JCTC Journal of Chemical Theory and Computation

Starting-Condition Dependence of Order Parameters Derived from Molecular Dynamics Simulations

Samuel Genheden,[†] Carl Diehl,[‡] Mikael Akke,[‡] and Ulf Ryde*,[†]

Department of Theoretical Chemistry, Lund University, Chemical Centre, P.O. Box 124, SE-221 00 Lund, Sweden and Center for Molecular Protein Science, Biophysical Chemistry, Lund University, P.O. Box 124, SE-221 00 Lund, Sweden

Received December 28, 2009

Abstract: We have studied how backbone N−H S^2 order parameters calculated from molecular dynamics simulations depend on the method used to calculate them, the starting conditions, and the length of the simulations. Using the carbohydrate binding domain of galectin-3 in the free and lactose-bound states as a test case, we compared the calculated order parameters with experimental data from NMR relaxation. The results indicate that the sampling can be improved by using several starting structures, taking into account conformational heterogeneity reported in crystal structures. However, the improvement is rather limited, and for 93% of the dihedrals that have alternative conformations in the crystal structures, the conformational space is well sampled even if a single conformation is used as the starting structure. Moreover, the agreement with experimental data is improved when using several short simulations, rather than a single long simulation. In the present case, we find that ∼10 independent simulations provide sufficient sampling, and the ideal length of the simulations is ∼10 ns, which is ∼25% longer than the global correlation time for rotational diffusion. On the other hand, the equilibration time appears to be less important, and our results suggest that an equilibration time of 0.25 ns is sufficient. We have also compared four different methods to extract the order parameters from the simulations, namely, the autocorrelation function and isotropic reorientational eigenmode dynamics using three different window sizes. Overall, the four methods yield comparable results, but large differences between the methods may serve to pinpoint cases for which the calculated parameters are unreliable.

Introduction

Nuclear spin relaxation is a powerful experimental technique that provides site-specific information on dynamics and conformational entropy.[1−3] Such measurements are normally interpreted in the context of the model-free approach,[4−6] yielding a generalized order parameter (S^2) for each studied bond vector. Typically, NMR spectroscopic investigations of conformational dynamics focus on a relatively limited subset of bond vectors, although continuous method development aims to expand this set.[7−11] Thus, investigations based solely on experimental data inevitably undersample the conformational entropy of the system, although recent results suggest that order parameters for selected subsets of bond vectors actually capture conformational entropy quite well.[12]

Therefore, NMR spin relaxation experiments can favorably be combined with molecular dynamics (MD) simulations to augment the information content of experimental order parameters.[13,14] MD simulations provide a detailed picture of the motions of all atoms considered, with an accuracy and precision similar to that of NMR experiments.[15] Thus, MD simulations offer a route to interpret in greater detail the results from spin relaxation experiments, provided that the two techniques yield commensurate results.[16] In particular, MD simulations can provide the probability distribution of the conformational substates, including those degrees of

* Corresponding author. Tel: +46 - 46 2224502. Fax: +46 - 46 2228648. E-mail: Ulf.Ryde@teokem.lu.se.

[†] Department of Theoretical Chemistry.

[‡] Center for Molecular Protein Science.

10.1021/ct900696z
Published on Web 05/27/2010

freedom that are not probed by spin relaxation measurements. Once an MD-generated conformational ensemble has been validated by experimental NMR data, it is therefore possible to calculate the total conformational entropy of the system and to address other issues such as the degree of coupling between bond vector motions. In addition, MD simulations offer a high-resolution view of the motional mechanisms that cannot be determined directly from the NMR relaxation data. Conversely, comparisons of order parameters obtained with MD and NMR have frequently been used to judge and improve the quality of MD force fields.[13,17−19]

A major issue for the calculation of generalized S^2 order parameters[20] from MD simulations is the convergence—typically quite long simulations are needed to reach convergence. Related to this issue is how the results depend on the starting conditions of the MD simulations. Several studies have shown that results of MD simulations, e.g., order parameters, strongly depend on the starting structure.[21−23] It has been much discussed whether it is more favorable to run a single long simulation or several shorter ones.[24−26]

In this paper, we examine a related problem: Many crystal structures, especially those obtained at a high resolution, show residues with multiple conformations. This provides a practical problem for MD simulations, because only a single structure is normally treated in the simulations. Which of these conformations should be selected as the starting structure, and how do the results depend on this selection? Is it necessary to start from many different conformations to cover the conformational space appropriately? Can we use this information to speed up the convergence of calculated properties? In this paper, we provide a systematic investigation of these issues. In particular, we study how the calculated order parameters vary and how they compare to experimental NMR data.

As a model system, we have studied the carbohydrate-recognition domain of galectin-3 (Gal3), for which high-resolution X-ray structures are available.[27] Galectins represent a family of proteins that preferentially bind β-galactoside-containing glycans composed of N-acetyllactosamine.[28,29] They are involved in a wide variety of extracellular and intracellular processes, e.g., cancer,[30,31] immunity, inflammation,[32,33] and RNA splicing.[34,35] The Gal3 structure consists of two antiparallel β sheets of six and five strands (Figure 1).[27,28] The saccharide-binding site is defined by a shallow groove formed by the six-stranded β sheet and surrounding loops. Galectin−monosaccharide interactions are relatively weak, with dissociation constants on the order of 0.1−1 mM. The binding free energy is in general dominated by enthalpic contributions and has a minor unfavorable entropic contribution.[36] Typically, two to five hydrogen bonds are formed between the carbohydrate ligand and Gal3, in addition to favorable van der Waals interactions.

Methods

MD Simulations. The carbohydrate-recognition domain of the protein galectin-3 (Gal3) was studied both in the unbound form (Gal3-apo) and in complex with lactose (Gal3-Lac). The simulations were based on an unpublished 1.08 Å structure of Gal3-apo and on a 1.35 Å structure of Gal3-Lac (PDB code 2nn8).[27] The two structures are very similar, with a backbone RMSD of 0.22 Å. In both crystals, several residues are reported with two conformations (27 for Gal3-apo and 16 for Gal3-Lac), all with an occupancy of 0.5. In the MD simulations, we need to select one of those conformations for the starting structure, but the choice is totally arbitrary and might possibly bias the final results. To investigate the effect of the selected conformation, residues were divided into five groups of nearby residues, and the 32 possible permutations of these groups were prepared in which residues in the same group had the same conformation, A or B. The groups are specified in Tables S1 and S2 in the Supporting Information. For about a third of the residues with alternative conformations, there is a change in the hydrogen-bond pattern around that residue, as is also specified in Tables S1 and S2.

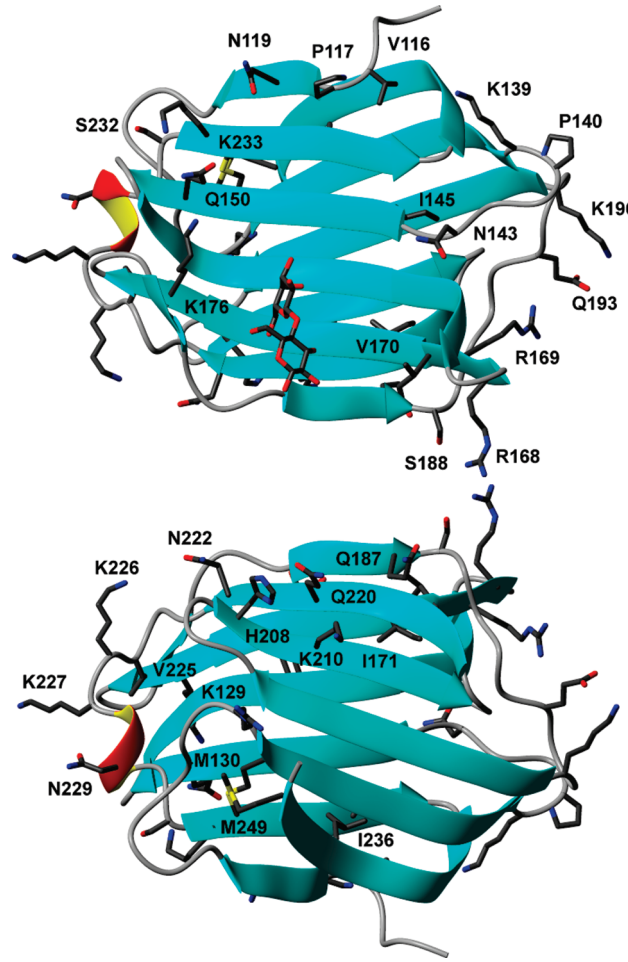

Figure 1. Structure of Gal3 with important residues indicated: (top) the side with the binding site, (bottom) the back side.

All simulations were run using the Amber 10 sander module.[37] The lactose molecule was described with the glycam06 force field and the protein with the Amber99SB force field.[17] Protons were added with the leap module of Amber, and the protonation states were as described previously.[38] The systems were solvated in an octahedral box of TIP4P-Ewald waters,[39] extending at least 9 Å from the protein. The SHAKE algorithm[40] was used to constrain bonds involving hydrogen atoms, making a 2 fs time step

possible. The temperature was kept constant at 300 K using Langevin dynamics[41] with a collision frequency of 2.0 ps^{-1}. The pressure was kept at 1 atm using a weak-coupling approach,[42] with isotropic position rescaling and a relaxation time of 1 ps. Long-range electrostatics were treated with the particle-mesh Ewald approach[43] with a fourth-order B-spline interpolation and a tolerance of 10^{-5}. The nonbonded cutoff was 8 Å, and the nonbonded pair list was updated every 50 fs.

The systems were energy minimized for 1000 steps, restraining all water molecules and heavy atoms to their start positions with a force constant of 418 kJ mol^{-1} $Å^{-2}$. This was followed by a 20 ps equilibration with the same restraints and constant pressure, 50 ps equilibration without any restraints at constant pressure, and 200 ps equilibration at constant volume and no restraints. Finally, a 20 ns production run was performed, still at a constant volume. Coordinates were saved every 1 ps for the calculation of order parameters. On the basis of the stability of the backbone RMSD, the first 5 ns were discarded from subsequent analysis, unless otherwise stated.

We also performed 10 independent simulations of the proteins, started with all residues in the A conformation. The simulation protocol was as described above, but the production simulation was extended to 40 ns. The independent simulations were generated by using different random starting velocities.

Thus, we have run 32 simulations of 20 ns length, starting from different conformations, and 10 simulations of 40 ns length, starting from the same A conformation, but with different velocities. In the following, we will discuss the results obtained from different subsets of these simulations. These subsets will be referred to by the number of simulations, followed by the letter M for mixed conformations or A for the A conformation and then by the length of the simulation in nanoseconds, preceded by "×". For example, 10 simulations of 5 ns length, started from different conformations will be denoted 10M×5.

MD-Derived Order Parameters. Two different methods were used to calculate order parameters from the MD simulations. In the first, order parameters were estimated from the plateau value ($\lim(t \rightarrow \infty)$ C_2 (t)) of the following time autocorrelation function (ACF):

$$C_2(t) = A\langle 3(\vec{\mu}(\tau)\vec{\mu}(\tau + t))^2 - 1\rangle \quad (1)$$

where A is a constant (including the length of the N−H vector) and the average was calculated over the trajectory.[44] The unit vectors $\mu(\tau)$ and $\mu(\tau + t)$ describe the orientation of the N−H vector of interest at times τ and $\tau + t$ in relation to a fixed reference frame. This ACF was calculated using the Amber 10 ptraj module, and the overall tumbling was removed by fitting the backbone heavy atoms to the first snapshot. It is not fully straightforward to determine the plateau value of C_2, because C_2 becomes noisy at large values of the time delay, t, owing to the finite sampling time. Therefore, C_2 was only calculated for t up to ∼1/10 of the total simulation time.[45−47] The order parameters were then obtained by fitting $C_2(t)$ to an exponential function of the form

$$A + B\,e^{-Ct} + D\,e^{-Et} \quad (2)$$

where A, B, C, D, and E are fitted coefficients[20] and the order parameter can be identified with A.[44,13] Statistical errors of the order parameters were estimated using a bootstrap procedure on the residuals from the exponential fit, using 1000 samples.[48]

Alternatively, order parameters were extracted using the isotropic reorientational eigenmode dynamics (iRED) approach.[49] In this approach, the following covariance matrix

$$M_{ij} = \frac{1}{2}\langle 3(\vec{\mu}_i\vec{\mu}_j)^2 - 1\rangle \quad (3)$$

of $\mu(\tau)$ for different N−H vectors was calculated using the Amber 10 ptraj module. The eigenvalues, λ_m, and eigenvectors $\vec{m}$ were then obtained by diagonalization, and the order parameters for residue i were calculated from

$$S_i^2 = 1 - \sum_{m=6}^{n} \lambda_m |m_i|^2 \quad (4)$$

where the sum runs over all internal modes, i.e., all except those with the five largest eigenvalues, and m_i is the ith element of $\vec{m}$. Order parameters were calculated either by using the entire trajectory or by averaging over 1 or 5 ns windows. The latter was tested because, if the length of a simulation exceeds the overall tumbling correlation time of the protein, S^2 parameters computed over the whole trajectory can include motions that would not be reflected in the experimental S^2 values, leading to a bias in the computed S^2 values.[16,50]

Dihedral Distributions. To identify the distribution of dihedral angles of interest, we employed a Gaussian-mixture model (GMM).[51] This approach models the total distribution as a sum of Gaussian (normal) distributions. Each of these distributions will be referred to as a state. In this study, we are only interested in one-dimensional distributions, and hence we employ univariate Gaussian distributions. The probability that a data point (a dihedral angle, denoted y in the following formulas) comes from state k is denoted π_k, and the distribution of each class is

$$p(y|\text{ from class } k, \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{-(y - \mu_k)^2}{2\sigma_k^2}\right) \quad (5)$$

where μ_k and σ_k^2 are the mean and variance of state k. The total distribution is

$$p(y|\pi, \mu, \sigma) = \sum_k \pi_k p(y|\text{ from class } k, \mu_k, \sigma_k) \quad (6)$$

To determine which state each data point belongs to and the values of the parameters π_k, μ_k, and σ_k^2, we use an expectation-maximization algorithm.[52] This is an iterative algorithm that starts with an initial guess of the parameters and then iteratively updates the parameters until convergence. Initially, we assume that there are four states that are uniformly distributed between −180 and +180, and that all states have equal probability. If the probability of a state in any iteration falls below 0.001, that state is discarded. In each iteration, the parameters are updated as follows:

$$\pi_k = \frac{1}{n}\sum_{i=1}^{n} w_{ik}$$
$$\mu_k = \frac{1}{n\pi_k}\sum_{i=1}^{n} w_{ik} y_i \quad (7)$$
$$\sigma_k^2 = \frac{1}{n\pi_k}\sum_{i=1}^{n} w_{ik}(y_i - \mu_k)^2$$

where the weight w_{ik} is determined from the data and the values of the parameters in the previous iteration (old)

$$w_{ik} = \frac{p(y|\text{ from class } k, \pi_k^{\text{old}}, \mu_k^{\text{old}}, \sigma_k^{\text{old}})\pi_k^{\text{old}}}{\sum_k p(y|\text{ from class } k, \pi_k^{\text{old}}, \mu_k^{\text{old}}, \sigma_k^{\text{old}})\pi_k^{\text{old}}} \quad (8)$$

Statistical Analysis. Twelve quality measures were employed to judge how well the calculated order parameters (S^2_{MD}) reproduce the measured ones (S^2_{NMR}), viz., the median, the correlation coefficient (r^2), the root mean squared deviation (RMSD), the mean signed deviation (MSD), the mean absolute deviation (MAD), the mean absolute deviation with systematic error removed (MADtr, i.e., after subtraction of the MSD), the mean quote (MQ; $S^2_{\text{MD}}/S^2_{\text{NMR}}$), and the Q value ($Q = (\sum_i(S^2_{i,\text{MD}} - S^2_{i,\text{NMR}})^2)/(\sum_i(S^2_{i,\text{NMR}})^2)$).[53] We also calculated how many of the experimental order parameters fall outside the range of the calculated order parameters among the set of simulations of the same type. This measure was also calculated when the range was extended by 0.01, 0.05, and 0.1 in each direction.

Errors in the various qualities were estimated from the standard deviations in S^2_{MD} and S^2_{NMR} by performing a random simulation: S^2_{MD} and S^2_{NMR} for each residue was assigned a random number from a normal distribution, with the mean and standard deviation obtained in the MD simulations or NMR measurements. Then, we calculated all the quality measures and repeated this procedure 10 000 times. The standard deviations within these sets are reported as the standard error of the quality estimates.

For the comparison of various methods or simulation protocols, we estimate the significance of each prediction by calculating the probability that a certain method will be best the observed number of times or more, using a binomial distribution, assuming equal probability for all methods or simulations. In this calculation, quality measures that give the same results for all methods were omitted.

NMR Relaxation Data. The acquisition and analysis of the NMR relaxation data for the backbone N−H groups have been described.[38] In comparing order parameters from NMR and MD, it should be kept in mind that the former depends on assumptions regarding the N−H bond length and chemical shift anisotropy of the ^{15}N nucleus.[13] Residue-specific variations in these parameters are not captured by the present approach. Furthermore, for the purposes of the present comparisons, we have also considered the potential effects of additional systematic errors, as follows.

Accurate interpretation of relaxation rates in terms of order parameters requires high-resolution structural information if the protein exhibits anisotropic global rotational diffusion, because the relaxation rates depend on the orientation of the N−H bond vector in the molecular frame. In the case of Gal3-apo, the loops surrounding the saccharide-binding site have different conformations in the low-resolution NMR structure[54] and the high-resolution X-ray structure, which can be attributed to intermolecular contacts in the crystal.[27] In principle, this discrepancy suggests that the experimental S^2 values determined for the loop residues in question might suffer from systematic errors. However, Gal3-apo has a modest anisotropy of 1.07, indicating that the potential errors in S^2 should be less than 3%.

The presence of conformational exchange contributions to R_2 requires that the model-free optimization includes an exchange term, R_{ex}. Deviation of the fitted R_{ex} from the actual exchange contribution leads to inaccuracy of the fitted S^2 values. To account for this, in some cases, we have omitted those residues that have been fitted with R_{ex} terms. However, in the case of Gal3, the model-free optimizations appear to be robust. Using reduced data sets excluding R_2, we obtain nearly identical order parameters for both Gal3-apo and Gal3-lac, with a weighted RMSD versus the full data sets (including R_2) of 0.007 in both cases.

Result and Discussion

Method to Calculate Order Parameters. Before studying the starting-condition dependence, we addressed which method to use to calculate order parameters. As described in the Methods section, we tested both the ACF and iRED approaches. In the latter case, order parameters were calculated either by using the entire trajectory or by averaging over 1 or 5 ns windows (these methods will be called iRED-full, iRED-1, and iRED-5 in the following). To compare the four methods, we used the 12 quality measures described in the Methods section to judge how well the calculated order parameters (S^2_{MD}) reproduce the measured ones (S^2_{NMR}), taken from our previous investigations of Gal3.[38] These comparisons were done both for Gal3-apo and Gal3-Lac. We also studied the difference in order parameters between Gal3-apo and Gal3-Lac, ΔS^2. All the results are collected in Tables S3−S5, in the Supporting Information.

Unfortunately, the various quality measures give different results, as do the simulations on different proteins. The correlation coefficient is in general highest with the iRED-5 method, but the correlation is rather poor for all methods, up to 0.35 and 0.43 for Gal3-apo and Gal3-Lac, respectively, and less than 0.07 for ΔS^2. Such a correlation is worse than what has been observed in most previous studies, in which S^2_{NMR} and S^2_{MD} have been compared, 0.22−0.93.[17,18,23,55−60] The reason for this is that r^2 for Gal3 strongly depends on a few residues with a low S^2, which often are poorly determined by NMR (as will be discussed more below). On the other hand, the RMSD, median, and MAD are actually better than observed in the great majority of previous studies: The RMSD is 0.04−0.06, compared to 0.02−0.26, in the previous studies, with an average of 0.09. In fact, only one investigation in our survey gave an RMSD lower than in the present comparison, 0.02−0.04.[58] Likewise, both the median (−0.02 to 0.03) and the MAD (0.03−0.04) are lower than in previous studies (0.04 and 0.06−0.11,[55,56] respec-

Table 1. Residues for which the Four Methods to Calculate S^2_{MD} Give a Range Larger than 0.05 in the Different Simulations

simulation	32M×20[a]		10A×40[a]		10A×20[a]		1A×40[a]		1A×20[a]		32M×20[b]		10M×10[b]	
residue	Lac	Apo	Lac	Apo	Lac	Apo	Lac	Apo	Lac	Apo	Lac	Apo	Lac	Apo
Ile115	0.13	0.12	0.18	0.16	0.13	0.13	0.26	0.11	0.16	0.13	0.13	0.12	0.16	0.25
Val116	0.12	0.07	0.12	0.11	0.11	0.09	0.12	0.09	0.12	0.06	0.12	0.07	0.15	0.15
Gly125	0.06	0.06	0.07		0.06		0.09	0.10	0.09	0.12			0.14	0.24
Val126			0.07	0.06					0.07					
Ala142														0.09
Asp154									0.12					
Val155	0.06	0.10	0.10	0.12	0.06	0.08	0.09	0.12		0.07	0.06	0.10	0.11	0.15
Arg168								0.06						
Arg169					0.06				0.07	0.08				
Leu177	0.09	0.08	0.10	0.16	0.09	0.14		0.13		0.23	0.09	0.08	0.18	0.16
Asn179										0.08				
Asn180										0.06				
Arg183														0.08
Glu184				0.06					0.12					0.06
Arg186										0.08				
Val189								0.11		0.11			0.06	0.08
Phe192						0.06			0.06				0.08	0.07
Asp207										0.07				
Val213										0.08				
Ala216													0.06	
Leu219										0.08				
Arg224									0.07	0.06				
Lys227			0.07	0.08						0.07			0.08	0.07
Leu228		0.06	0.07	0.06	0.00	0.06		0.13	0.09	0.06			0.06	0.06
Ile231									0.09					
Ser232	0.10	0.14	0.16	0.15	0.11	0.12	0.24	0.15	0.17	0.13	0.09	0.14	0.18	0.18
Ser246										0.07				
Ile250							0.10	0.10	0.09	0.18				0.07

[a] Results based on all four methods. [b] Results based only on the iRED-1 and iRED-full methods.

tively). Thus, the accuracy of the present investigation seems to be similar or better than in previous studies.

The number of NMR values outside the simulated range is typically lowest for the ACF method, but this criterion may favor methods with a poor precision. For the other quality measures, the iRED-1 method gives the best results, at least for the Gal3-apo simulation and the difference. For RMSD, which gives a high weight to outliers, the iRED-full method works better for Gal3-Lac, and for the median, MSD, and MQ, which give a low weight to outliers, ACF performs better for Gal3-Lac. For MAD and MADtr, which give an intermediate weight to outliers, iRED-1 is always best. On the basis of these results, it is hard to point out a single method as the best. In the following, we will use iRED with 1 ns windows, simply because it had the best average performance.

Most importantly, the four different methods give closely similar results for most of the order parameters. In fact, only for seven residues (out of 127), the largest difference among the four methods is larger than 0.05 in the 32 simulation using different starting conformations (32M×20). These residues are listed in Table 1. This shows that, for the great majority of the residues, it does not matter what method is used, whereas for a few residues, different methods give differing results, indicating problems to accurately determine S^2_{MD}. Thus, a large difference between the four methods can be used as a criterion to decide what residues have a poorly determined S^2_{MD}, and these could then be excluded from comparisons. As the order parameters calculated with ACF deviate most from those calculated with iRED-1, it is sufficient to calculate order parameters with these two methods to decide which residues have a poorly determined S^2_{MD}. However, if the simulation time is short, many of the ACFs will not be converged. Therefore, we recommend using iRED without windowing as a second opinion. From Table 1 (sixth set of columns), it can be seen that this only slightly changes the results.

Conformational Sampling. Next, we turn to simulations started at different structures, based on the alternative conformations in the crystal structure. As detailed in the Methods section and described in Tables S1 and S2 (Supporting Information), we have run 32 simulations of 20 ns length for both Gal3-apo and Gal3-Lac, based on a permutation of five groups of alternative configurations observed in the crystal structure.

A natural question is whether the protein stays in the same conformation during the simulations or if it moves between the various conformations freely. To answer this question, we defined a set of 30 dihedral angles that describe the differences of most of the alternative conformations observed in the crystal structure. They are shown in Table 2.

These dihedral angles were followed throughout the MD simulations. We used a Gaussian-mixture model (GMM)[51] to identify the number of maxima in the distribution function, the dihedral angle at the maxima, and the percent of the time the system spent in each conformation. A typical example is shown in Figure 2.

All 30 dihedral angles describe rotations around a C−C single bond. Therefore, three distinct conformations are expected, rather than the two conformations modeled into the crystal structures. This is confirmed by the simulations: 21 of the 30 angles showed three conformations with a significant probability (>1%; the conformational states

Table 2. Definition of Dihedral Angles Used to Characterize the Conformational Sampling[a]

			Apo		Lac				
residue	dihedral angle	structure	A	B	A	B	S1	S2	S3
Val 116	N-CA-CB-CG2	Apo	−89	90	−57		−61	60	
Pro 117	N-CA-CB-CG	Apo	−6	20	23		−23	27	
Asn 119	N-CA-CB-CG	Lac	178		−64	178	−171	−71	77
Arg 129	CA-CB-CG-CD	Lac	−167		**175**	**−153**	−180	13	53
Met 130	N-CA-CB-CG	Apo	**−160**	**175**	−172		−174	−88	83
Ile 134	CA-CB-CG1-CD1	both	96	172	−55	94	−62	89	170
Lys 139	CB-CG-CD-CE	Apo	−177	71	171		−67	68	179
Pro 140	N-CA-CB-CG	Apo	−28	25	−18		−21	25	
Asn 143	CA-CB-CG-ND2	both	−106	138	−121	132	−161	85	
Gln 150	CA-CB-CG-CD	both	−95	146	**−154**	**153**	173		
Arg 168	CB-CG-CD-NE	both	−168	176	**−180**	**167**	−69	69	180
Arg 169	CG-CD-NE-CZ	Lac	164		170	−86	−167	−87	148
Val 170	N-CA-CB-CG2	Apo	−175	69	−171		−168	−56	59
Ile 171	N-CA-CB-CG1	Apo	**−44**	**−76**	−59		−57	68	
Lys 176	CB-CG-CD-CE	Lac	−164		69	−162	−176	−64	70
Gln 187	CB-CG-CD-NE2	Apo	−122	80	111		−88	96	182
Ser 188	C-CA-CB-OG	both	−70	138	−69	152	−69	174	
Glu 193	CA-CB-CG-CD	Apo	−69	171	−175		−180	−65	59
Lys 196	CA-CB-CG-CD	Lac	−161		**−152**	**171**	−175	−64	69
His 208	CA-CB-CG-ND1	Lac	−100		−99	82	−110	97	
Lys 210	CB-CG-CD-CE	Apo	71	−106	76		−66	66	178
Gln 220	C-CA-CB-CG	both	173	−72	172	−61	−62	65	170
Asn 222	CA-CB-CG-ND2	Lac	−71		66	172	−79	72	189
Lys 226	CA-CB-CG-CD	Apo	−64	−179	178		−179	−61	66
Lys 227	CA-CB-CG-CD	Apo	**−167**	**163**	166		−180	−67	70
Asn 229	CA-CB-CG-ND2	Apo	**−172**	**168**	179		−90	88	162
Ser 232	C-CA-CB-OG	Lac	173		−67	171	−69	179	
Lys 233	C-CA-CB-CG	Apo	**165**	**−163**	−145		−61	62	166
Ile 236	CA-CB-CG1-CD1	Apo	176	118	145		−72	91	175
Met 249	C-CA-CB-CG	both	−62	110	−60	−160	−69	71	177

[a] The table shows the residue and the atoms in each dihedral angle, the dihedral angles observed in the two conformations in the two crystal structures, as well as the (up to) three conformations observed in the MD simulations, according to the Gaussian-mixture model. Conformations in the crystal structures that belong to the same MD states are marked in bold face. All dihedral angles are in degrees.

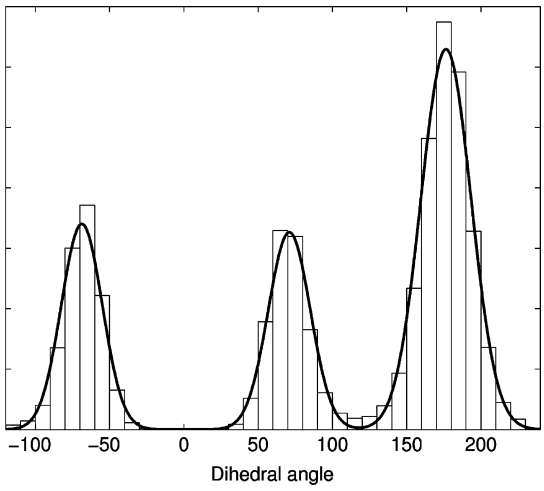

Figure 2. A typical example of the Gaussian-mixture model (GMM) fit for the dihedral angle in Met249. The underlying distribution is shown as a histogram with a bin size of 72 and the GMM is shown as a solid line.

identified with the GMM will be called S1−S3 in the following, whereas the conformations observed in the crystal structures are called A and B). Only for one residue, Gln150, did we find a single significant state (Figure 1 shows the location of important residues in the crystal structure of Gal3). All conformations identified by the GMM are listed in Table 2. It can be seen that most of the conformations identified in the MD simulations are separated by ∼120°, as expected for a 3-fold rotation. However, nine dihedrals had rotamers that were closer than 100°, down to 41° for S2 and S3 of Arg129.

In general, the dihedral angles with maximum probabilities obtained from the MD simulations are fairly close to those observed in the crystal structure. However, in ∼5 cases, the difference is over 30°, and for Gln150 in Gal3-apo, crystal conformation A was not observed in the MD simulation. In nine cases, the difference between the two conformations observed in the crystal structures is so small that they belong to the same state in the MD simulations.

If we compare the results of simulations started with all residues in either the A or B conformation, it can be seen that there is a quite large difference between the average values for the 30 dihedral angles in the two simulations, especially for Gal3-apo (Tables S6 and S7, Supporting Information): The averages differ by up to 121°, and the MADs are 30° and 17° for Gal3-apo and Gal3-Lac, respectively. Apparently, the estimated standard errors (∼1°) for a single simulation grossly overestimate the precision of the averages.

The percentage of the time spent in the various conformations is more stable, but it still shows differences of up to 65% units, with MADs of 11−13 and 8−11% units for Gal3-apo and Gal3-Lac, respectively. This shows that there are significant differences between simulations started from different structures. However, there are also significant dynamics for the studied dihedrals. Only 2−5 dihedrals show a single conformation in the simulations, and for 2−7

additional dihedrals, over 90% of the time is spent in a single conformation. Thus, there is a decent sampling of at least two conformations for most of the dihedrals.

Next, we consider the 32 simulations with the permutations of different conformations. Again, the results (Tables S8 and S9, Supporting Information) show that there is an extensive variation in the results obtained with different starting structures: There are simulations that give completely opposite results for the percentage of the various conformations in the simulations (i.e., some simulations give 100% S1 and others give 100% S2 or S3). This shows that it is mandatory to run several independent simulations to obtain reliable results (or use simulation times much longer than 20 ns).

Interestingly, there is little correlation between what conformations are observed in the crystal structures and the populations of the dihedrals sampled in the simulations. Only in three cases does a residue that has a single conformation in the crystal also populate primarily (>90%) the same conformation in the simulations (Asn119 and His208 in Gal3-apo and Ile171 in Gal3-Lac). For another three dihedrals, MD samples only a single conformation, which essentially covers the two states observed in the crystal structures (i.e., the two conformations observed in the crystal structures are so close that they belong to the same MD conformation). On the other hand, there are three residues that show almost only one conformation in the simulations, but two conformations in the crystal structures (Val116 in Gal3-apo and Asn119 and Asn222 in Gal3-Lac). All of the other residues show two or three conformations in the MD simulations, independently of the number of conformations observed in the crystal structures. The reason for this may be that the resolution of the structures is too low to discern several conformations (some of which have a low occupancy), that crystal packing effects may stabilize certain conformations, or that the low temperature conditions during the X-ray diffraction experiments (employing liquid nitrogen) restricts the number of populated conformational states.

There are some conspicuous differences between Gal3-apo and Gal-Lac. In particular, Asn222 is almost entirely in the S3 state in the Lac simulation, whereas it is only 1% of the time in that conformation for Gal3-apo. Glu193 and His208 also show quite large differences. However, for most of the dihedrals, the occupancy of the various conformations is quite similar, with a MAD of only 8% units.

The prime question in this investigation is whether it is necessary to use different starting structures to obtain a proper sampling or if similar results can be obtained with different means. To this end, we compare the 32 simulations started from different conformations with 10 simulations started from the same structure (all residues in the A conformation), but with different starting velocities. The results in Tables S8 and S9, Supporting Information, show that the 10 simulations started from the same conformation show a slightly smaller sampling. For example, the average range of the percentages (maximum − minimum) of the three states of the dihedrals is 24−32% units for the 10 simulations, but 37−50% units for the 32 simulations. On the other hand, the average standard deviation of the percentage of the three conformations is similar or slightly larger for the 10 simulations, 2−3% units. However, the MAD between the 32 and 10 simulations is only ∼4% units for all three states in both proteins, and these numbers are dominated by three residues from each simulation, (Pro140, Glu193, and Asn222 for Gal3-apo and His208, Gln220, and Lys233 for Gal3-Lac), which show differences of 14−30% units. Thus, there is some advantage of starting the simulations from different conformations, but the effect is rather small.

Another interesting question is how long simulations are needed for converged results. In Tables S10 and S11 (Supporting Information), we compare the results obtained for the 10 independent simulations after 20 and 40 ns of simulation time. It can be seen that the results in general are similar, with MADs of 2−3% units for both complexes, and with maximum differences of up to 7−11% units. They also give similar differences, compared to the 32 simulations. Thus, we can conclude that the conformational sampling is reasonably converged already after 20 ns.

Related to this issue is the time-scale of the conformational changes studied. If it is short, compared to the simulation time, the results should be converged, and then it should also be possible to estimate the equilibrium constants from the observed percentage and the activation barriers from the time-scale. In Tables S12 and S13 (Supporting Information), we therefore list how long it takes before the protein changes the conformation of the various dihedrals. It can be seen that the time varies from 4 ps for Ser188 to 2.8 ns for His208, with an average of ∼0.7 ns for both systems. This indicates that the sampling of 20−40 ns should be appropriate, although the sampling of states with a low occupancy can be worse. For most residues, the simulations starting from 10 different velocities or 32 different structures give a similar result, but for 3−4 residues, the difference is large, up to 6 ns.

Starting-Condition Dependence of Order Parameters. Next, we consider order parameters obtained from the 32 + 10 different simulations of Gal3-apo and Gal3-Lac. Our prime question is how to perform simulations that give the best results, compared to experiments. The results of the 12 different quality measures used to compare S^2_{NMR} to S^2_{MD} are listed in Table 3 for five different sets of simulations, viz., the 32 simulations with different starting structures (20 ns length; 32M×20), the 10 independent simulations starting from the same structure (of either 40 or 20 ns length; 10A×40 and 10A×20), and a single simulation of either 40 or 20 ns length (1A×40 and 1A×20). Figure 3 shows the ranking of the various simulations, i.e., the number of times each of the simulations rank first, second, and so on, for the various quality measures and systems.

From Figure 3, it clear that the 32M×20 simulation is best: It gives the best results for all measures, except the median, for both complexes, as well as for the difference. If we take into consideration the uncertainties in the various quality estimates (both from NMR and MD), the 32M×20 simulation gives significantly better results at the 95% level (according to a Student's *t* test) for 2−6 quality measures compared to the other four simulations (Table S14, Supporting Information).

Table 3. Comparison of S^2_{MD} and S^2_{NMR} for Gal3-apo, Gal3-Lac, and the Difference between the Two Proteins[a]

simulation	RMSD	r^2	MAD	MADtr	MSD	median	MQ	Q	$n \pm 0$	$n \pm 0.01$	$n \pm 0.05$	$n \pm 0.1$
						Gal3-apo						
32M×20	**0.038**	**0.28**	**0.029**	**0.029**	**−0.004**	0.003	**1.00**	**0.002**	**63**	**46**	**6**	**0**
10A×40	0.039	0.24	0.030	0.031	**−0.004**	**0.000**	**1.00**	**0.002**	80	61	9	1
10A×20	0.039	0.26	0.030	0.030	**−0.004**	−0.001	**1.00**	**0.002**	79	57	8	**0**
1A×40	0.041	0.21	0.031	0.031	**−0.004**	**0.000**	**1.00**	**0.002**				
1A×20	0.044	0.20	0.033	0.033	**−0.004**	0.003	**1.00**	0.003				
						Gal3-Lac						
32M×20	**0.062**	**0.35**	**0.041**	**0.036**	**0.028**	0.026	**1.04**	**0.005**	**76**	**59**	**15**	**2**
10A×40	0.063	0.33	**0.041**	0.037	**0.028**	0.026	**1.04**	**0.005**	88	73	23	4
10A×20	0.063	0.33	**0.041**	0.037	**0.028**	**0.025**	**1.04**	0.006	83	70	20	4
1A×40	0.065	0.29	0.042	0.037	0.029	0.026	**1.04**	0.006				
1A×20	0.066	0.28	0.043	0.038	0.030	**0.025**	**1.04**	0.006				
						difference						
32M×20	**0.054**	**0.05**	**0.036**	**0.027**	**0.032**	0.025						
10A×40	0.057	0.01	**0.036**	**0.028**	**0.032**	0.026						
10A×20	0.057	0.01	**0.036**	0.028	**0.032**	0.026						
1A×40	0.059	0.02	0.038	0.030	**0.032**	**0.024**						
1A×20	0.062	0.02	0.040	0.031	0.034	0.026						

[a] The 12 different quality measures listed are the root-mean-squared-deviation (RMSD), Pearson's correlation coefficient (r^2), the mean absolute deviation (MAD), the mean absolute deviation when removing the systematic error (MADtr), the mean signed deviation (MSD), the median, the mean quote (MQ; S^2_{MD}/S^2_{NMR}), the Q value, and the number of residues for which the S^2_{NMR} value falls outside the range of the S^2_{MD} values (when there are several simulations; $n \pm 0$). The latter measure is also calculated when the MD range is extended by 0.01, 0.05, and 0.1 in each direction ($n \pm 0.01$, $n \pm 0.05$, and $n \pm 0.1$). The best result for each quality measure for each system is marked in bold face. The iRED method with 1 ns windows was used to obtain S^2_{MD}, and the equilibration time was 5 ns.

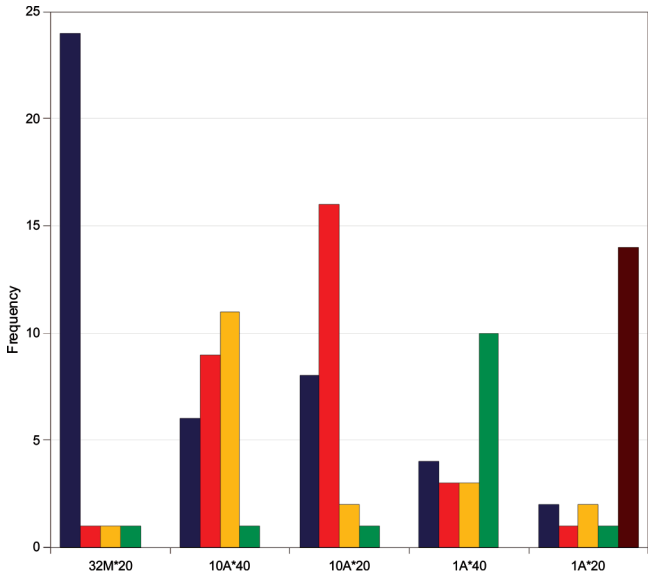

Figure 3. Ranking of the five simulations for the 12 quality measures and three systems (Gal3-apo, Gal3-Lac, and the difference between the two systems). The figure shows the number of times each simulation ranks as number one, two, three, four, or five (from left to right) for each quality measure and simulation.

However, it is also notable how small the differences are: If we instead use a single simulation of 20 ns length, the RMSD increases by up to 0.006, the MAD by 0.004, the MSD by 0.002, and Q by 0.001, whereas the median and MQ hardly change. It is only the correlation coefficient that increases by a larger amount, up to 0.08. Compared to the 10 independent simulations, the differences are even smaller, and the main difference is seen for the total range of the simulated values; that is, it is appreciably larger for the 32 different starting structures (illustrated by a decreased number of S^2_{NMR} values outside the S^2_{MD} range).

Interestingly, there seems to be little gain in running the 10 independent simulations for a longer time (40 ns, rather than 20 ns): Only two quality measures are improved and two become worse, in all cases by a minimal amount. Moreover, the number of S^2_{NMR} values outside the range of the calculated S^2_{MD} range increases, although this only illustrates that it is a poor quality measure, favoring simulations with a poor precision. However, for the single simulation, most of the quality measures are improved if the simulation is prolonged from 20 to 40 ns. Thus, we can conclude that there is small, but consistent, improvement in the results as more simulations are performed. It seems more favorable to run several shorter simulations than one long one.

The analysis above is based on sets of simulations that have different total simulation times. To make a more fair comparison, we devised new sets of simulations, which have a total simulation time of either 20 or 40 ns. Until now, we have used an equilibration time of 5 ns, a decision that was made on the basis of a rather qualitative analysis of the backbone RMSD fluctuations. We therefore studied the effect of equilibration times that ranged from 0 to 19 ns for the 32 simulations of mixed conformations. The results in Table 4 show that the $S^2_{NMR} - S^2_{MD}$ differences are rather insensitive to the equilibration time: Only the correlation coefficient (and RMSD for $\Delta\Delta S^2$) show significant differences between 0 and 19 ns equilibration times. This is because the correlation coefficient is very sensitive to the actual S^2 value of a few vectors (all except 1−4 order parameters are between 0.7 and 1.0 for both NMR and MD). For example, r^2 for $\Delta\Delta S^2$ can increase from 0.04 to 0.24 upon a change of only 0.08 for a single residue in one of the simulations. On the basis of these results, we decided to use an equilibration time of 0.25 ns.

Next, we created seven new sets of simulations, all using an equilibration time of 0.25 ns: Three of them have a total

Table 4. Equilibration-Time Dependence of the Quality Measures for Gal3-apo, Gal3-Lac, and the Difference, Compared to NMR[a]

	Gal3-apo								Gal3-Lac								Difference					
time (ns)	RMSD	r^2	MAD	MADtr	MSD	Med	MQ	Q	RMSD	r^2	MAD	MADtr	MSD	Med	MQ	Q	RMSD	r^2	MAD	MADtr	MSD	Med
0	0.038	0.409	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.358	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.040	0.036	0.027	0.032	0.025
0.1	0.038	0.409	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.357	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.037	0.036	0.027	0.032	0.025
0.2	0.038	0.409	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.357	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.035	0.036	0.027	0.032	0.025
0.3	0.038	0.408	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.357	0.041	0.036	0.028	0.027	1.040	0.005	0.054	0.037	0.036	0.027	0.032	0.025
0.4	0.038	0.408	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.357	0.041	0.036	0.028	0.027	1.040	0.005	0.054	0.037	0.036	0.027	0.032	0.025
0.5	0.038	0.408	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.357	0.041	0.036	0.028	0.027	1.040	0.005	0.054	0.038	0.036	0.027	0.032	0.025
0.7	0.038	0.409	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.358	0.041	0.036	0.028	0.027	1.040	0.005	0.054	0.047	0.036	0.027	0.031	0.025
0.9	0.038	0.409	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.358	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.040	0.036	0.027	0.032	0.025
1	0.038	0.410	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.357	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.046	0.036	0.027	0.032	0.024
2	0.038	0.410	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.356	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.050	0.036	0.027	0.032	0.024
3	0.038	0.409	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.355	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.057	0.036	0.027	0.032	0.024
4	0.038	0.408	0.029	0.029	−0.003	0.002	1.000	0.002	0.062	0.353	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.057	0.036	0.027	0.032	0.024
5	0.038	0.408	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.351	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.057	0.036	0.027	0.032	0.024
6	0.038	0.409	0.029	0.029	−0.004	0.002	1.000	0.002	0.062	0.348	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.052	0.036	0.027	0.032	0.024
7	0.038	0.408	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.345	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.044	0.036	0.027	0.032	0.024
8	0.038	0.408	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.343	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.044	0.036	0.027	0.032	0.024
9	0.038	0.407	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.340	0.041	0.037	0.028	0.026	1.040	0.005	0.054	0.045	0.036	0.027	0.032	0.023
10	0.038	0.403	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.339	0.041	0.037	0.028	0.026	1.040	0.005	0.054	0.050	0.036	0.027	0.032	0.023
11	0.038	0.405	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.338	0.041	0.037	0.028	0.026	1.040	0.005	0.054	0.050	0.036	0.027	0.031	0.023
12	0.038	0.405	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.344	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.047	0.036	0.027	0.032	0.023
13	0.038	0.401	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.348	0.041	0.036	0.028	0.026	1.040	0.005	0.054	0.060	0.036	0.027	0.032	0.024
14	0.038	0.402	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.349	0.041	0.036	0.028	0.025	1.040	0.005	0.054	0.082	0.036	0.027	0.032	0.023
15	0.038	0.407	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.349	0.041	0.036	0.028	0.025	1.040	0.005	0.054	0.084	0.036	0.027	0.032	0.023
16	0.038	0.400	0.029	0.029	−0.004	0.002	1.000	0.002	0.062	0.349	0.041	0.036	0.028	0.026	1.040	0.005	0.053	0.108	0.036	0.027	0.031	0.023
17	0.038	0.395	0.029	0.029	−0.004	0.003	1.000	0.002	0.062	0.351	0.041	0.036	0.028	0.026	1.040	0.005	0.053	0.122	0.036	0.027	0.031	0.024
18	0.039	0.374	0.029	0.030	−0.003	0.002	1.000	0.002	0.062	0.354	0.041	0.036	0.028	0.026	1.040	0.005	0.052	0.172	0.035	0.026	0.031	0.023
19	0.039	0.370	0.030	0.030	−0.003	0.003	1.000	0.002	0.061	0.381	0.040	0.036	0.028	0.026	1.040	0.005	0.051	0.245	0.036	0.026	0.032	0.024

[a] The quality measures are the same as in Table 3.

Table 5. Comparison of S^2_{MD} and S^2_{NMR} for Gal3-apo, Gal3-Lac, and the Difference between the Two Proteins[a]

	Gal3-apo								Gal3-Lac								Difference					
set	RMSD	r^2	MAD	MADtr	MSD	Med	MQ	Q	RMSD	r^2	MAD	MADtr	MSD	Med	MQ	Q	RMSD	r^2	MAD	MADtr	MSD	Med
													20 ns									
1A×20	0.041	**0.23**	0.031	**0.031**	−0.003	**0.003**	1.00	0.002	0.066	0.28	0.043	0.038	**0.030**	0.027	**1.04**	0.006	**0.062**	0.02	0.040	0.032	**0.033**	0.027
10A×2	0.041	0.18	0.031	0.032	−0.003	0.004	1.00	0.002	0.065	0.30	0.043	0.037	0.031	**0.026**	**1.04**	0.006	0.061	**0.08**	**0.038**	0.029	0.034	0.027
10M×2	**0.039**	0.20	**0.030**	**0.031**	**−0.002**	0.004	1.00	0.002	**0.063**	0.36	0.042	0.035	0.031	0.028	1.04	0.006	0.057	0.02	**0.038**	0.028	0.033	0.026
													40 ns									
1A×40	0.039	0.22	**0.030**	**0.030**	−0.003	**0.001**	1.00	0.002	0.065	0.29	0.042	0.037	**0.029**	0.026	1.04	0.006	0.059	0.03	0.038	0.030	**0.032**	**0.023**
10A×4	0.040	0.21	0.031	0.031	−0.003	0.002	1.00	0.002	0.064	0.33	0.042	0.036	0.030	**0.025**	1.04	0.006	0.061	**0.11**	**0.037**	0.029	0.033	0.025
32M×1	0.039	0.19	**0.030**	0.031	−0.003	**0.001**	1.00	0.002	0.063	0.36	**0.041**	0.035	0.031	0.027	1.04	**0.005**	0.058	0.00	0.038	**0.028**	0.033	0.025
10M×4	**0.038**	**0.23**	**0.030**	**0.030**	−0.003	0.003	1.00	0.002	**0.062**	**0.37**	**0.041**	**0.035**	0.030	0.027	1.04	0.005	**0.057**	0.02	**0.037**	**0.028**	0.033	0.026
													50 ns									
10A×5	0.039	0.38	0.030	0.031	−0.003	**0.002**	1.00	0.002	0.064	0.33	0.042	0.036	0.030	**0.025**	1.04	0.006	0.061	**0.16**	0.037	0.029	0.033	**0.023**
5A×10	0.039	0.37	0.030	**0.030**	−0.003	0.004	1.00	0.002	0.063	0.34	**0.041**	0.036	0.030	**0.025**	1.04	0.006	0.060	0.08	0.037	0.029	**0.032**	0.025
10M×5	**0.038**	0.40	**0.029**	**0.030**	−0.003	0.003	1.00	0.002	**0.062**	0.37	**0.041**	0.035	0.030	0.026	1.04	**0.005**	0.056	0.02	0.037	**0.027**	0.033	0.026
5M×10	**0.038**	**0.40**	**0.029**	**0.030**	−0.003	**0.002**	1.00	0.002	**0.062**	0.37	0.041	0.036	**0.029**	0.026	1.04	**0.005**	**0.055**	0.05	0.037	0.028	**0.032**	0.025
													100 ns									
10A×10	0.039	0.38	0.030	0.030	**−0.003**	0.003	1.00	0.002	0.063	0.34	0.041	**0.036**	0.029	**0.025**	1.04	**0.005**	0.059	**0.10**	**0.036**	0.029	0.032	**0.023**
5A×20	0.039	0.38	0.030	0.030	−0.004	**0.001**	1.00	0.002	0.063	0.33	0.041	0.037	0.029	**0.025**	1.04	0.006	0.058	0.03	0.037	0.028	0.032	0.025
10M×10	**0.038**	**0.41**	**0.029**	**0.029**	**−0.003**	0.002	1.00	0.002	**0.061**	**0.37**	0.041	**0.036**	0.029	0.026	1.04	**0.005**	**0.055**	0.04	**0.036**	**0.027**	0.032	0.025
5M×20	**0.038**	0.40	**0.029**	0.030	**−0.003**	0.002	1.00	0.002	0.062	0.35	0.041	**0.036**	**0.028**	0.026	1.04	**0.005**	**0.055**	0.04	**0.036**	0.028	0.032	0.025
													160 ns									
32M×5	0.038	0.40	0.029	0.029	−0.003	0.003	1.00	0.002	0.062	0.37	0.041	**0.035**	0.030	**0.025**	1.04	0.005	0.056	0.00	0.037	0.027	0.032	0.025
16M×10	0.038	**0.41**	0.029	0.029	−0.003	**0.002**	1.00	0.002	**0.061**	**0.37**	0.041	0.036	**0.029**	0.026	1.04	0.005	**0.055**	**0.03**	**0.036**	0.027	0.032	0.025
													320 ns									
32M×10	**0.037**	**0.41**	0.029	0.029	−0.003	0.003	1.00	0.002	**0.061**	**0.37**	0.041	0.036	0.029	0.026	1.04	0.005	0.055	0.03	0.036	0.027	0.032	0.025
16M×20	0.038	0.41	0.029	0.029	−0.003	**0.002**	1.00	0.002	0.062	0.36	0.041	0.036	**0.028**	0.026	1.04	0.005	**0.054**	**0.04**	0.036	0.027	0.032	0.025

[a] The quality measures are the same as in Table 3. The best results within each group are marked in bold face. The equilibration time was always 0.25 ns.

simulation time of 20 ns. In the first, we take a single 20 ns simulation, started with the all-A conformation (1A×20). In the second, we instead take 10 independent simulations of 2 ns, all started from the A conformations (10A×2). In the third set, we take 10 simulations of 2 ns, started from different conformations (10M×2). These 10 simulations can be selected in many ways from the 32 simulations we have run with different starting conformations. We simply selected 10 simulations out of these 32 at random and repeated this 50 times to obtain a stable average. From the results in Table 5, it can be seen that the third set (10M×2) gives slightly better results than the other two sets: It gives the best result for 16 of the 22 quality criteria examined (we did not consider here the number of residues for which the S^2_{NMR} value falls outside the range of the S^2_{MD} values because our previous results indicated that it is a poor quality measure). The probability that we would get such a result if the distribution was completely random is less than 3%. The other two sets were best only for seven or nine quality measures. Thus, it is better to run 20 short simulations than one long one, and it is also better to start from several different conformations than a single one.

Likewise, we constructed four sets of simulations with a total length of 40 ns. The first is a 40-ns simulation started from a single conformation (1A×40). The second is 10 independent simulations of 4 ns, all started from the same conformation (10A×4). The third is 32 simulations of 1.25 ns, started from different conformations, (32M×1), whereas the fourth is 10 simulations of 4 ns length, started from different conformations (10M×4; again an average over 50 different random selections of 10 simulations out of the available 32 different simulations). The results in Table 5 indicate that there is a slight advantage to start with different conformations: The 10M×4 simulations gave the best results for 16 quality measures (89% significance), whereas the second best methods 1A×40 and 32M×1.25 are best for 10 quality measures. The last method, 10A×4 is best for seven quality measures. This also indicates that 1.25 ns is a too short a time for the simulation of order parameters—the 4 ns simulations give better results, even if fewer simulations are run. This result is expected, because the experimentally determined correlation time for the rotational diffusion is 7−8 ns for both proteins.[38]

Comparing the results with 20 or 40 ns total simulation time, there is a clear improvement when using the longer simulation time for 12 of the quality criteria, and only one becomes worse (>99% significance). Likewise, there is a clear improvement in the results going from the best set of simulations with a total time of 40 ns and the 32M×20 simulations in Table 3: 11 quality criteria are improved, especially for the difference between the two proteins, whereas only three become worse (97% significance). This shows that the order parameters can be improved by extending the simulations, although the convergence is very slow.

This observation led us to continue the investigation with simulations of a total length of 50, 100, 160, and 320 ns. The results are also included in Table 5. It can be seen that, for a total simulation time of 50 ns, it is better to run five 10 ns simulations than 10 5-ns simulations (95% significance), irrespective of whether they are started from a single or many different conformations, although the latter gives the best results (89% significance). On the other hand, for a total simulation time of 100 ns, it is better to run 10 10-ns simulations than five 20 ns simulations (89−94% significance). Again, simulations started from several conformations give the best results (95% significance). These results are confirmed for the even longer total simulation time: It is better to run 16 simulations of a length of 10 ns than 32 simulations of 5 ns length (95% significance). Only for the longest simulation time (320 ns) do the results become inconclusive—there is no significant difference between 32 simulations of 10 ns length or 16 simulations of 20 ns length. However, the conclusion remains that there is no advantage of running the longer simulations. Therefore, we can with good confidence conclude that the optimum simulation length, at least for Gal3, is ∼10 ns.

It can also be seen from Table 5 that we reach convergence for the various quality measures. Between 50 and 100 ns total simulation time, there is an improvement for 10 of the quality measures, whereas only one is worse (for the two best methods; 98% significance). However, going from 100 to 160 or 320 ns total simulation time, there is no longer any clear improvement. In fact, by comparing the results in Tables 3 and 5, it can be seen that the 10M×10 simulations actually give better results than the full 32M×20 simulations for six of the quality measures and only two of them give worse results. This also confirms that we can use a short equilibration time of 0.25 ns. Therefore, we conclude that, for Gal3, the ideal simulation protocol involves 10 simulations of 10 ns, starting from different conformations.

There is also a difference between the various types of simulations with regard to the stability of the calculated S^2_{MD}, as estimated from the difference in results obtained with the four methods to estimate S^2_{MD}. From Table 1, it can be seen that, with a single 20 ns simulation, 24 residues have a range larger than 0.05 between the various methods, and most of them are only observed for one protein. This number is decreased to 10 if the simulation is extended to 40 ns, and a similar number is observed also for the 10 independent simulations started from the same structure, irrespective of whether they are 20 or 40 ns long. However, for the 32 simulations started from different conformations, only 6−7 residues have poorly determined S^2_{MD}'s. Thus, starting from several different conformations makes the results more stable and well-determined. However, this also depends on the length of the simulations. If we instead use only 10M×10 simulations (with 0.25 ns equilibration), there are 15 residues with poorly determined S^2_{MD}'s, although this set can be reduced to something similar to the 32M×20 set by using a threshold of ∼0.09 instead.

For seven residues, the range of S^2_{MD} obtained with different methods is large in all simulation sets, viz., Ile115, Val116, Gly125, Val155, Leu177, Leu228 (not in Gal3-Lac), and Ser232 (cf. Figure 1). Only two of these, Val116 and Ser232, have different conformations in the crystal structures, but they both reside mainly in one conformation (S1) during the MD simulations (74−91%).

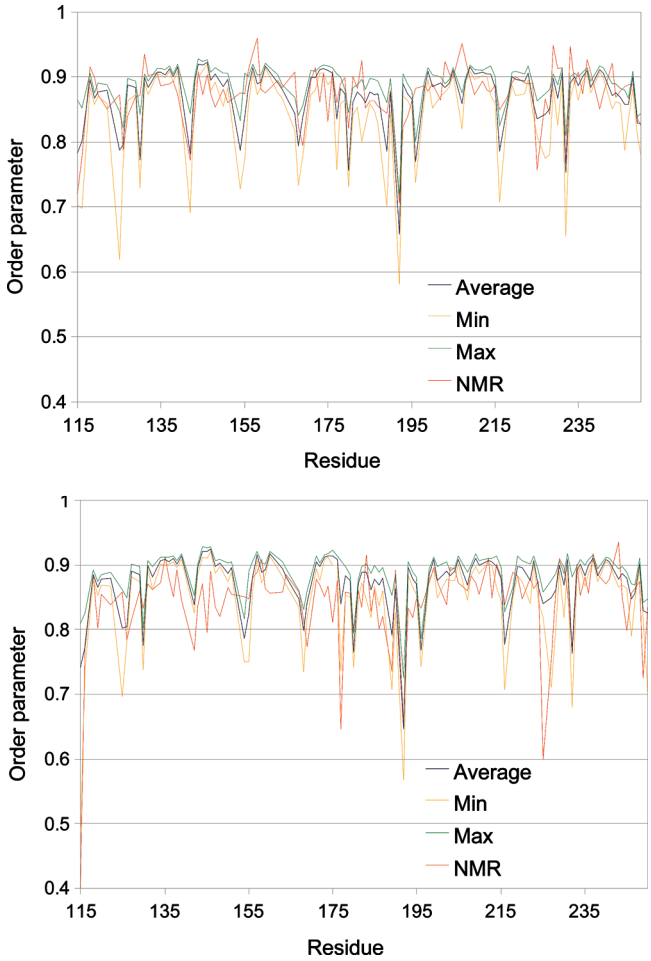

Figure 4. S^2_{MD} parameters from the 32M×20 ns simulations (average, minimum, and maximum over the 32 simulations), compared to the corresponding S^2_{NMR} parameters for (top) Gal3-apo and (bottom) Gal3-Lac. Only residues with S^2_{NMR} parameters in both states are shown.

In Figure 4, we compare the calculated and measured order parameters for Gal3-apo and Gal3-Lac. The maximum value of each order parameter over the 32 simulations is always close to the average, whereas the minimum shows a rather large variation for some of the residues. The residues that have the largest absolute difference between S^2_{MD} and S^2_{NMR} are Lys196, Met130, Asp207, Asp154, and Gly125 for Gal3-apo and Ile115, Val225, Leu177, Ile145, and Met249 for Gal3-Lac. As can be seen in Figure 5, large differences are primarily observed in loops in the protein structure, whereas the β sheets are well described. Three of the residues with large deviations, Ile115, Gly125, and Leu177, have poorly determined S^2_{MD}'s, whereas the other residues with poorly determined S^2_{MD}'s do not show any conspicuous errors. Interestingly, residues with large errors have a too high S^2_{MD} for Gal3-Lac, but a too low S^2_{MD} for Gal3-apo. The errors are also larger for Gal3-Lac (up to 0.35) than for Gal3-apo (up to 0.10).

If the residues with poorly determined S^2_{MD}'s are removed from the analysis, the results are significantly improved for Gal3-Lac and for $\Delta\Delta S^2$, as can be seen in Table 6 (97% significance). For example, the MADs decrease from 0.041 and 0.036 to 0.037 and 0.032. On the other hand, correlation coefficients become worse, simply because the poorly

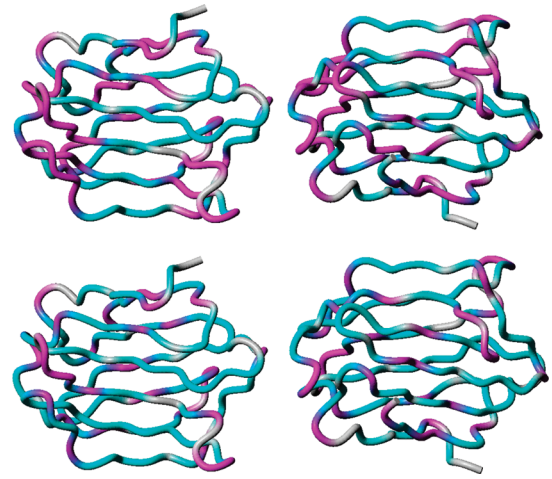

Figure 5. Mapping of the $S^2_{MD} - S^2_{NMR}$ difference onto the crystal structure of Gal3. Two pictures are shown of each Gal3-apo (top) and Gal3-Lac (bottom), related by a 180° rotation. The scale runs from dark blue ($S^2_{MD} < S^2_{NMR}$) via cyan ($S^2_{MD} = S^2_{NMR}$) to magenta. Gray color indicates that data are missing.

determined residues typically have low S^2's, contributing strongly to the correlation coefficient. Further improvement is seen if residues in loops surrounding the saccharide-binding site are omitted or if residues that have been fitted with a R_{ex} term in the NMR experiments are omitted, as is also seen in Table 6 (for example, r^2 increases to 0.54−0.70), but then, the number of considered residues becomes rather small (41). We have also checked whether residues that have two conformations in the crystal structure, or are located close to such residues, give worse results when S^2_{MD} is compared to S^2_{NMR}. However, we did not find any such trends.

Finally, we have also included in the table the results of the simple contact model, suggested by Zhang and Brüschweiler.[61] It can be seen that it gives worse results for all quality measures, except the RMSD, MSD, MQ, and median for Gal3-Lac and r^2 for $\Delta\Delta S^2$ (significance 97%). Thus, the simulations provide significantly improved predictions of the order parameters compared to the contact model.

Conclusions

In this paper, we have addressed a number of questions of importance in the calculation of backbone N−H order parameters from MD simulations. First, we have compared four different methods to extract the S^2_{MD} parameters, viz., ACF and iRED with three different window sizes. Different quality measures give different results, as do different simulated systems, so we cannot reach any definite conclusions. ACF seems to give the results with the largest spread; i.e., it seems to be more sensitive to the convergence of the simulations than the iRED approach. The iRED method with windows of 1 ns seems to give the best precision and the smallest outliers on average, but the median and correlation coefficient were sometimes worse than for other variants of iRED. However, the most important result was that the four methods gave similar results for most of the studied S^2_{MD} parameters, indicating that all methods give reliable results. In fact, if the four methods differ significantly (e.g., by more than 0.05), it indicates that

Table 6. Comparison of S^2_{MD} and S^2_{NMR} for Gal3-apo, Gal3-Lac, and the Difference between the Two Proteins When Various Residues Are Omitted from the Comparison[a]

simulation	*n*	RMSD	r^2	MAD	MADtr	MSD	median	MQ	*Q*
				Gal3-apo					
32M×20	109[b]	0.038	0.28	0.029	0.029	−0.004	0.003	1.00	0.002
	104[c]	0.038	0.34	0.029	0.029	−0.004	0.003	1.00	0.002
	82[d]	0.029	0.60	0.023	0.023	−0.001	0.004	1.00	0.001
	41[b]	0.030	0.68	0.023	0.023	−0.007	−0.004	0.99	0.001
10M×10	109[b]	0.038	0.41	0.029	0.029	−0.003	0.002	1.00	0.002
	104[c]	0.037	0.34	0.029	0.029	−0.003	0.002	1.00	0.002
	82[d]	0.029	0.59	0.023	0.023	0.000	0.004	1.00	0.001
	41[e]	0.029	0.70	0.023	0.023	−0.006	−0.003	0.99	0.001
contact model	109[b]	0.073	0.17	0.049	0.046	−0.034	−0.018	0.96	0.007
				Gal3-Lac					
32M×20	109[b]	0.062	0.35	0.041	0.036	0.028	0.026	1.04	0.005
	104[c]	0.050	0.31	0.037	0.032	0.025	0.026	1.03	0.003
	82[d]	0.052	0.48	0.031	0.028	0.023	0.022	1.03	0.004
	41[e]	0.066	0.50	0.035	0.034	0.025	0.022	1.04	0.006
10M×10	109[b]	0.061	0.37	0.041	0.036	0.029	0.026	1.04	0.005
	104[c]	0.050	0.31	0.037	0.032	0.025	0.026	1.03	0.003
	82[d]	0.052	0.49	0.032	0.027	0.024	0.022	1.03	0.004
	41[e]	0.065	0.54	0.035	0.033	0.027	0.022	1.04	0.006
contact model	109[b]	0.058	0.31	0.041	0.040	0.003	0.005	1.01	0.005
				difference					
32M×20	109[b]	0.054	0.05	0.036	0.027	0.032	0.025		
	104[c]	0.042	0.00	0.032	0.022	0.028	0.025		
	82[d]	0.044	0.22	0.029	0.021	0.024	0.019		
	41[e]	0.057	0.23	0.035	0.025	0.032	0.023		
10M×10	109[b]	0.055	0.04	0.036	0.027	0.032	0.025		
	104[c]	0.043	0.01	0.033	0.022	0.029	0.025		
	82[d]	0.045	0.17	0.029	0.021	0.024	0.021		
	41[e]	0.058	0.18	0.036	0.025	0.033	0.025		
contact model	109[b]	0.075	0.09	0.042	0.034	0.037	0.025		

[a] *n* is the number of residues included in the comparison. The quality measures are the same as in Table 3. [b] All residues are included. [c] Five poorly determined S^2_{MD}'s according to the 32M×20 simulations were omitted (cf. Table 1). [d] Residues in loops surrounding the saccharide-binding site are omitted. [e] Residues that have been fitted with a R_{ex} term are omitted.

there are problems with the convergence of the calculations; consequently, we suggest that it is good practice to exclude the affected residues from detailed interpretation.

Second, we have studied how the calculated S^2_{MD} parameters depend on the starting conditions of the MD simulations. It is clearly inappropriate to base the calculations on a single MD simulation. Better results are obtained if the results of several independent simulations are averaged. They can be obtained by simply using different starting velocities, but it is advantageous to use several different conformations, if present in the crystal structure.

Third, we have compared different lengths of the simulations. Our calculations show that, at least for Gal3, the results are better if the simulation length is increased from 5 to 10 ns, but there is no significant improvement if they are extended to 20 or 40 ns (keeping the total simulation time constant by running several independent simulations). Moreover, there is no significant improvement when extending the total simulation time over 100 ns, except that a few order parameters become better determined. Thus, our results indicate that the ideal simulation protocol is 10 independent simulations of 10 ns length, started from different conformations.

Fourth, even if the RMSD of the coordinates indicates that an equilibration time of 5 ns is needed to reach stable results, this has a small influence on the calculated S^2_{MD} parameters, at least when averaged over several independent simulations. In fact, 10 × 10 ns simulations with an equilibration time of only 0.25 ns give as good results as 32 × 20 ns simulations with 5 ns equilibration.

Fifth, it should be noted that, even after 400−640 ns simulation time, the correspondence between calculated and measured S^2 parameters is rather poor, with a correlation coefficient of less than 0.43, a MAD of over 0.029, and with a maximum error of up to 0.35.

Finally, although this study has concentrated on a comparison of calculated and measured S^2 order parameters, we are confident that most of our conclusions are applicable also to calculations of other properties from MD simulations, as other investigations indicate.[25]

Supporting Information Available: Description of the selection of alternative conformations; description of the alternative conformations and their distinct hydrogen-bond patterns; comparison of the four methods to obtain S^2_{MD}; description of the dihedral conformations observed in the various simulations; transition times between the various dihedral conformations; and comparison of the various simulations in Table 3, taking into account the statistical uncertainties in both S^2_{MD} and S^2_{NMR}. This information is available free of charge via the Internet at http://pubs.acs.org/.

Acknowledgment. This investigation has been supported by grants from the Swedish research council and from the Research School in Pharmaceutical Science. It has also

been supported by computer resources of Lunarc at Lund University and at HPC2N at Umeå University.

References

(1) Akke, M.; Brüschweiler, R.; Palmer, A. G. NMR order parameters and free energy: An analytical approach and its application to cooperative Ca^{2+} binding by calbindin D_{9k}. *J. Am. Chem. Soc.* **2003**, *115*, 9832–9833.

(2) Homans, S. W. Probing the binding entropy of ligand-protein interactions by NMR. *ChemBioChem* **2005**, *6*, 1585–1591.

(3) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem. Rev.* **2006**, *106*, 1672–1699.

(4) 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.

(5) 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 15N nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* **1990**, *112*, 4989–4991.

(6) Halle, B. The physical basis of model-free analysis of NMR relaxation data from proteins and complex fluids. *J. Chem. Phys.* **2009**, *131*, 224507.

(7) Palmer, A. G. NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 129–155.

(8) Boyd, J. Measurement of 15N Relaxation Data from the Side Chains of Asparagine and Glutamine Residues in Proteins. *J. Magn. Reson. B* **1995**, *107*, 279–285.

(9) Berglund, H.; Baumann, H.; Knapp, S.; Ladenstein, R.; Härd, T. Flexibility of an Arginine Side Chain at a DNA-Protein Interface. *J. Am. Chem. Soc.* **1995**, *117*, 12883–12884.

(10) Muhandiram, D. R.; Yamazaki, T.; Sykes, B. D.; Kay, L. E. Measurement of 2H T1 and T1.rho. Relaxation Times in Uniformly 13C-Labeled and Fractionally 2H-Labeled Proteins in Solution. *J. Am. Chem. Soc.* **1995**, *117*, 11536–11544.

(11) Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 1. Measurement of Five Relaxation Rates per Deuteron in 13C-Labeled and Fractionally 2H-Enriched Proteins in Solution. *J. Am. Chem. Soc.* **2002**, *124*, 6439–6448.

(12) Li, D.-W.; Brüschweiler, R. A Dictionary for Protein Side-Chain Entropies from NMR Order Parameters. *J. Am. Chem. Soc.* **2009**, *131*, 7226–7227.

(13) Case, D. A. Molecular Dynamics and NMR Spin Relaxation in Proteins. *Acc. Chem. Res.* **2002**, *35*, 325–331.

(14) Brüschweiler, R. New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Curr. Opin. Struct. Biol.* **2003**, *13*, 175–183.

(15) Philippopoulos, M.; Mandel, A. M.; Palmer, A. G.; Lim, C. Structure and dynamics of the M13 coat signal sequence in membranes by multidimensional high-resolution and solid-state NMR spectroscopy. *Proteins: Struct., Funct., Gen.* **1997**, *27*, 481–493.

(16) Showalter, S. A.; Brüschweiler, 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.

(17) 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: Struct., Funct. Bioinform.* **2006**, *65*, 712–725.

(18) Buck, M. S.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. Biophys. *J. Biophys. Lett.* **2006**, *90*, L36–L38.

(19) Soares, T.; Daura, X.; Oostenbrink, C.; Smith, L.; van Gunsteren, W. F. Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme. *J. Biomol. NMR* **2004**, *30*, 407–422.

(20) Lipari, G. A. Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.

(21) Horita, D. A.; Zhang, W.; Smithgall, T. E.; Gmeiner, W. H.; Byrd, R. A. Dynamics of the Hck-SH3 domain: comparison of experiment with multiple molecular dynamics simulations. *Protein Sci.* **2000**, *9*, 95–103.

(22) Andrews, B. K.; Romo, T.; Clarage, J. B.; Pettitt, B. M.; Phillips, G. N. Characterizing global substates of myoglobin. *Structure.* **1998**, *6*, 587–594.

(23) Koller, A. N.; Schwalbe, H.; Gohlke, H. Starting Structure Dependence of NMR Order Parameters Derived from MD Simulations: Implications for Judging Force-Field Quality. Biophys. *J. Biophys. Lett.* **2008**, *95*, L04–L06.

(24) Smith, P. E.; Petitt, B. M.; Karplus, M. Stochastic dynamics simulations of the alanine dipeptide using a solvent-modified potential energy surface. *J. Phys. Chem.* **1999**, *97*, 6907–6913.

(25) Genheden, S.; Ryde, U. A Comparison of Different Initialization Protocols to Obtain Statistically Independent Molecular Dynamics Simulations. *J. Comput. Chem.* In press. DOI: 10.1002/jcc.21546.

(26) Lawrenz, M.; Baron, P.; McCammon, J. A. Independent-Trajectories Thermodynamic-Integration Free-Energy Changes for Biomolecular Systems: Determinants of H5N1 Avian Influenza Virus Neuraminidase Inhibition by Peramivir. *J. Chem. Theory Comput.* **2009**, *5*, 1106–1116.

(27) Collins, P. M.; Hidari, K. I. P. J.; Blanchard, H. Slow diffusion of lactose out of galectin-3 crystals monitored by X-ray crystallography: possible implications for ligand-exchange protocols. *Acta Crystallogr., Sect. D* **2007**, *63*, 415–419.

(28) Houlzelstein, D.; Goncalves, I. R.; Fadden, A. J.; Sidhu, S. S.; Cooper, D. N.; Drickamer, K.; Leffler, H.; Poirer, F. Phylogenetic Analysis of the Vertebrate Galectin Family. *Mol. Biol. Evol.* **2004**, *21*, 1177–1187.

(29) Leffler, H.; Carlsson, S.; Hedlund, M.; Quian, Y. Introduction to galectins. *Glycoconjugate J.* **2002**, *19*, 433–440.

(30) Liu, F.-T.; Rabinovich, G. A. Galectins as modulators of tumour progression. *Nat. Rev. Cancer* **2005**, *5*, 29–41.

(31) Nakahara, S.; Oka, N.; Raz, A. On the role of galectin-3 in cancer apoptosis. *Apoptosis* **2005**, *10*, 267–275.

(32) Liu, F.-T. Regulatory Roles of Galectins in the Immune Response. *Int. Arch. Allergy Immunol.* **2005**, *136*, 385–400.

(33) Ilarrgui, J. M.; Bianco, G. A.; Toscano, M. A.; Rabinovich, G. A. New targets III: The coming of age of galectins as immunomodulatory agents: impact of these carbohydrate binding proteins in T cell physiology and chronic inflammatory disorders. *Ann Rheum. Dis.* **2005**, *64*, 96–103.

(34) Patterson, R. J.; Wang, W.; Wang, J. L. Understanding the biochemical activities of galectin-1 and galectin-3 in the nucleus. *Glycoconj. J.* **2002**, *19*, 499–506.

(35) Dumic, J.; Dabelic, S.; Flögel, M. Galectin-3: An open-ended story. *Biochim. Biophys. Acta* **2006**, *1760*, 616–635.

(36) Bachhawat-Sikder, K.; Thomas, C. J.; Suriola, A. Thermodynamic analysis of the binding of galactose and poly-N-acetyllactoseamine derivatives to human galectin-3. *FEBS Lett.* **2001**, *500*, 75–79.

(37) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California: San Francisco, 2008.

(38) Diehl, C.; Genheden, S.; Modig, K.; Ryde, U.; Akke, M. Conformational entropy changes upon lactose binding to the carbohydrate recognition domain of galectin-3. *J. Biomol. NMR* **2009**, *45*, 157–169.

(39) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(40) 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.

(41) Wu, X.; Brooks, B. R. Self-guided Langevin dynamics simulation method. *Chem. Phys. Lett.* **2003**, *381*, 512–518.

(42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(43) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(44) Buck, M.; Karplus, M. Internal and overall peptide group motion in proteins: molecular dynamics simulations for lysozyme compared with results from X-ray and NMR spectroscopy. *J. Am. Chem. Soc.* **1999**, *121*, 9645–9658.

(45) Zwansig, R.; Ailawadi, N. K. Statistical Error Due to Finite Time Averaging in Computer Experiments. *Phys. Rev.* **1969**, *1982*, 280–282.

(46) Lu, C.-Y.; Bout, D. A. V. Effect of finite trajectory length on the correlation function analysis of single molecule data. *J. Chem. Phys.* **2006**, *125*, 124701–124709.

(47) Madsen, H. *Time Series Analysis*; Chapman & Hall/CRC: New York, 2008.

(48) Efron, B.; Tibshirani, R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statist. Sci.* **1986**, *1*, 54–77.

(49) Prompers, J. J.; Brüschweiler, R. General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation. *J. Am. Chem. Soc.* **2002**, *124*, 4522–4534.

(50) Nederveen, A. J.; Bonvin, A. NMR relaxation and internal dynamics of ubiquitin from a 0.2 microsec MD simulation. *J. Chem. Theory Comput.* **2005**, *1*, 363–374.

(51) Bowers, K.; Devolder, B.; Yin, L.; Kwan, T. A maximum likelihood method for linking particle-in-cell and Monte-Carlo transport simulations. *Comput. Phys. Commun.* **2004**, *164*, 311–317.

(52) Dempster, A. P.; Laird, N. M.; Rubin, M. D. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Stat. Soc. B* **1977**, *39*, 1–38.

(53) Meiler, J.; Promper, J. J.; Peti, W.; Griesinger, C.; Bruschweiler, R. Model-Free Approach to the Dynamic Interpretation of Residual Dipolar Couplings in Globular Proteins. *J. Am. Chem. Soc.* **2001**, *123*, 6098–6107.

(54) Umemoto, K.; Leffler, H.; Venot, A.; Valafar, H.; Prestegard, J. H. Conformational differences in liganded and unliganded states of Galectin-3. *Biochemistry* **2003**, *42*, 3688–3695.

(55) Pfeiffer, S.; Fushman, D.; Cowburn, D. Simulated and NMR-Derived Backbone Dynamics of a Protein with Significant Flexibility: A Comparison of Spectral Densities for the βARK1 PH Domain. *J. Am. Chem. Soc.* **2001**, *112*, 3021–3036.

(56) Kanibolotsky, D. S.; Ivanova, O. S.; Lisnyak, V. V. Comparison of NMR and MD NZH bond order parameters: example of HIV-1 protease. *Mol. Sim.* **2006**, *32*, 1155–1163.

(57) MacRaild, C. A.; Daranas, A. H.; Bronowska, A.; Homans, S. W. Global Changes in Local Protein Dynamics Reduce the Entropic Cost of Carbohydrate Binding in the Arabinose-binding Protein. *J. Mol. Biol.* **2007**, *368*, 822–832.

(58) Markwick, P. R. L.; Bouvignies, G.; Blackledge, M. S. Exploring Multiple Timescale Motions in Protein GB3 Using Accelerated Molecular Dynamics and NMR Spectroscopy. *J. Am. Chem. Soc.* **2007**, *129*, 4724–4730.

(59) Maragkis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. O.; Xu, H.; Trbovis, N.; Friesner, R. A.; Plamer III, A. G.; Shaw, D. E. Microsecond Molecular Dynamics Simulation Shows Effect of Slow Loop Dynamics on Backbone Amide Order Parameters of Proteins. *J. Phys. Chem. B* **2008**, *112*, 6155–6158.

(60) Tong, Y.; Ji, C. G.; Mei, Y.; Zhang, J. Z. H. Simulation of NMR Data Reveals That Proteins' Local Structures Are Stabilized by Electronic Polarization. *J. Am. Chem. Soc.* **2009**, *131*, 8636–8641.

(61) Zhang, F.; Brüschweiler, R. Contact Model for the Prediction of NMR N-H Order Parameters in Globular Proteins. *J. Am. Chem. Soc.* **2002**, *124*, 12654–12655.

CT900696Z