JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Iterative Optimization of Molecular Mechanics Force Fields from NMR Data of Full-Length Proteins

Da-Wei Li and Rafael Brüschweiler*

Chemical Sciences Laboratory, Department of Chemistry and Biochemistry and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306, United States

ABSTRACT: High quality molecular mechanics force fields of proteins are key for the quantitative interpretation of experimental data and the predictive understanding of protein function based on computer simulations. A strategy is presented for the optimization of protein force fields based on full-length proteins in their native environment that is guided by experimental NMR chemical shifts and residual dipolar couplings (RDCs). An energy-based reweighting approach is applied to a long molecular dynamics trajectory, performed with a parent force field, to efficiently screen a large number of trial force fields. The force field that yields the best agreement with the experimental data is then used as the new parent force field, and the procedure is repeated until no further improvement is obtained. This method is demonstrated for the optimization of the backbone φ,ψ dihedral angle potential of the Amber ff99SB force field using six trial proteins and another 17 proteins for cross-validation using ^{13}C chemical shifts with and without backbone RDCs. The φ,ψ dihedral angle potential is systematically improved by the inclusion of correlation effects through the addition of up to 24 bivariate Gaussian functions of variable height, width, and tilt angle. The resulting force fields, termed ff99SB_$\varphi\psi$(g24;CS) and ff99SB_$\varphi\psi$(g8;CS,RDC), perform significantly better than their parent force field in terms of both NMR data reproduction and Cartesian coordinate root-mean-square deviations between the MD trajectories and the X-ray crystal structures. The strategy introduced here represents a powerful addition to force field optimization approaches by overcoming shortcomings of methods that are solely based on quantum-chemical calculations of small molecules and protein fragments in the gas phase.

INTRODUCTION

Molecular dynamics (MD) simulations of biomolecules play an important complementary role with respect to experiments. In principle, computer simulations can provide a complete time-resolved single-molecule picture of a protein's behavior in atomic detail. On the other hand, experimental studies are invariably incomplete since certain properties are hard to measure. In recent years, MD has been increasing in popularity in the study of protein behavior enabled by the continuing exponential increase in computational power[1] that allows the simulation of both larger systems and longer time scales and due to progress in the quality of the underlying molecular mechanics force fields.[2,3] Indeed, critical assessment of protein force fields has greatly benefited from the ability to sample conformational space of proteins more thoroughly.[4] Still, there remains much room for further improvement to enable increasingly accurate *in silico* studies of protein properties and function. Currently, MD simulations into the hundreds of nanoseconds regime are becoming routine, which permit a fully quantitative comparison between computation and certain types of experiments, such as heteronuclear NMR spin relaxation.[4,5] Other NMR parameters, such as chemical shifts, scalar J couplings, and residual dipolar couplings (RDCs), reflect a wider range of time scales from picoseconds to milliseconds, covering motional regimes on which many biologically important events occur. In order to validate force fields on the full range of time scales, the use of longer simulations represents a natural choice.[6−8]

The development of better force fields is a complex task that is both labor-intensive and time-consuming. Essentially, all modern force fields have been parametrized on the basis of both extensive quantum-chemical calculations and experimental data.[2,9−27] The complexity associated with the fitting of a large number (well over 100) of force field parameters is reflected in different parametrization philosophies for different force fields. A common approach is the fitting of force field parameters to gas-phase quantum-chemical calculations and experimental data of small molecules, including amino acid analogs and small peptides, in terms of average conformation, vibrational spectra, solvation free energies, and relative energies of different conformations. The fine-tuning of the CHARMM force field in the final step through MD simulations of protein crystals is an exception.[3]

Not only can NMR data of full-length proteins in their native environment be used to cross-validate MD trajectories, but these data can also serve to directly improve the molecular mechanics force field itself. Using an energy-based reweighting scheme, we recently demonstrated this strategy by improving the backbone dihedral angle potential of the Amber ff99SB[2,28] force field using experimental NMR chemical shift data from a set of proteins.[29] The force field was improved through modification of the coefficients of a fourth order Fourier series expansion in the backbone dihedral angles. Here, we present an alternative strategy that uses a set of bivariate Gaussian functions that are added to the backbone dihedral angle potential to optimize agreement with respect to both protein ^{13}C chemical shifts and backbone RDCs measured in multiple alignment media.

Received: February 9, 2011
Published: May 05, 2011

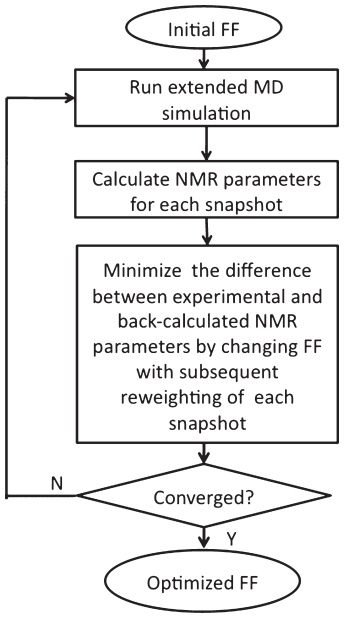

Figure 1. Schematic presentation of the iterative optimization procedure of protein molecular mechanics force fields using experimental data (in the present work, NMR chemical shifts or RDCs) of intact full-length proteins as input.

METHODS

Iterative Reweighting Strategy. The computational strategy used in this work is outlined in Figure 1. It starts with the back-calculation of experimental parameters, such as chemical shifts and RDCs, for each snapshot of the MD trajectory based on the parent force field (V_{old}) of one or several protein systems, which are stored for subsequent analysis. We use here Cα, Cβ, and C′ chemical shifts predicted for each MD snapshot using the program SHIFTS[30,31] as well as RDCs (see below). Time- and ensemble-averaged chemical shifts are calculated with equal weights, $p_{old}(i) = 1/N$, for all N snapshots and compared with the experimental chemical shifts by means of the root-mean-square deviation (RMSD) in units of parts per million. We then reweight a parent trajectory performed with the original force field V_{old} for a new test force field V_{new} by using Boltzmann's relationship:

$$p_{new}(i) = p_{old}(i)\, e^{-V_{new}(i)/kT} / e^{-V_{old}(i)/kT} \quad (1)$$

where $p_{old}(i)$ and $p_{new}(i)$ are the relative weights and $V_{old}(i)$ and $V_{new}(i)$ are the potential energies of snapshot i for the old and new force field, respectively, k is Boltzmann's constant, and T is the simulation temperature that is kept constant. The force field is then optimized iteratively as follows. For each new trial force field V_{new}, the new weights $p_{new}(i)$ are used to compute new NMR parameters. The optimized force field is the one that minimizes the overall discrepancy between the back-calculated and experimental parameters. In previous work,[29] we applied downhill simplex minimization for optimization. In order to search force field space more comprehensively, we use here a Monte Carlo simulated annealing method followed by simplex minimization.

Energy-based reweighting is a well-known tool in biomolecular simulations,[32] whose utility for force field optimization has only been demonstrated recently.[29] Because reweighting does not create new conformations, its effectiveness critically depends on the overlap of the parent trajectory with the reweighted trajectory. We therefore use a collectivity parameter[33]

$$\kappa = \frac{100\%}{N} \exp[-\sum_{i=1}^{N} p_{new}(i) \log p_{new}(i)] \quad (2)$$

where $p_{new}(i)$ are the normalized populations of eq 1. κ = 100% indicates that all snapshots in the parent trajectory contribute equally to the reweighted trajectory, whereas a collectivity parameter near zero means that very few snapshots dominate the reweighted trajectory. The latter situation is undesirable, as it leads to a statistically poor representation of the native ensemble. Throughout this work, we require that κ > 50% to ensure that the parent and the reweighted trajectories significantly overlap. Since this requirement limits the allowed range of changes between the new and the parent force field, the above reweighting procedure is iteratively repeated (Figure 1) until no further improvement is achieved.

Local Reweighting Method. As every protein possesses a large number of degrees of freedom, even a moderately large change of a single force field term can produce a change in the potential energy that amounts to minimal overlap between the parent and the new trajectory (low κ value). To overcome this issue, we employ a local residue-based reweighting scheme. For the evaluation of the ensemble-averaged chemical shift of atom j from the reweighted trajectory, rescaled energy changes $\Delta E_j = \Sigma_{k=1}^{N_{res}} \exp(-r_{kj}/r_0)\Delta E_{kj}$ are used where ΔE_{kj} is the dihedral angle energy difference between the parent and trial force field of residue k with their Cα atoms separated by the distance r_{kj}, N_{res} is the number of residues, and r_0 is set to 9 Å. This assumes that the effect of residue k on the local structure of atom j decreases with increasing distance. MD simulations with the new force field are then carried out to verify its performance.

Chemical Shift RMSD Calculations. The new predicted ensemble averaged chemical shifts $\langle\delta_m\rangle$ are calculated as

$$\langle\delta_m\rangle = \frac{\sum_{i=1}^{N} p_{new}(i)\, \delta_m(i)}{\sum_{i=1}^{N} p_{new}(i)} \quad (3)$$

where $\delta_m(i)$ is the chemical shift of a given nucleus m predicted for snapshot i. The root-mean-square deviation (RMSD) between predicted chemical shifts $\langle\delta_m\rangle$ and experimental chemical shifts $\delta_{m,exp}$ is given by

$$RMSD = \sqrt{\frac{1}{N_{nuclei}} \sum_{m=1}^{N_{nuclei}} (\langle\delta_m\rangle - \delta_{m,exp})^2} \quad (4)$$

in units of parts per million (ppm) where N_{nuclei} is the number of nuclei of each type, Cα, Cβ, or C′. Total RMSDs are obtained by averaging of the RMSDs of the different types of nuclei. All experimental chemical shifts were taken from the BioMagRes Bank[34] with the databank entry codes given in column 1 of Table 2.

Residual Dipolar Coupling Calculations. RDCs represent another type of NMR parameter, which reflects protein structure and dynamics in a way that is complementary to chemical shifts.[35−38] RDCs are observed as cross-peak splittings due to weak alignment of the proteins in an anisotropic environment. The simple geometric dependence of RDCs on protein structure makes them suitable for the rigorous assessment of the quality of

MD trajectories.[39,40] For the same reason, force-field optimization based on RDC information is attractive. However, extension of the residue-based local reweighting method from chemical shifts to RDCs is not straightforward because back-calculation of RDCs involves fitting of an alignment tensor by singular value decomposition,[39,41] which simultaneously involves all RDCs across a protein. In analogy to the local chemical shift reweighting (see above), we assign a relative weight $\exp(-r/r_0)$ to each individual RDC in the global fitting when RDCs of internuclear vectors that belong to one particular residue are back-calculated from the MD trajectory where r is the distance between the Cα atoms of the two residues and r_0 is set to 9 Å. The agreement between back-calculated RDCs, $D_{i,\mathrm{back}}$, and experimental values, $D_{i,\mathrm{exp}}$ (i = 1, ..., N_{RDC}), is then expressed by the Q value:[39]

$$Q = \left[\frac{\sum_{i=1}^{N_{\mathrm{RDC}}} (D_{i,\mathrm{back}} - D_{i,\mathrm{exp}})^2}{\sum_{i=1}^{N_{\mathrm{RDC}}} D_{i,\mathrm{exp}}^2} \right]^{1/2} \quad (5)$$

Q values obtained for different alignments are linearly averaged, resulting in $\langle Q \rangle$. One limitation associated with force field optimization using RDCs only is that high-quality experimental data for proteins with known high-resolution 3D structures are presently available for only a few systems, which poses the risk of overfitting. Therefore, we combine RDCs with chemical shifts where the optimization target for force field improvement is the weighted sum of the average chemical shift RMSD and the average RDC-derived $\langle Q \rangle$, with weights of 1.0 and 2.5 for chemical shifts and RDCs, respectively.

Bivariate Gaussian Potential Dihedral Angle Energy Terms. In the course of force field optimization, backbone dihedral angle terms are often optimized in the final step because they do not directly affect the interaction between amino acids and between amino acids and explicit water. In our previous work,[29] we limited ourselves to the modification of individual backbone dihedral angle energy terms. Here, we explicitly include non-amino acid specific backbone dihedral angle *cross terms* involving $\varphi = C'-N-C_\alpha-C_\beta$ and $\psi = C_\beta-C_\alpha-C'-N$ (except for glycines) with the goal of further improving the force field quality. Cross terms between φ and ψ, which are part of the CMAP correction of the CHARMM 22 force field,[3,42] have previously not been used in the Amber family of force fields. To keep the number of fitting parameters in our reweighting scheme reasonably small, we allow the addition of eight bivariate Gaussian potential energy terms (GPETs) to the 2D backbone dihedral angle energy surface where each GPET is defined as

$$\Delta V = f \cdot \exp\{ -a(\varphi - d)^2 - b(\varphi - d)(\psi - e) - c(\psi - e)^2 \} \quad (6)$$

where the parameters a, b, c, d, e, and f defining the center, widths, tilt angle, and height of each GPET are used as fitting parameters. Initially, three GPETs are placed in the α-helical region of φ,ψ space, three GPETs are placed in the β-strand region, and two GPETs are placed in the α_R regions. The half widths (standard deviations) of all GPETs are originally set to 10° along both φ and ψ, and all heights are set to zero. With each additional round of optimization, a new set of eight GPETs is added to the existing ones.

Protein Systems for Force Field Optimization and Validation. Protein GB3 (PDB code 1IGD) together with a pool of 22 proteins with variable sizes and topologies used previously[29] were selected for this study. GB3 is added because of its extensive set of backbone RDCs. All of these proteins were determined by X-ray crystallography (except for the NMR structure 2EA9) with a resolution of 2.1 Å or better, except for 1HIK and 3ILE. All NMR chemical shifts data are directly taken from the BMRB database.[34] The chemical shifts of GB1 were obtained from S. Grzesiek, and the ones of GB3 are taken from the examples provided with CS-ROSETTA.[43] Six proteins (1UBQ, 1IGD, 1HIK, 1ENH, 1SMX, and 1QZM) were selected as trial proteins in our force field optimization based on chemical shifts, while all others were used for validation. A comprehensive set of experimental backbone $N-H^N$ residual dipolar couplings measured in 23 different alignment media for ubiquitin[44] and a set of backbone $N-H^N$, N−Cα, Cα−C′, and $C\alpha-H^N$ RDCs measured in five different alignment media for GB3[45] have been reported in the literature. Backbone $N-H^N$ RDCs in two different alignment media for GB1 were taken from the BMRB. Three proteins (1UBQ, 1IGD, and 1PGA) were selected for force field optimization based on RDCs.

To test the dynamics properties of the proteins under the new force field, NMR relaxation backbone $N-H^N$ S^2 order parameters[46,47] of the three proteins, ubiquitin (1UBQ), lysozyme (6LYT), and interleukin-4 (1HIK), are back-calculated and compared with experimental results using the iRED[48] approach (without correction for zero-point vibrations[49]).

Unfolded Peptide System. To test the new force field on unfolded and intrinsically disordered polypeptides, the Ala_3 peptide was simulated at 300 K for 1 μs in explicit water, and back-calculated vicinal scalar J-coupling constants were compared with experimental results. Both termini of the Ala_3 peptide are protonated to be consistent with experimental conditions (pH 2).[50] Hence, the peptide has a net charge of $+e$, which is balanced by the addition of a Cl^- ion during the simulation. Because the charge distribution for the protonated, uncapped C-terminus of Ala_3 is not available in the literature, charges of the carboxyl group were taken from the side-chain of protonated Glu in the Amber ff99SB force field, and the charge of the C-terminal Cα atom was adjusted to ensure a total peptide charge of $+e$. We find that the details of the charge model of the C-terminus can have a significant influence on the structural propensity of this short peptide during the MD simulation.

As in previous work, the deviation between simulations and experiments is calculated as[51]

$$\chi^2 = N_{\mathrm{Jcoup}}^{-1} \sum_{i=1}^{N_{\mathrm{Jcoup}}} (\langle J_i \rangle_{\mathrm{sim}} - J_{i,\mathrm{exp}})^2 / \sigma_i^2 \quad (7)$$

where N_{Jcoup} is the total number of J couplings, $\langle J_i \rangle_{\mathrm{sim}}$ is the average coupling constant back-calculated from the MD trajectory, and $J_{i,\mathrm{exp}}$ is the corresponding experimental value. The coupling constants used here are ^3J(HN,Hα), ^3J(HN,C′), ^3J(Hα,C′), ^3J(C′,C′), ^3J(HN,Cβ), ^1J(N,Cα), ^2J(N,Cα), and ^3J(HN,Cα). The ^1J(N,Cα) coupling of the C-terminal residue was not included in eq 7 due to its strong sensitivity to the precise charges and their distribution in the C-terminus. The corresponding Karplus parameters, the experimental values, and errors σ_i included in eq 7 were directly taken from the literature.[50−52]

MD Simulations. All MD simulations were performed using the Gromacs 4.5 package[53−56] with its built-in support of dihedral angle cross terms. Backbone cross terms were added to the Gromacs topology file manually, after it was generated by

Table 1. Average Chemical Shift RMSDs (ppm) for the Six Trial Proteins after Each Round of Optimization for Both Reweighted Trajectories ("reweighted") As Well As New MD Simulations ("new")

		round 1		round 2		round 3	
PDB	99SB	reweighted MD	new MD	reweighted MD	new MD	reweighted MD	new MD
1UBQ	3.14	3.04	3.06	2.99	3.02	2.98	2.99
1IGD	3.43	3.39	3.43	3.41	3.39	3.38	3.36
1ENH	3.86	3.78	3.83	3.74	3.62	3.61	3.73
1HIK	3.30	3.17	3.09	3.08	3.23	3.21	3.26
1SMX	3.98	3.80	3.70	3.66	3.78	3.77	3.64
1QZM	2.13	1.97	2.02	1.96	1.91	1.87	1.93

Gromacs' PDB2GMX module. All protein MD simulations (i.e., both parent trajectories and validation trajectories) were run for 100 ns at 300 K using the TIP3P water model. The simulation of the Ala_3 peptide was run for 1 μs. The proteins as well as the peptide are uncapped. The integration time step was set to 2 fs with all bond lengths involving hydrogen atoms constrained by the SETTLE algorithm. Electrostatic interactions were cut off at 10 Å, and the long-range electrostatic interactions were calculated using the PME algorithm with 1.2 Å spacing. van der Waals interactions were cut off at 8 Å. The initial protein structures listed in Table 2 were taken from the Protein Data Bank.[57] The corresponding PDB codes are listed in the first column of Table 2. Standard minimization and heating procedures described previously[58] were applied before the final production runs at a constant temperature and pressure (NPT ensemble) of 300 K and 1 atm, respectively.

RESULTS

Force Field Optimization Using Chemical Shifts. Application of the reweighting strategy depicted in Figure 1 to the six trial proteins using only NMR chemical shift information leads to a significant drop in the chemical shift RMSD. Subsequent application of the modified force field to the validation set indicates that the force field has indeed improved. Further improvement is achieved by two additional rounds of optimization, but a fourth iteration did not yield any further improvement. The chemical shift RMSDs obtained from the reweighted trajectories as well as from new sets of MD runs are summarized in Table 1 for the first three rounds of optimization. The force field obtained after three rounds is named ff99SB_$\varphi\psi$(g24;CS), where ff99SB indicates the parent force field, $\varphi\psi$ stands for "backbone dihedral potential", g24 stands for the 24 Gaussian potential functions added, and CS stands for chemical shift information as input for refinement. (According to this terminology, the force field ff99SBnmr1[29] becomes ff99SB_$\varphi\psi$(f32;CS), where f32 stands for the 32 Fourier series coefficients.)

The modification made to the backbone dihedral φ,ψ energy surface from the new ff99SB_$\varphi\psi$(g24;CS) force field is shown in Figure 2A. The new force field fine-tunes the energy surface both in the α-helical and β-strand basins and makes the left-handed α-region (α_L) more stable. In the α-helical basin (α_R), the new force field renders conformations centered around (φ,ψ) = $(-50°, -50°)$ more stable, while it slightly destabilizes conformations centered around (φ,ψ) = $(-50°, -25°)$. In the β-strand basin, the new force field stabilizes both β-sheet conformations and polyproline II conformations while adding a small barrier between them.

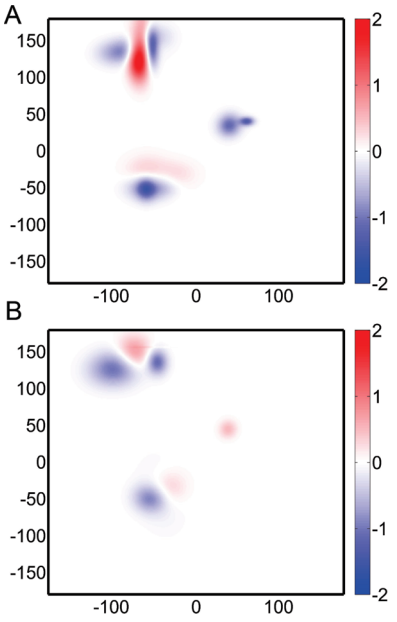

Figure 2. Optimization results following the scheme of Figure 1 for backbone φ,ψ dihedral angle potential modifications using Cα, Cβ, and C′ chemical shifts of six proteins (1UBQ, 1IGD, 1ENH, 1HIK, 1QZM, and 1SMX) (A) without RDCs (force field ff99SB_$\varphi\psi$(g24;CS)) and (B) with RDCs of three proteins (1UBQ, 1IGD, and 1PGA) as input (force field ff99SB_$\varphi\psi$(g8;CS,RDC)).

Next, the new force field is validated with the 17 proteins that were excluded during optimization. As shown in Figure 3A, the average improvement found for the validation protein set is similar to that of the six trial proteins, which provides evidence that the new backbone dihedral cross term is (i) not the result of overfitting and (ii) is transferable to a broad range of globular proteins of variable topology and size, as illustrated in Figure 3D. It is noted that the overall improvement is larger than that obtained for ff99SBnmr1 (see Figure 4).[29] The only exception is protein 2RNJ, whose structure was determined by NMR. A comparison of the performance of ff99SB and its optimized variants is provided in Table 2.

Besides the accurate reproduction of chemical shifts, a favorable protein force field is expected to stabilize native state protein structure. This property of ff99SB_$\varphi\psi$(g24;CS) is assessed here on the basis of average backbone RMSDs (excluding flexible N- and C-termini) of 100 ns MD trajectories with respect to initial PDB structures. A comparison of these results with the corresponding results obtained for ff99SB are shown in Figure 5A.

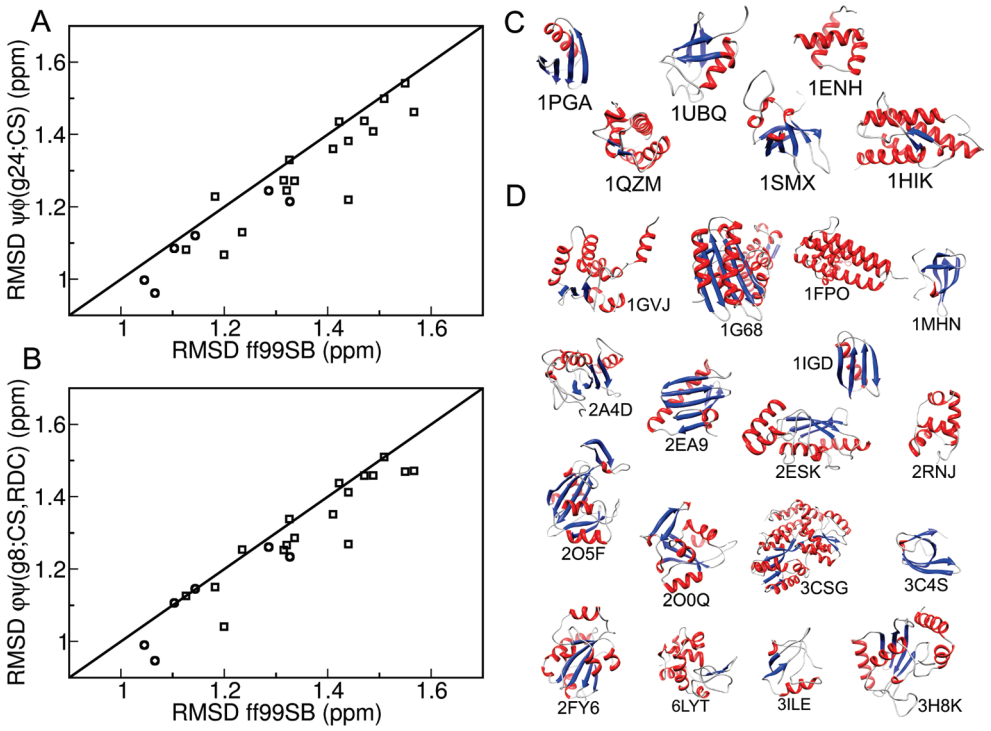

Figure 3. Test of the performance of the new force fields (A) ff99SB_$\varphi\psi$(g24;CS) and (B) ff99SB_$\varphi\psi$(g8;CS,RDC) using the combined Cα, Cβ, and C′ chemical shift RMSD from 100 ns MD trajectories of 23 proteins shown in C and D. The average chemical shift RMSDs (eq 3) are compared with the ones obtained for the parent force field ff99SB. The circles belong to the six proteins used during optimization, and the 17 squares belong to proteins used only for validation. Proteins below the diagonal show improved performance with the new force field. Ribbon models of (C) six trial proteins and (D) 17 validation proteins together with PDB codes (α-helices are in red and β-strands in blue).

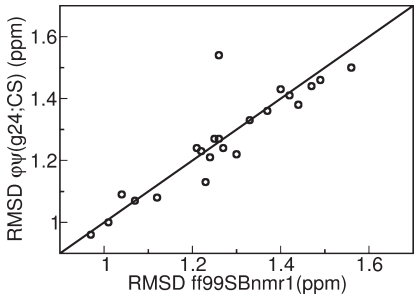

Figure 4. Comparison of the performance of the new force field ff99SB_$\varphi\psi$(g24;CS) and ff99SBnmr1[29] (i.e., ff99SB_$\varphi\psi$(f32;CS)) in terms of the combined Cα, Cβ, and C′ chemical shift RMSD of the 23 proteins depicted in Figure 3C,D.

The new force field ff99SB_$\varphi\psi$(g24;CS) leads to a significantly reduced average RMSD: for 13 out of the 23 proteins, it yields an RMSD that is lower than for ff99SB, whereas only six proteins show a weak opposite trend.

The force field can be cross-validated also by RDC data, as such data were not used during optimization. For ubiquitin, the average Q value (excluding the flexible C-terminal residues 72−76) improves from $\langle Q \rangle$ = 0.27 for ff99SB to $\langle Q \rangle$ = 0.22 for ff99SB_$\varphi\psi$(g24;CS), confirming the improved performance of the new force field. Considering that dipolar couplings and NMR chemical shifts reflect protein geometry and fluctuations in highly complementary ways, this underscores that the new force field more accurately captures structural-dynamic properties of this protein. For GB3, back-calculation of the RDCs yields Q values for ff99SB_$\varphi\psi$(g24;CS) (ff99SB) of 0.14 (0.12), 0.22 (0.21), 0.12 (0.11), and 0.23 (0.22) for Cα−C′, C′−N, Cα−Hα, and $N-H^N$ couplings, respectively. The new Q values are just slightly higher than the ones of ff99SB and on average are still very low. Similarly, back-calculated RDCs for GB1 yield a Q value of 0.14, which is only slightly larger than the original one of ff99SB, which is 0.12. Hence, the performance of the new force field is successful, as reflected by the low Q values when reproducing these RDCs.

Not only do modifications of force fields alter the distribution of conformer populations, but they can also affect the dynamics properties of proteins. NMR relaxation data, particularly S^2 order parameters, are very well-suited to evaluate the performance of the force field for protein dynamics on pico- to nanosecond time scales. Backbone $N-H^N$ S^2 order parameters are calculated for ubiquitin[59] (1UBQ), lysozyme[60] (6LYT), and interleukin-4[61] (1HIK) for both ff99SB and ff99SB_$\varphi\psi$(g24;CS) using iRED,[48] with a detailed comparison shown in Figure 6. The overall level of agreement is almost identical for the two force fields, although relatively minor differences can be found for specific regions.

Combined Force Field Optimization Using Chemical Shifts and RDCs. Chemical shift and RDC information can be directly combined in the scheme of Figure 1 for the optimization of force fields. After one round of optimization with chemical shifts of the six proteins used above and RDCs of the three proteins with a total of eight bivariate Gaussian corrections, a modified φ,ψ energy surface is obtained, termed ff99SB_$\varphi\psi$(g8;CS,RDC), which is shown in Figure 2B. A second round of optimization did not yield further improvement. The changes are similar to the ones in Figure 2A, except that ff99SB_$\varphi\psi$(g8;

Table 2. Comparison of the Performance of the ff99SB Force Field and Its Variants for the Prediction of ^{13}C Chemical Shifts, Expressed As RMSDs (eq 4), for Six Trial Proteins and 17 Validation Proteins

PDB code	resolution (Å)	BMRB code	ff99SB	ff99SB_$\varphi\psi$(g24;CS)	ff99SB_$\varphi\psi$(g8;CS,RDC)	ff99SBnmr1 (ff99SB_$\varphi\psi$(f32;CS))
1UBQ[a]	1.8	6475	1.04	1.00	0.99	1.01
1IGD[a]	1.1	n/a	1.14	1.12	1.14	n/a
1ENH[a]	2.1	15536	1.28	1.24	1.26	1.21
1HIK[a]	2.6	4094	1.10	1.09	1.11	1.04
1SMX[a]	1.8	6122	1.33	1.21	1.23	1.24
1QZM[a]	1.9	5107	1.07	0.96	0.95	0.97
1PGA	2.1	n/a	1.20	1.07	1.04	1.07
6LYT	1.9	4562	1.49	1.41	1.46	1.42
1G68	2.0	6838	1.44	1.38	1.41	1.44
1MHN	1.8	4899	1.18	1.23	1.15	1.22
2EA9	2.1	15088	1.51	1.50	1.51	1.56
2RNJ	NMR	11024	1.55	1.54	1.47	1.26
2ESK	1.4	6277	1.32	1.33	1.34	1.33
3CSG	1.8	4986	1.47	1.44	1.46	1.47
2FY6	1.9	6709	1.34	1.27	1.29	1.26
1GVJ	1.5	5991	1.41	1.36	1.35	1.37
3C4S	1.7	15604	1.32	1.24	1.27	1.27
3H8K	1.8	6711	1.31	1.27	1.25	1.25
3ILE	2.9	16325	1.23	1.13	1.25	1.23
1FPO	1.8	15541	1.13	1.08	1.12	1.12
2A4D	1.7	7219	1.44	1.22	1.27	1.30
2O0Q	1.8	15281	1.42	1.43	1.44	1.40
2O5F	1.9	5570	1.57	1.46	1.47	1.49

[a] These proteins were used during force field optimization, while all others were strictly used for validation purposes only.

CS,RDC) slightly destabilizes the α_L region. Under the new force field, the Q values of two proteins and chemical shifts RMSDs of four proteins drop significantly while the chemical shifts RMSDs of the remaining two proteins are essentially unchanged (Table 3a and b). (It is noted that the Q values obtained with local reweighting cannot be directly compared with Q values obtained from MD.)

Validation results with chemical shifts of proteins, not included during optimization, are given in Figure 3B and Table 2. They demonstrate the improvement of the new force field over ff99SB and a performance that is comparable to that of ff99SB_$\varphi\psi$(g24;CS). The protein stability test (Figure 5B) shows backbone RMSDs similar to ff99SB_$\varphi\psi$(g8;CS,RDC) and ff99SB. Backbone $N-H^N$ S^2 order parameters calculated for ubiquitin (1UBQ), lysozyme (6LYT), and interleukin-4 (1HIK) are displayed in Figure 6. The overall quality of the agreement is remarkably high, similar to the trajectories obtained with ff99SB_$\varphi\psi$(g24;CS) and ff99SB.

Force Field Validation for Ala_3 Peptide. Modification of backbone dihedral angle terms is expected to significantly affect also the properties of intrinsically disordered proteins and short peptides, as these systems are less stabilized by other interactions such as intramolecular hydrogen bonds and hydrophobic interactions. Experimental J-coupling constants[50] of the Ala−Ala−Ala peptide are utilized to evaluate the three force fields. The χ^2 values (eq 7) obtained for a 1 μs trajectory of Ala_3 are 1.69, 1.55, and 1.88 for ff99SB, ff99SB_$\varphi\psi$(g24;CS), and ff99SB_$\varphi\psi$(g8;CS,RDC), respectively. A plot of the comparison between experimental and predicted J-scalar coupling constants in Figure 7 shows that for essentially all couplings the predictions fall within the experimental error bar, consistent with the low χ^2 values. Considering this fact, together with the uncertainties of the associated Karplus parameters used for J-coupling back-calculation, the agreement is very good. This result demonstrates that the new force fields adequately represent the free energy balance between the different secondary structural basins.

DISCUSSION

The availability of better force fields is not only beneficial for computational chemists and biophysicists. Experimental structural biologists, in particular, protein X-ray crystallographers and NMR spectroscopists, have a longstanding tradition in using computer simulations of proteins that include energy restraints derived from experimental data to refine protein structures. Recently, an increasing number of experimental biochemists are using MD simulations as "*in silico* experiments" without experimental restraints, to interpret their experimental data from a variety of sources, including calorimetry, small-angle X-ray scattering, magnetic resonance, and optical spectroscopy.[62−68] Taken together, these demands make the development of more accurate biomolecular force fields a timely task.

Traditionally, force field parametrizations have been primarily based on quantum-chemical energy calculations of peptide fragments in a vacuum.[2,14,16] Despite the success of this approach in the past, the potential for further force field improvements using ever higher level quantum-chemical calculations and larger basis sets is unclear. An accurate representation of protein interactions in a vacuum does not guarantee that the energetics are correctly reproduced when the protein is placed in its native environment (water, ions, lipid bilayers, etc.). This obstacle can be overcome, in principle, by optimizing the force field directly

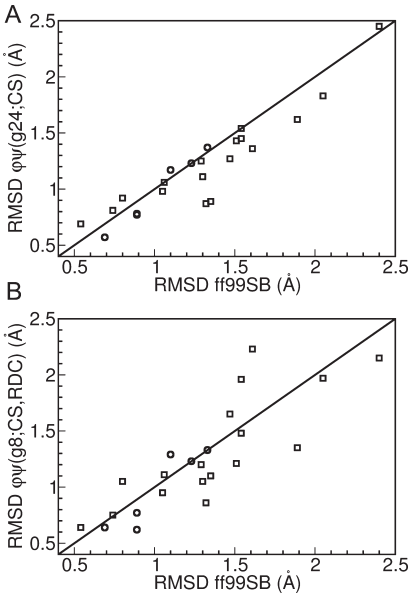

Figure 5. Test of the performance of the new force fields (A) ff99SB_$\varphi\psi$(g24;CS) and (B) ff99SB_$\varphi\psi$(g8;CS,RDC) relative to ff99SB by comparing the average Cartesian coordinate backbone RMSD of the 100 ns validation MD simulations with respect to the X-ray crystal structures. The circles belong to the six proteins used during optimization, and the squares belong to the 17 proteins used only for validation.

against experimental data of full-length proteins in the native environment.

This approach, which is pursued in this work using NMR data of intact globular proteins, permits by means of superposition of multiple bivariate Gaussian correction terms versatile and specific modifications of positions, shapes, and correlations of the favorable and unfavorable energy regions in φ,ψ space. On the other hand, it is unlikely that the method changes the relative average weights between distinct regions, because during the simulation of a globular protein it is quite rare that a residue visits both α and β regions. Hence, free energy differences between these regions are mostly inconsequential for the back-calculation of NMR data of globular proteins. The global balance between the α-helical and β-strand basins in the optimization of backbone dihedral angle terms is best achieved[18] using experimental data of (partially) unfolded systems.

In our previous work,[29] the backbone dihedral angle potential was modified through adjustment of the Fourier series coefficients truncated at the fourth order. The truncation at relatively low order was imposed by practical considerations to keep the number of fitting parameters manageable. However, it gives each Fourier component, as well as any of their superpositions, a significant nonlocal character and thereby has the tendency to introduce unwanted "wiggles" in the φ,ψ energy map. For the present work, we therefore decided to allow a "basis set" of potential energy correction terms in the form of bivariate Gaussian functions that, depending on the widths, can have both a local as well as a more global character.

The chosen length of the protein MD trajectories, which has been set to 100 ns, strikes a balance between computational affordability and convergence. In previous work, the comparison between experimental ^{13}C carbon chemical shifts and predicted ones from a microsecond MD simulation of ubiquitin showed the onset of convergence around 100 ns.[69] Moreover, the use of six to

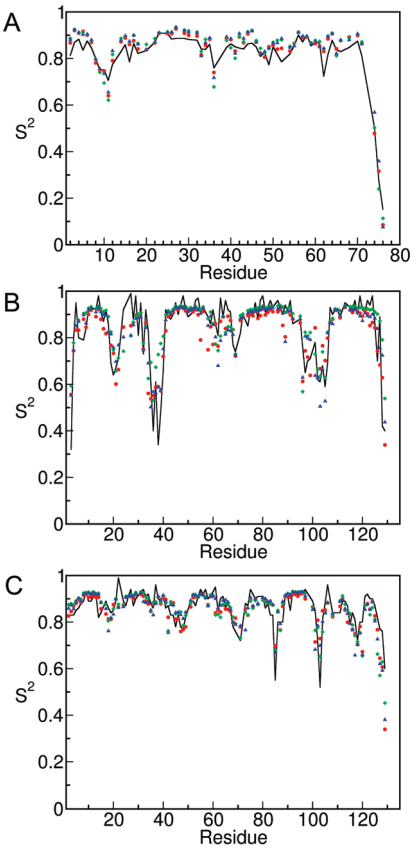

Figure 6. Comparison of the new force fields for the calculation of NMR N−H^N S^2 order parameters of (A) ubiquitin, (B) interleukin-4, and (C) lysozyme. The S^2 values were computed by iRED from 100 ns MD trajectories with averaging done over (A) 4, (B) 6, and (C) 8 ns time windows. Black lines belong to experimental values. Red circles, green diamonds, and blue triangles belong to ff99SB, ff99SB_$\psi\varphi$(g24;CS), and ff99SB_$\psi\varphi$(g8;CS,RDC), respectively.

eight proteins during optimization, as opposed to four proteins used previously, further enhances the stability of the force field parameter fitting. Chemical shifts represent the most abundant experimental NMR parameter of proteins. Their predictions from a given structural model (or ensemble) have dramatically improved over the past decade, but they are still not fully quantitative. This explains, at least in part, the observed RMSD offset observed even when the force field optimization protocol has converged. RDCs provide a complementary source of structural and dynamics information, and they are well suited for the optimization task at hand. However, the number of RDC sets reported in the literature for proteins for which a well-resolved X-ray crystal structure is available is still very sparse. Therefore, RDCs of only three proteins were included during optimization. The resulting force field ff99SB_$\varphi\psi$(g8;CS,RDC) performs similarly well to the one optimized with chemical shift data only. The average backbone RDC Q value of ubiquitin of a free MD simulation using this force field now drops to 0.19, which is only slightly higher than Q_{free} reported in an RDC optimized accelerated MD simulation.[70] As high-quality RDC sets of more proteins with known X-ray crystal structures are becoming available, their utility for force-field optimization holds significant promise.

Our optimization protocol can be further expanded to incorporate other NMR data, including T_1, T_2, and nuclear

Table 3. Force Field Optimization Results Using Both Chemical Shifts and RDCs for (a) Chemical Shift Agreement[a] and (b) RDC *Q* Values for ff99SB Trajectories, Reweighted Trajectories, and New MD Simulations

(a) Chemical shift RMSDs

PDB	ff99SB	reweighted MD	new MD
1UBQ	3.14	3.03	2.97
GB3	3.43	3.40	3.44
1ENH	3.80	3.77	3.78
1HIK	3.30	3.20	3.32
1QZM	2.12	1.96	1.89
1SMX	3.98	3.86	3.70

(b) RDC *Q* Values

PDB		ff99SB	Q_{local} ff99SB	Q_{local} reweighted	*Q* new
1UBQ		0.26	0.55	0.51	0.19
1PGA		0.12	0.32	0.31	0.13
1IGD	NH	0.12	0.42[b]	0.42[b]	0.12
	CH	0.21			0.20
	CN	0.22			0.18
	CαC	0.11			0.11

[a] Average chemical shift RMSDs (ppm) for both reweighted trajectories ("reweighted") and new MD simulations ("new"). [b] Q_{local} corresponds to the average over the 4 internuclear vectors.

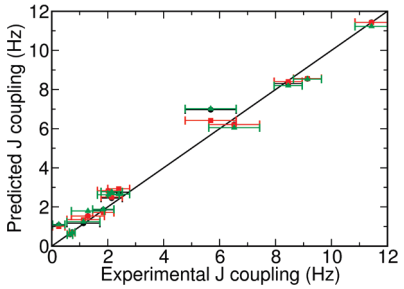

Figure 7. Comparison of back-calculated and experimental scalar J-coupling constants of Ala_3 listed below eq 7 for different force fields: ff99SB (black), ff99SB_$\varphi\psi$(g24;CS) (red), and ff99SB_$\varphi\psi$(g8;CS, RDC) (green).

Overhauser enhancement (NOE) spin-relaxation parameters of ^{15}N spins,[71] ^{13}C spins,[72] and ^{2}H spin relaxation parameters[73] of ^{13}C-D groups and partially deuterated $^{13}CDH_2$-methyl moieties of methyl-bearing side chains.[74,75] These parameters can be expressed in terms of model-free S^2 order parameters and intramolecular correlations times τ in the pico- to nanosecond range, which can be used to actively guide force field improvements using methods similar to the ones presented here. Although we focused here on NMR data, obviously our approach can be combined with other types of experimental data too. The average Cartesian coordinate RMSD between the MD trajectory and the high-resolution X-ray structure has been found to correlate well with the quality of the force field. Although in Figure 5 this information serves for cross-validation purposes only, it can also be used to actively guide force field improvements. In this case, one should either simulate proteins in their full crystalline environment[76] or exclude during optimization those atoms that are significantly affected by crystal packing.[77] Hence, the Cartesian RMSD of selected sets of atoms can be employed as a metric to assess and improve force fields. Meanwhile, crystallographic B factors, which report on both dynamics and static disorder in protein crystals, represent another measure that is worth exploring for force field validation and refinement.

The present work focuses on backbone dihedral angles, although the extension to side chain dihedral angles is rather straightforward. Most side-chain potentials have not been updated for well over a decade. A recent quantum chemical study showed that amino acid specific potentials for ILDN residues can provide improvements.[78] Although NMR parameters of side-chains are less abundant than for the backbone, side-chain assignments (and thus chemical shifts) are becoming available for an increasing number of proteins and, thus, can be used in full analogy to backbone chemical shifts. Our optimization protocol can be also applied to force field terms other than backbone dihedral angle terms, such as Coulomb interactions, van der Waals interactions, and explicit hydrogen-bonding potentials. The latter terms were used in the early phase of force field development,[11] subsequently abolished, and have been revisited more recently.[79,80] Similarly, it should be possible to further improve the parametrization of explicit water models to better reflect structural, dynamic, and thermodynamic properties of protein solutes in their native solvent. Finally, with the imminent advent of polarizable force fields[81−84] for biomacromolecules, the optimization approach presented here should find useful application in this emerging field.

CONCLUSION

MD simulations have been limited in the past by the accuracy of force fields and limited sampling of conformational space.[85] While sampling improves with every new generation of computer hardware, the development of better force fields has been a relatively slow, labor-intensive task mostly based on small molecules and protein fragments in a vacuum. However, the performance of a force field when applied to an entire protein in explicit solvent is more than the sum of its isolated parts. The work presented here shows how the confluence of efficient optimization protocols and protein structural and NMR databases make the development of force fields on intact full-length proteins feasible and computationally tractable. These advances enable a plethora of opportunities to render current molecular mechanics force fields increasingly quantitative, which remains a key prerequisite for the predictive understanding of biomolecular properties and function.

AUTHOR INFORMATION

Corresponding Author

*Tel.: 850-644-1768. Fax: 850-644-8281. E-mail: bruschweiler@magnet.fsu.edu.

ACKNOWLEDGMENT

This work was supported by grant MCB-0918362 of the National Science Foundation.

REFERENCES

(1) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120–127.

(2) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.

(3) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
(4) Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961–975.
(5) Trbovic, N.; Cho, J. H.; Abel, R.; Friesner, R. A.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **2009**, *131*, 615–622.
(6) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. *Biophys. J.* **2008**, *94*, L75–L77.
(7) Freddolino, P. L.; Park, S.; Roux, B.; Schulten, K. *Biophys. J.* **2009**, *96*, 3772–3780.
(8) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; Wriggers, W. *Science* **2010**, *330*, 341–346.
(9) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
(10) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(11) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(12) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(13) MacKerell, A. D.; Banavali, N.; Foloppe, N. *Biopolymers* **2000**, *56*, 257–265.
(14) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
(15) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(16) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
(17) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
(18) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
(19) Warshel, A. *Isr. J. Chem.* **1973**, *11*, 709–717.
(20) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
(21) Levitt, M.; Hirshberg, M.; Sharon, R.; Daggett, V. *Comput. Phys. Commun.* **1995**, *91*, 215–231.
(22) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490–519.
(23) Arnautova, Y. A.; Jagielska, A.; Scheraga, H. A. *J. Phys. Chem. B* **2006**, *110*, 5025–5044.
(24) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
(25) Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; Van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719–1751.
(26) Nevins, N.; Lii, J. H.; Allinger, N. L. *J. Comput. Chem.* **1996**, *17*, 695–729.
(27) Langley, C. H.; Allinger, N. L. *J. Phys. Chem. A* **2003**, *107*, 5208–5216.
(28) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 915–920.
(29) Li, D. W.; Brüschweiler, R. *Angew. Chem., Int. Ed.* **2010**, *49*, 6778–6780.
(30) Xu, X. P.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321–333.
(31) Xu, X. P.; Case, D. A. *Biopolymers* **2002**, *65*, 408–423.
(32) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.
(33) Brüschweiler, R. *J. Chem. Phys.* **1995**, *102*, 3396–3403.
(34) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. *Nucleic Acids Res.* **2008**, *36*, D402–408.
(35) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1114.
(36) Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371–424.
(37) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 1416–1424.
(38) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098–6107.
(39) Showalter, S. A.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 4158–4159.
(40) Lange, O. F.; van der Spoel, D.; de Groot, B. L. *Biophys. J.* **2010**, *99*, 647–655.
(41) Losonczi, J. A.; Andrec, M.; Fischer, M. W.; Prestegard, J. H. *J. Magn. Reson.* **1999**, *138*, 334–342.
(42) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. *Biophys. J.* **2006**, *90*, L36–L38.
(43) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 4685–4690.
(44) Lakomek, N. A.; Walter, K. F. A.; Fares, C.; Lange, O. F.; de Groot, B. L.; Grubmuller, H.; Brüschweiler, R.; Munk, A.; Becker, S.; Meiler, J.; Griesinger, C. *J. Biomol. NMR* **2008**, *41*, 139–155.
(45) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179–9191.
(46) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559–4570.
(47) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
(48) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522–4534.
(49) Brüschweiler, R. *J. Am. Chem. Soc.* **1992**, *114*, 5341–5344.
(50) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.
(51) Wickstrom, L.; Okur, A.; Simmerling, C. *Biophys. J.* **2009**, *97*, 853–856.
(52) Case, D. A.; Scheurer, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2000**, *122*, 10390–10397.
(53) Berendsen, H. J. C.; van der Spoel, D.; van der Unen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
(54) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
(55) van der Spoel., D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
(56) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
(57) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(58) Li, D. W.; Meng, D.; Brüschweiler, R. *J. Am. Chem. Soc.* **2009**, *131*, 14610–14611.
(59) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870–9879.
(60) Buck, M.; Boyd, J.; Redfield, C.; Mackenzie, D. A.; Jeenes, D. J.; Archer, D. B.; Dobson, C. M. *Biochemistry* **1995**, *34*, 4041–4055.
(61) Redfield, C.; Smith, L. J.; Smith, R. A. G.; Dobson, C. M. *Biochemistry* **1992**, *31*, 10431–10437.
(62) Abbate, S.; Barlati, S.; Colombi, M.; Fornili, S. L.; Francescato, P.; Gangemi, F.; Lebon, F.; Longhi, G.; Manitto, P.; Recca, T.; Speranza, G.; Zoppi, N. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4668–4677.
(63) Lim, C. C.; Yang, H.; Yang, M.; Wang, C. K.; Shi, J.; Berg, E. A.; Pimentel, D. R.; Gwathmey, J. K.; Hajjar, R. J.; Helmes, M.; Costello, C. E.; Huo, S.; Liao, R. *Biophys. J.* **2008**, *94*, 3577–3589.

(64) El-Turk, F.; Cascella, M.; Ouertatani-Sakouhi, H.; Narayanan, R. L.; Leng, L.; Bucala, R.; Zweckstetter, M.; Rothlisberger, U.; Lashuel, H. A. *Biochemistry* **2008**, *47*, 10740–10756.
(65) Feng, J. W. A.; Kao, J.; Marshall, G. R. *Biophys. J.* **2009**, *97*, 2803–2810.
(66) Pelikan, M.; Hura, G. L.; Hammel, M. *Gen. Physiol. Biophys.* **2009**, *28*, 174–189.
(67) Daneshgar, P.; Moosavi-Movahedi, A. A.; Norouzi, P.; Ganjali, M. R.; Madadkar-Sobhani, A.; Saboury, A. A. *Int. J. Biol. Macromol.* **2009**, *45*, 129–134.
(68) Lerbret, A.; Affouard, F.; Bordat, P.; Hedoux, A.; Guinet, Y.; Descamps, M. *J. Chem. Phys.* **2009**, 131.
(69) Li, D. W.; Brüschweiler, R. *J. Phys. Chem. Lett.* **2010**, *1*, 246–248.
(70) Markwick, P. R. L.; Bouvignies, G.; Salmon, L.; McCammon, J. A.; Nilges, M.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 16968–16975.
(71) Jarymowycz, V. A.; Stone, M. J. *Chem. Rev.* **2006**, *106*, 1624–1671.
(72) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. *Chem. Rev.* **2006**, *106*, 1672–1699.
(73) Sheppard, D.; Li, D. W.; Brüschweiler, R.; Tugarinov, V. *J. Am. Chem. Soc.* **2009**, *131*, 15853–15865.
(74) Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6439–6448.
(75) Skrynnikov, N. R.; Millet, O.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6449–6460.
(76) Cerutti, D. S.; Freddolino, P. L.; Duke, R. E.; Case, D. A. *J. Phys. Chem. B* **2010**, *114*, 12811–12824.
(77) Li, D. W.; Brüschweiler, R. *Biophys. J.* **2009**, *96*, 3074–3081.
(78) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–1958.
(79) Wroblewska, L.; Jagielska, A.; Skolnick, J. *Biophys. J.* **2008**, *94*, 3227–3240.
(80) Huang, J.; Meuwly, M. *J. Chem. Theory Comput.* **2010**, *6*, 467–476.
(81) Warshel, A.; Kato, M.; Pisliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.
(82) Lopes, P. E. M.; Roux, B.; MacKerell, A. D. *Theor. Chem. Acc.* **2009**, *124*, 11–28.
(83) Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; Wang, J. M. *J. Phys.: Condens. Matter* **2009**, *21*, 333102.
(84) Ponder, J. W.; Wu, C. J.; Ren, P. Y.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
(85) Yang, W.; Nymeyer, H.; Zhou, H. X.; Berg, B.; Brüschweiler, R. *J. Comput. Chem.* **2008**, *29*, 668–672.