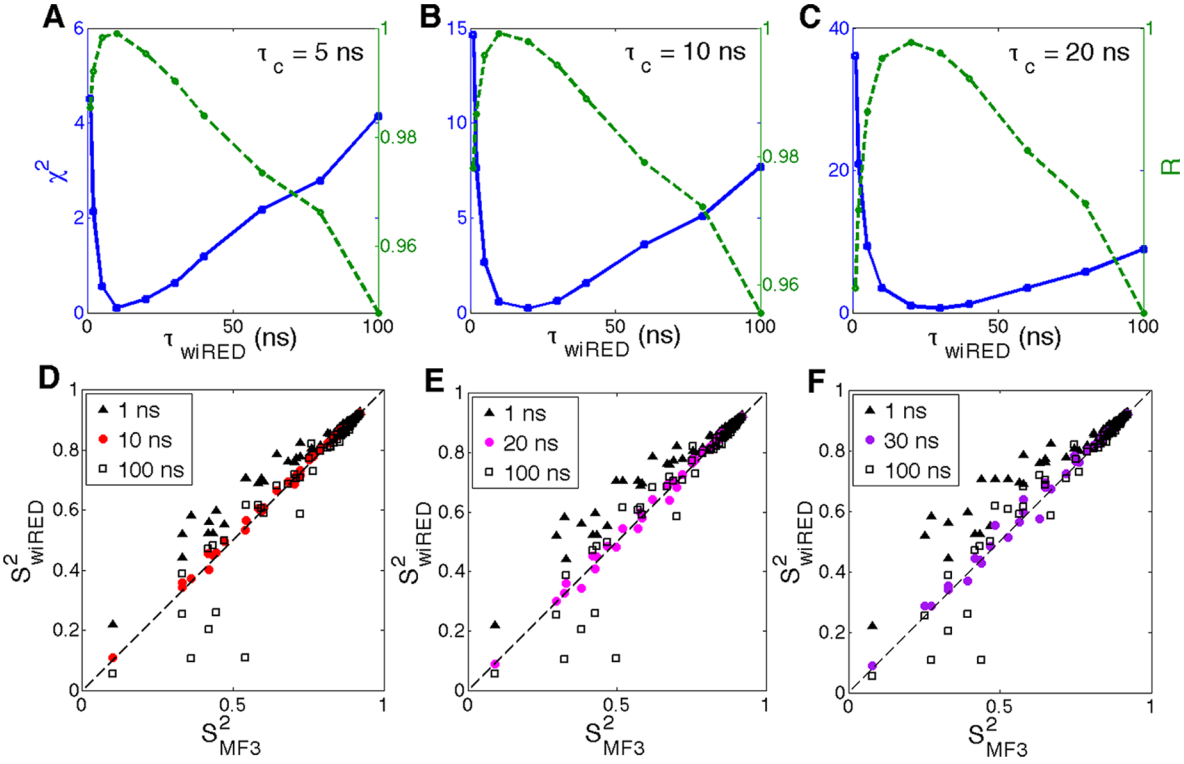


Figure 5. Plot of χ^2 error (blue solid line) and correlation coefficient R (green dashed line) between S^2_{wiRED} with different memory times and S^2_{MF3} at (A) $\tau_c = 5$ ns, (B) $\tau_c = 10$ ns, and (C) $\tau_c = 20$ ns for protein Csp A (PDB 1MJC). Scatter plots of S^2_{MF3} and S^2_{wiRED} for memory times of 1 ns (triangles) and 100 ns (squares), and the optimal memory time window (filled circles): (D) $\tau_{wiRED} = 10$ ns for $\tau_c = 5$ ns, (E) $\tau_{wiRED} = 20$ ns for $\tau_c = 10$ ns, and (F) $\tau_{wiRED} = 30$ ns for $\tau_c = 20$ ns.

dx.doi.org/10.1021/ct500181v | J. Chem. Theory Comput. 2014, 10, 2599−2607

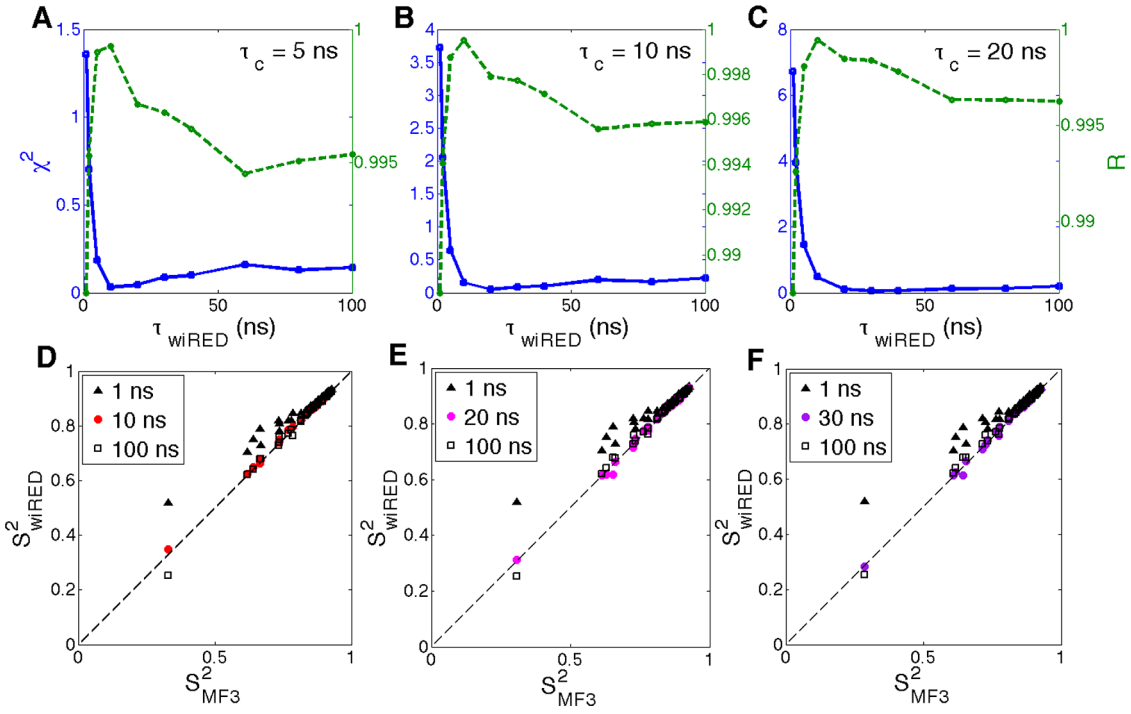


Figure 6. Plot of χ^2 error (blue solid line) and correlation coefficient R (green dashed line) between S^2_{wiRED} with different memory times and S^2_{MF3} at (A) $\tau_c = 5$ ns, (B) $\tau_c = 10$ ns, and (C) $\tau_c = 20$ ns for protein GB3 (PDB 1IGD). Scatter plots of S^2_{MF3} and S^2_{wiRED} for memory times of 1 (triangles) and 100 ns (squares), and the optimal window sizes (filled circles): (D) $\tau_{wiRED} = 10$ ns for $\tau_c = 5$ ns, (E) $\tau_{wiRED} = 20$ ns for $\tau_c = 10$ ns, and (F) $\tau_{wiRED} = 30$ ns for $\tau_c = 20$ ns.

These results are corroborated for additional eight proteins described in Figures S2−S9 in the Supporting Information. They all show the same trend as the one observed for protein Csp A and expressed by eqs 15 and 16, underlining the importance of the proper choice of the iRED window (and wiRED memory time) to accurately reproduce generalized order parameters extracted from NMR relaxation data by model-free analysis. For proteins that display dynamics predominantly on the sub-nanosecond time scale, the choice of τ_{wiRED} is less critical as is shown for protein GB3 (PDB 1IGD) (see Figure 6). The χ^2 error of S^2 still follows the relationship of eq 15, but displays a much shallower behavior once the wiRED memory time exceeds the optimal size. Hence, the precise choice of τ_{wiRED} is less critical as long as τ_{wiRED} is comparable to or larger than $2\tau_c$. Figure 6D−F illustrates the decrease of S^2_{wiRED} for increasing τ_{wiRED} until $\tau_{wiRED} \cong 2\tau_c$. When τ_{wiRED} is further increased to 100 ns, i.e., when $5\tau_{wiRED}$ corresponds to the full trajectory length, the S^2_{wiRED} values do not change noticeably.

4. DISCUSSION

NMR spin-relaxation-derived S^2 values play an important role for validating the amount of dynamics observed in MD simulations with direct consequences for the assessment of the suitability of the computational protocol, the starting protein structure, and the quality of MD force fields.[26,51,52] Once a MD trajectory has been validated, one can identify new information about biomolecular properties from the simulation(s) that are hard to obtain by experiments alone, such as correlated motions and protein recognition behavior. Furthermore, such information can help actively improve the quality of molecular dynamics force fields similar to the use of NMR chemical shifts and residual dipolar couplings of proteins.[3,36]

Among the different methods for determining NMR order parameters from MD trajectories, iRED and wiRED possess advantages as both approaches can be easily implemented without requiring the explicit elimination of overall tumbling effects for each snapshot. Moreover, iRED and wiRED analyses do not require the computation of autocorrelation functions and their conversion to spectral densities for the back-calculation of R_1, R_2, and NOE parameters (equations eqs 4 − 7), although it is always recommended to validate the convergence of the trajectories for a meaningful interpretation of the computed S^2 values. Our results show that, by choosing a suitable averaging window, iRED and wiRED are both capable of reproducing with remarkable accuracy the NMR relaxation-derived order parameters S^2_{MF3} using the model-free approach.

In a previous study, iRED-derived S^2 values averaged over a short ∼1 ns MD time window were reported to produce the lowest root-mean-square deviation with respect to the experimental data for a flexible region of lysozyme.[26] In another study, it was found that the optimal iRED window is $1.25\tau_c$ based on the comparison of S^2_{iRED} from MD trajectories with experimentally determined S^2 values for a single protein (Gal3).[27] In both these studies the focus was on the accurate reproduction of experimental order parameters by MD simulation, rather than the determination of order parameters that are consistent with the ones extracted from spin relaxation parameters computed from the same trajectory. Compared with experimental results, no significant difference between calculated S^2 values from a 30 ns trajectory and a 1 ns averaged trajectory has been observed for lysozyme starting from PDB structures 6LYT and 1IEE. In the Gal3 study the reproduction of the experimental S^2 profile by MD proved to be a considerable challenge. Therefore, the optimal iRED window was dominated to a significant extent by the ability to reproduce the baseline S^2

dx.doi.org/10.1021/ct500181v | J. Chem. Theory Comput. 2014, 10, 2599−2607

values belonging to regular secondary structures. Such a comparison not only depends on the MD force field, but also on the choice of the average N−H bond length used in the model-free approach. By contrast, the conclusions in our study, by comparing for a set of 10 proteins S^2_{iRED} with S^2_{MF3} values extracted from the same MD trajectory, are largely independent of the details of the force field and the N−H bond length. When comparing MD-derived and experimental S^2 values, one needs to keep in mind that the experimental results, but not the MD-derived ones, include zero-point vibrational effects due to local high-frequency bending motions involving the amide proton.[8,53,54] These effects, which cause a systematic decrease of the experimentally determined S^2_{fast} values by a few percent, can be treated, for example, by a modified effective N−H bond length during model-free analysis of the experimental data in order to make the experimentally derived order parameters directly comparable to the ones predicted by MD.

5. CONCLUSION

With the advent of MD trajectories of proteins in free solution that now extend routinely into the hundreds of nanoseconds and microseconds range, the question about the proper protocol for the back-calculation of NMR relaxation-derived order parameters is more pertinent than ever. In this work, we explored dominant factors that contribute to the accurate back-calculation of order parameters from MD trajectories. The specific mathematical dependence of NMR spin relaxation parameters on internal motional time scales has the consequence that S^2_{MF3} values do not sense internal motions on time scales much slower than τ_c even if such motions are present. To ensure a meaningful comparison, the computed S^2_{iRED} values need to take this into account by limiting the window size T_{iRED} over which they are averaged. With the recent availability of trajectories into the hundreds of nanoseconds range and beyond, this can cause a substantial underestimation of S^2_{iRED} for certain residues. Conversely, if T_{iRED} is chosen too short, e.g., in the hundreds of picoseconds and low nanoseconds range, the resulting S^2_{iRED} values can be overestimated. Clearly, the optimal T_{iRED} should grow with increasing τ_c. In this study, we find that the optimal T_{iRED} is approximately 5-fold τ_c provided that internal motions in the nanosecond range are present. It should be noted that the internal correlation times τ_{int} fitted during MF3 analysis are typically much shorter than $5\tau_c$. These correlation times, which represent effective time scales reflecting motions experienced by a N−H vector on multiple time scales, tend to be biased toward faster time scales.[55]

Furthermore, we find that the use of up to five iRED vectors per amino acid, instead of only the backbone N−H vectors, provides a more stable assessment of the global motion especially for small proteins, thereby preventing the potential underestimation of internal motional amplitudes. The wiRED method is introduced here as an alternative approach to iRED, which exponentially weights the contributions of snapshots as a function of the elapsed simulation time. The iRED and wiRED methods show equivalent performance in terms of their accuracy for the computation of order parameters. For both methods, the choice of the averaging window and memory time, respectively, is critical to accurately match the order parameters determined by model-free analysis. The optimal window length and memory time was found to be $5\tau_c$ for iRED and $2\tau_c$ for wiRED, respectively. The same considerations apply when order parameters are computed based on eq 10. Proper averaging is particularly important for the interpretation of spin relaxation data from protein parts displaying nanosecond time scale dynamics including time scales that are comparable to or somewhat slower than τ_c. The methods described here can be applied to the refinement of protein force fields so that they not only match fluctuation amplitudes, such as order parameters, but also kinetic data, including internal motional time scales.

■ ASSOCIATED CONTENT

Supporting Information

Table S1: list of proteins used in this study. Figure S1: weighting functions for iRED and wiRED analysis. Figures S2−S9: plots of RMS error and Pearson's correlation coefficient between S^2_{wiRED} and S^2_{MF3} for various τ_{wiRED} windows for additional eight proteins. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bruschweiler.1@osu.edu. Tel.: (614) 688-2083.

Funding

This work was supported by the National Science Foundation (Grant MCB-1360966).

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 788−788.
(2) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(3) Li, D. W.; Brüschweiler, R. *Angew. Chem., Int. Ed.* **2010**, *49*, 6778−6780.
(4) Huang, J.; MacKerell, A. D. *J. Comput. Chem.* **2013**, *34*, 2135−2145.
(5) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2013**, *23*, 58−65.
(6) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120−127.
(7) Levy, R. M.; Karplus, M.; Wolynes, P. G. *J. Am. Chem. Soc.* **1981**, *103*, 5998−6011.
(8) Palmer, A. G.; Case, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 9059−9067.
(9) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870−9879.
(10) Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989**, *28*, 8972−8979.
(11) Palmer, A. G., 3rd. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 129−55.
(12) Case, D. A. *Acc. Chem. Res.* **2002**, *35*, 325−331.
(13) Brüschweiler, R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 175−183.
(14) Villa, A.; Stock, G. *J. Chem. Theory Comput.* **2006**, *2*, 1228−1236.
(15) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(16) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559−4570.
(17) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1990**, *112*, 4989−4991.
(18) Chandrasekhar, I.; Clore, G. M.; Szabo, A.; Gronenborn, A. M.; Brooks, B. R. *J. Mol. Biol.* **1992**, *226*, 239−250.
(19) Richter, B.; Gsponer, J.; Varnai, P.; Salvatella, X.; Vendruscolo, M. *J. Biomol. NMR* **2007**, *37*, 117−135.
(20) Markwick, P. R. L.; Bouvignies, G.; Salmon, L.; McCammon, J. A.; Nilges, M.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 16968−16975.
(21) Wong, V.; Case, D. A. *J. Phys. Chem. B* **2008**, *112*, 6013−6024.
(22) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522−4534.
(23) Bernado, P.; Fernandes, M. X.; Jacobs, D. M.; Fiebig, K.; de la Torre, J. G.; Pons, M. *J. Biomol. NMR* **2004**, *29*, 21−35.
(24) Musselman, C.; Al-Hashimi, H. M.; Andricioaei, I. *Biophys. J.* **2007**, *93*, 411−422.

(25) Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961−975.
(26) Koller, A. N.; Schwalbe, H.; Gohlke, H. *Biophys. J.* **2008**, *95*, L4−L6.
(27) Genheden, S.; Diehl, C.; Akke, M.; Ryde, U. *J. Chem. Theory Comput.* **2010**, *6*, 2176−2190.
(28) Lindert, S.; Kekenes-Huskey, P. M.; Huber, G.; Pierce, L.; McCammon, J. A. *J. Phys. Chem. B* **2012**, *116*, 8449−8459.
(29) Rosenman, D. J.; Connors, C. R.; Chen, W.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2013**, *425*, 3338−3359.
(30) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. *Biophys. J.* **2006**, *90*, L36−L38.
(31) Lange, O. F.; van der Spoel, D.; de Groot, B. L. *Biophys. J.* **2010**, *99*, 647−655.
(32) Berendsen, H. J. C.; van der Spoel, D.; Vandrunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(33) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306−317.
(34) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(35) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(36) Li, D. W.; Brüschweiler, R. *J. Chem. Theory Comput.* **2011**, *7*, 1773−1782.
(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(38) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.
(39) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(40) Bremi, T.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1997**, *119*, 4272−4284.
(41) Wangsness, R. K.; Bloch, F. *Phys. Rev.* **1953**, *89*, 728−739.
(42) Bloch, F. *Phys. Rev.* **1956**, *102*, 104−135.
(43) Redfield, A. G. *IBM J. Res. Dev.* **1957**, *1*, 19−31.
(44) Abragam, A. *The Principles of Nuclear Magnetism*; Clarendon Press: Oxford, U.K., 1961.
(45) Palmer, A. G.; Rance, M.; Wright, P. E. *J. Am. Chem. Soc.* **1991**, *113*, 4371−4380.
(46) Dosset, P.; Hus, J. C.; Blackledge, M.; Marion, D. *J. Biomol. NMR* **2000**, *16*, 23−28.
(47) Clore, G. M.; Schwieters, C. D. *J. Mol. Biol.* **2006**, *355*, 879−886.
(48) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 7305−7313.
(49) Schindelin, H.; Jiang, W. N.; Inouye, M.; Heinemann, U. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 5119−5123.
(50) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. O.; Xu, H. F.; Trbovic, N.; Friesner, R. A.; Palmer, A. G.; Shaw, D. E. *J. Phys. Chem. B* **2008**, *112*, 6155−6158.
(51) Trbovic, N.; Kim, B.; Friesner, R. A.; Palmer, A. G. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 684−694.
(52) Soares, T. A.; Daura, X.; Oostenbrink, C.; Smith, L. J.; van Gunsteren, W. F. *J. Biomol. NMR* **2004**, *30*, 407−422.
(53) Brüschweiler, R. *J. Am. Chem. Soc.* **1992**, *114*, 5341−5344.
(54) Sun, H. C.; Long, D.; Brüschweiler, R.; Tugarinov, V. *J. Phys. Chem. B* **2013**, *117*, 1308−1320.
(55) Chen, J.; Brooks, C. L., 3rd; Wright, P. E. *J. Biomol. NMR* **2004**, *29*, 243−257.