JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2019, 15, 6769−6780

pubs.acs.org/JCTC

Well-Balanced Force Field *ff03CMAP* for Folded and Disordered Proteins

Yangpeng Zhang,[†,∥] Hao Liu,[†,∥] Sheng Yang,[†,∥] Ray Luo,*[,‡] and Hai-Feng Chen*[,†,§]

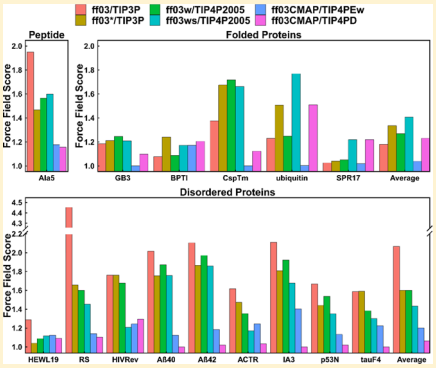
[†]State Key Laboratory of Microbial metabolism, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

[‡]Departments of Molecular Biology and Biochemistry, Chemical and Molecular Engineering, and Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine, California 92697, United States

[§]Shanghai Center for Bioinformation Technology, Shanghai 200235, China

Supporting Information

ABSTRACT: Molecular dynamics simulation as an important complement of experiment is widely used to study protein structures and functions. However, previous studies indicate that the current force fields cannot, simultaneously, provide accurate descriptions of folded proteins and intrinsically disordered proteins (IDPs). Therefore, a correction maps (CMAP)-optimized force field based on the Amber *ff*03 force field (termed *ff*03*CMAP* herein) was developed for a balanced sampling of folded proteins and IDPs. Extensive validations of short peptides, folded proteins, disordered proteins, and fast-folding proteins show that simulated chemical shifts, *J*-coupling constants, order parameters, and residual dipolar couplings (RDCs) with the *ff*03*CMAP* force field are in very good agreement with nuclear magnetic resonance measurements and are more accurate than other *ff*03-series force fields. The influence of solvent models was also investigated. It was found that the combination of *ff*03*CMAP*/TIP4P-Ew is suitable for folded proteins, and that of *ff*03*CMAP*/TIP4PD is better for disordered proteins. These findings confirm that the newly developed force field *ff*03*CMAP* can improve the balance of conformer sampling between folded proteins and IDPs.



INTRODUCTION

Both folded proteins and disordered proteins are related to important biological processes. Folded proteins are easier to study because they are ordered and stable. But disordered proteins also need exploring. In eukaryotes, more than 30% of proteins contain disordered regions with more than 50 consecutive residues.[1] Proteins with disordered regions or overall intrinsically disordered proteins (IDPs) have been proved to have important biological functions, such as molecular recognition, molecular assembly, protein modification, and so on.[2] Furthermore, IDPs are associated with many human diseases, such as Alzheimer's disease, Parkinson's disease, Huntington's disease, cancer, and cardiovascular disease, to name a few.[3,4] IDPs are more flexible and unstable with little secondary structures than structured or ordered proteins. "Intrinsically disordered" implies a sequence-dependent nature in IDPs that tend to be lack of ordered structures.[5] Many experimental methods have been utilized to study IDPs, such as electron paramagnetic resonance, X-ray diffraction, nuclear magnetic resonance (NMR), and small-angle X-ray scattering (SAXS).[6−8]

Because of their important biological functions, IDPs have become common topics in molecular dynamics studies in recent years. Standard protein force fields were developed based on lots of parameters from folded proteins, and with the development of techniques, significant discrepancies are shown in the comparison between the MD simulation and experimental measurements.[9] Due to limitation for none of the tested force fields such as ff99SB/TIP3P, C36m, C22*/TIP3P, ff03ws, and ff99SB/TIP4P-D, simultaneously provided accurate descriptions of folded proteins and disordered proteins. A set of special-purpose force fields has been developed for simulating IDPs, such as *ff99IDPs*, *ff14IDPs*, *ff14IDPSFF*, *ff03ws*, RSFF2, a99SB-disp, CHARMM-M36IDPSFF, and so on.[9−17] In addition, the D. E. Shaw group also modified the dispersion interaction of the TIP4P water model (TIP4P-D) to improve the quality of IDP simulations.[18] However, it remains elusive to reach a good balance between ordered and disordered states with either standard or special-purpose force fields.

*ff*03 is a new-generation Amber force field that has become widely used in biomolecular simulation studies.[19] Based on *ff*03, the best group modified backbone dihedral potentials in the context of the TIP4P/2005 water model. Their efforts lead to three *ff*03 variants: *ff*03*, *ff*03*w*, and *ff*03*ws*.[14,20,21] These modifications were shown to partially improve the performance of conformer sampling of IDPs and folded proteins. The

Received: June 21, 2019
Published: October 27, 2019

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769−6780

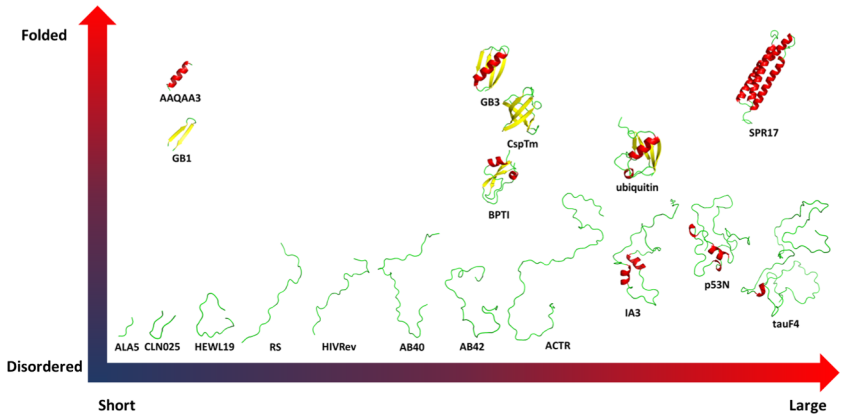


Figure 1. Test system from short peptides to proteins. Five representative folded proteins such as third immunoglobulin binding domain of protein G (GB3, α/β), free bovine pancreatic trypsin inhibitor (BPTI, α/β), cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima* (CspTm, all-β),[25] and ubiquitin of human (α/β)[26] and chicken brain α spectrin repeat 17 (SPR17, all-α).[27] Nine typical disordered protein, including 19 length peptide of hen egg-white lysozyme (HEWL19),[28] phosphorylated SRSF1 (RS),[29] HIV-1 Rev ARM peptide (HIVRev),[30] 40 length amyloid-β-peptides (AB40),[31] 42 length amyloid-β-peptides (AB42),[31] activation domain of the nuclear hormone receptor coactivator (ACTR),[32] an aspartic proteinase inhibitor for *Saccharomyces cerevisiae* (IA3),[33] p53 N-terminal transactivation domain (p53N),[34] and tau protein fragment (TauF4).[35] Three fast-folding proteins including 15-residue helix-forming peptide Ac-$(AAQAA)_3$-NH_2 (AAQAA3),[36] β-hairpin B1 domain of protein G (GB1),[37,38] and Chignolin, a 10 residue-folded peptide designed by segment statistics (CLN025).[39]

effect of the solvent model was found to be important in the sampling of IDPs in their studies.[14,20,21] TIP3P is the most commonly used water model in earlier force fields.[22] Improvement of water models has always been concurrent with protein force field developments. Besides TIP4P-D, TIP4P-Ew and TIP4P/2005 are also two commonly used four-site water models and should be investigated in any force field improvement efforts.[23,24] These four-site water models have been found to reproduce well the hydrophobic effect and water density in a wide temperature range.

In this development, we systematically analyzed the original *ff*03 force field, its published variants, and explored a new variant based on the grid-based energy correction map (CMAP) method, named *ff*03*CMAP*. The combinations of force fields and the corresponding solvent models are *ff*03 with TIP3P, *ff*03* with TIP3P, *ff*03*w* with TIP4P/2005, and *ff*03*ws* with TIP4P/2005 as in the published efforts. Based on our analyses, we recommend the combinations of *ff*03*CMAP* with TIP4P-Ew and TIP4P-D, respectively, suitable for folded proteins and IDPs.[14,19−21] To evaluate the performance of these force fields, 18 proteins were simulated to probe the quality of various force field/water model combinations in reproducing experimental observables. The tested short peptides and proteins are shown in Figure 1.

MATERIALS AND METHODS

Molecular Dynamics Simulations. Initial structures were built by the LEaP module in the AMBER 14 suite if not available,[40] which was also used to conduct MD simulations.[41] All systems were neutralized and solvated in boxes of different water models.[22] All bonds involving hydrogen atoms were constrained with the SHAKE algorithm.[42] The particle mesh Ewald (PME) algorithm was used to calculate long-range electrostatic interactions.[43] Initial structures were relaxed with 10 000 steps of minimization, then subjected to heating for 20 ps and equilibration for 10 ps in the NPT ensemble with PMEMD. The CUDA version of PMEMD was used to accelerate the simulations.[44] The simulation temperature and ion strength were set according to their respective experimental conditions. The number of replica-exchange molecular dynamics (REMD) replicas and temperatures were set by an online temperature predictor for parallel tempering simulations.[45] All simulation conditions are shown in Table S1.

Benchmark of PDB Coil Structures. The coil database was built and extracted from PDB. The DSSP program was utilized to classify the secondary structures and extract main-chain dihedrals from these structures.[46,47] Amino acids (2 611 450) without secondary structures were collected. The counts of amino acids in the coiled database are shown in Figure S1. The Ramachandran plots for the database were used as the benchmark for the optimization of dihedral distribution.

CMAP Method. Grid-based energy correction maps (CMAP) is a useful method for automatically correcting dihedral distribution for the additive force field, which is based on the backbone dihedral distribution and has been used to develop IDP-specific force fields.[12,48] CMAP was first published to modify the CHARMM force field and was transferred into Amber software.[11−13,49−52] We minimized the main-chain dihedral distribution differences between the MD simulation and the benchmark for each of the 20 amino acids. A 576-(24 × 24) grid was used to cover the phi/psi map. The tetrapeptide models (Nme-Ala-X-Ala-Ace, where X represents one of the 20 natural amino acids, Nme for aminomethyl, and Ace for acetyl) were utilized in the CMAP optimization via MD simulations in the TIP4PEw water model. Ten cycles of CMAP optimization were conducted to minimize the distribution differences between the MD simulation and the benchmark. In each cycle, the solvated tetrapeptides were simulated for 200 nanoseconds. After the CMAP optimization, we added an additional structural factor, which is the partial dihedral energy distribution of "S" fragments predicted by DSSP in the PDB coil database to avoid overestimation of the disordered state when using the new force field. Root-mean-square deviations of population (RMSp) is calculated to compare the difference between the MD simulation and the benchmark with eq 1.

$$\mathrm{RMSp}=\sqrt{\frac{\sum_{i=1}^{576}(P_i^{\mathrm{DB}}-P_i^{\mathrm{MD}})^2}{576}} \quad (1)$$

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769−6780

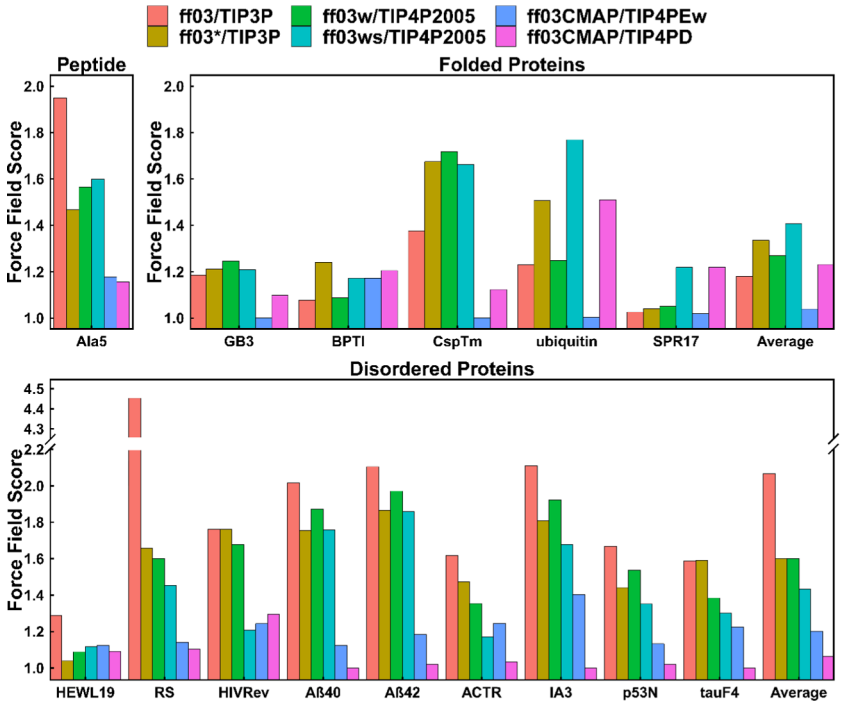


Figure 2. FF scores for short peptide, disordered proteins, and folded proteins. Average scores are also shown for folded proteins and disordered proteins. Fifteen tested systems included 1 short peptide, 5 folded proteins, and 9 disordered proteins, which are sorted by the residue length in each category.

where P_i^{DB} is the population of the ith grid in the database benchmark and P_i^{MD} is the population of the ith grid in the MD simulation.

Quantification in the Evaluation for Force Fields. To quantitatively compare different *ff*03 variants for folded proteins and disordered proteins with experimental measurements, we utilized the normalized force-field score.[9] For folded proteins, eq 2 was used to calculate the average normalized RMSD from each class of the experimental data as follows

$$\text{folded protein FF}_{\text{score}} = \frac{1}{N}\sum_{N}^{i=1}\frac{\text{FF}_{\text{rmsd}}}{\text{rmsd}_{\text{norm}}} \quad (2)$$

where N is the number of classes of experimental measurements, FF_{rmsd} is the RMSD of the ith class for simulated and experimental values, and $\text{rmsd}_{\text{norm}}$ is the lowest RMSD of ith class in all force fields. According to this metric, FF_{score} is always greater than or equal to 1, and 1 is the best score theoretically, which means that this force field perfectly reproduced the experimental data. For the disordered protein, we divided the experimental measurements into two groups because there are fewer experimental measurements than the folded proteins, for which there are chemical shifts and other NMR measurements. If the experimental data for both chemical shifts and NMR measurements are available, FF_{score} is calculated with eq 3.

$$\text{disordered protein FF}_{\text{score}} = \frac{\text{CS}_{\text{score}} + \text{NMR}_{\text{score}}}{2} \quad (3)$$

and if there is only a chemical shift, the FF_{score} is calculated with eq 4.

$$\text{disordered protein FF}_{\text{score}} = \text{CS}_{\text{score}} \quad (4)$$

where CS_{score} and $\text{NMR}_{\text{score}}$ are calculated the same as the score of class in folded protein FF_{score}.

Calculation of Experimental Observables. Backbone chemical shifts were calculated by SHIFTX2 for Cα, Cβ, C, N, Hα, and HN atom types.[53] Backbone scalar coupling constants were calculated using published Karplus relations for $^3J_{\text{HNH}\alpha}$, $^3J_{\text{H}\alpha\text{C}}$, $^3J_{\text{HNC}}$, $^3J_{\text{CC}}$, $^3J_{\text{HNC}\alpha}$, $^3J_{\text{HNC}\beta}$, $^2J_{\text{C}\alpha\text{N}}$, $^1J_{\text{C}\alpha\text{N}}$, $^1J_{\text{H}\alpha\text{C}\alpha}$, and $^1J_{\text{C}\alpha\text{C}\beta}$[54−62] and side-chain scalar coupling constants with Karplus relations for $^3J_{\text{CC}\gamma}$ and $^3J_{\text{NC}\gamma}$.[63] Backbone residual dipolar couplings (RDCs) were calculated using PALES with a local alignment window of 15 residues.[64,65] Backbone amide and the side-chain methyl axis S^2 order parameters were calculated with the direct method described in Trbovic et al.[66] Small-angle X-ray scattering (SAXS) curves were calculated using the FoXS package.[67] Cα RMSD and radius of gyrations (R_g) were calculated using CPPTRAJ in AmberTools.[40] Conformational clustering was performed with the kClust program in the MMTSB tool.[68] MDTraj, a python package, was also used for miscellaneous calculations.[69] The PyMOL molecular visualization system was used to show three-dimensional structures for all proteins.[70] All experimental measurements are listed in Table S2.

The biphasic exponential decay model was used to analyze the IDPs sampling convergence for force fields. Equations 5 and 6 were used to calculate the fitting half-time.[71]

$$\Delta\text{C}\alpha \text{ chemical shift } (N_t) = A_1\, e^{-(x/\tau_1)} + A_2\, e^{-(x/\tau_2)} + N_0 \quad (5)$$

$$t_{1/2} = \tau \ln(2) \quad (6)$$

where N_0 is the plateau of an observable, $t_{1/2}$ is the half-life time, A and τ are constants, respectively. In this model, the decay consists of two stages for the fast stage and slow stage. The slower τ_2 value was utilized to calculate the slow stage half-life and evaluate the convergence rate of IDP simulation.

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769−6780

Table 1. RMSD's of Secondary Chemical Shifts, *J*-Coupling Constants and Force-Field Scores of *ff*03-Series Force Fields for Ala5[a]

	ff03[19]/TIP3P[22]	ff03*[20]/TIP3P	ff03w[21]/TIP4P2005[24]	ff03ws[14]/TIP4P2005	ff03CMAP/TIP4PEw[23]	ff03CMAP/TIP4PD[18]
Cα	0.321	0.182	0.192	0.173	0.162	0.183
Cβ	0.372	0.474	0.499	0.514	0.192	0.204
C	0.297	0.453	0.471	0.490	0.346	0.348
N	1.633	1.022	1.112	1.177	0.417	0.415
Hα	0.045	0.057	0.055	0.057	0.024	0.021
HN	0.352	0.375	0.381	0.384	0.346	0.348
$^3J_{HNH\alpha}$	0.443	0.292	0.281	0.297	0.319	0.259
$^3J_{HNC\alpha}$	0.104	0.047	0.046	0.047	0.047	0.042
$^3J_{HNC\beta}$	0.374	0.234	0.264	0.275	0.359	0.296
$^3J_{H\alpha C}$	0.120	0.131	0.135	0.137	0.124	0.154
$^3J_{HNC}$	0.224	0.133	0.168	0.179	0.188	0.160
$^1J_{C\alpha N}$	0.265	0.092	0.141	0.136	0.166	0.179
$^2J_{C\alpha N}$	0.435	0.217	0.222	0.217	0.222	0.228
CS_{score}	1.995	1.888	1.950	1.997	1.049	1.061
NMR_{score}	1.905	1.047	1.178	1.202	1.307	1.250
FF_{score}	1.950	1.468	1.564	1.599	1.178	1.155

[a]Chemical shifts are in ppm, *J*-coupling constants are in Hz and the scores are unitless.

Table 2. FF Scores of Five Folded Proteins for Six *ff*03-Series Force Fields

protein	ff03/TIP3P	ff03*/TIP3P	ff03w/TIP4P2005	ff03ws/TIP4P2005	ff03CMAP/TIP4PEw	ff03CMAP/TIP4PD
GB3	1.185	1.212	1.246	1.209	1.001	1.099
BPTI	1.078	1.241	1.088	1.171	1.173	1.205
CspTm	1.376[a]	1.674	1.718	1.663	1.000	1.123
Ubiquitin	1.231	1.506	1.249	1.769	1.003	1.509
SPR17	1.026	1.041	1.052	1.219	1.019	1.220

RESULTS AND DISCUSSION

CMAP Optimization. Ten cycles of CMAP optimization were performed for each amino acid. In the first cycle (CMAP0), the initial φ/ψ distribution was obtained from the standard *ff*03 force field, where the lowest RMSp is 0.234% among 20 amino acids. In contrast, after 10 cycles of optimization, the lowest RMSp is less than 0.064%, as shown in Figure S2. Comparison of the distributions of CMAP0 and the benchmark database, we found that there is almost no left-handed helix distribution, except for MET, GLY, and LEU. In addition, an obvious energy barrier exists between the β-sheet region and the α-helix region, hence it would be difficult to sample both types of structures. After optimization, these limitations are no longer present. The parameters for the best RMSp for each amino acid were selected as the final CMAP values. These parameters and structural factors were integrated with the standard *ff*03 force field to generate the new force field *ff*03*CMAP*.

Evaluation of *ff*03-Series Force Fields. We assessed the performance of *ff*03-series force fields in reproducing the experimental data. The same conditions were used in all MD simulations among all tested force fields. The FF scores for short peptides, IDPs, and fold proteins are shown in Figure 2 and specific values are listed in Supporting Information Table S3. Figure 2 suggests that the combination of *ff*03*CMAP*/*TIP4PD* agrees the best with experiments for short peptides and IDPs. In addition, the combination of *ff*03*CMAP*/*TIP4PEw* leads to the best agreement with experiment for folded proteins. In summary, the use of *ff*03*CMAP* can yield more accurate sampling of the conformers for short peptide, IDPs, and folded proteins.

Short Peptide Ala_5. Table 1 shows the RMSD's of secondary chemical shifts, *J*-coupling constants, and force-field score for Ala_5. There are 6 types of secondary chemical shifts and 7 types of *J*-coupling constants. For the CS score, the performance of *ff*03*CMAP*/*TIP4PEw* is much better than all other force fields. However, the performance of *ff*03*/*TIP3P* is the best for the NMR score. If we combine CS and NMR scores (i.e., overall FF score), *ff*03-driven force fields are significantly improved over the origin *ff*03 and the *ff*03*CMAP*/*TIP4PEw* is the best. The detail RMSD's of secondary chemical shifts and *J*-coupling constants are shown in Figures S3 and S4.

Folded Proteins. To evaluate the stability of folded proteins when modeled with *ff*03*CMAP*, five representative folded proteins were simulated: GB3 (α/β), BPTI (α/β), CspTm (all-β), ubiquitin (α/β), and SPR17 (all-α). The initial structures are extracted from PDB and the simulation time is 1 μs for each system.

Table 2 shows the FF scores of the five tested proteins for all *ff*03-series force fields. It is obvious that the FF score for the combination of *ff*03*CMAP*/*TIP4P-Ew* is the best among all tested force fields and the value close to 1, except for BPTI. This suggests that *ff*03*CMAP*/*TIP4PEw* indeed can be used to simulate folded proteins, and we also found that the original *ff*03 force field performs better than other revised *ff*03 force fields. It is no surprise that the performance of *ff*03*CMAP*/*TIP4PD* is a little worse than that of *ff*03*CMAP*/*TIP4PEw* because the TIP4P-D water model would destabilize the folded states of proteins as reported.[18] The details of the FF score composition for five proteins are shown in Tables S4–S8.

To quantify the fluctuation in simulations, Cα RMSD's of three folded proteins are shown in Figure 3. For CspTm and

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769–6780

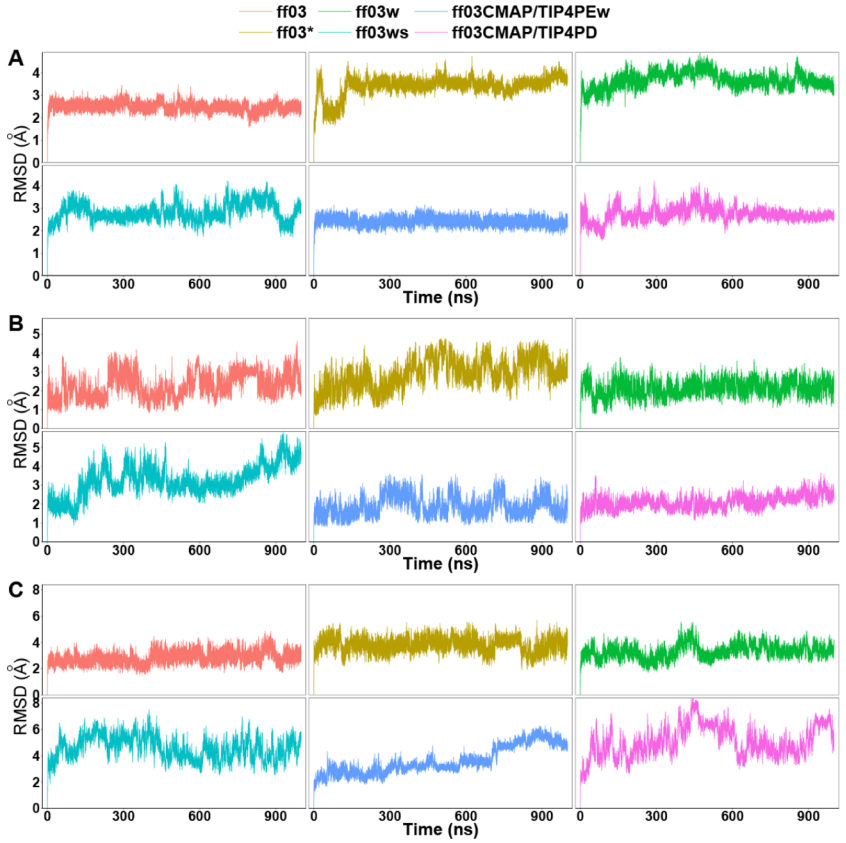


Figure 3. Cα RMSD of six *ff*03-series force fields. (A) CspTm, (B) ubiquitin, and (C) SPR17.

ubiquitin, the RMSDs in the *ff*03*CMAP*/*TIP4PEw* simulations are small and stable, which is consistent with the FF score. However, the RMSD's for SPR17 rise over 4 Å after 700 ns in the *ff*03*CMAP*/*TIP4PEw* simulation. The RMSDs of *ff*03*ws* for three folded proteins fluctuate quite significantly, implying less stable folded states for the tested proteins.

More detailed analyses were conducted for ubiquitin to compare the performances of these force fields. Secondary chemical shifts and backbone scalar coupling constants of ubiquitin (Figures S10 and S11) suggest that the *ff*03*CMAP*/*TIP4PEw* simulation agrees the best with the experimental data. The same can be said for the side-chain scalar coupling constants as shown in Tables S18 and S19. RDCs of backbone N−HN, Cα−Hα, Cα−C, C−N, and C−HN were also calculated as shown in Figure S12. Similar to chemical shifts and scalar coupling constants, the performance of the *ff*03*CMAP*/*TIP4PEw* simulation also agrees among the best, along with the *ff*03 simulation, while the *ff*03*ws* simulation agrees the worse. The order parameters for backbone amide and side-chain methyl groups are shown in Figure 4 and Table S20, respectively. The order parameters in the *ff*03*CMAP*/*TIP4PEw* simulation are also in good agreement with the experimental data. Figure 4 further indicates that the order parameters of the loop regions in the *ff*03*CMAP*/*TIP4PD* and *ff*03*ws* simulation are much lower than those from other force fields, which suggest that these two force fields overestimate disordered characters for this system, especially on some loop regions. Except for backbone order parameters, *ff*03*CMAP*/*TIP4PEw* exhibited similar behavior to experiment for side-chain order parameters (Table S20).

The RMSDs of the secondary chemical shift, *J*-coupling, order parameters, and RDCs are gathered in Table 3. The

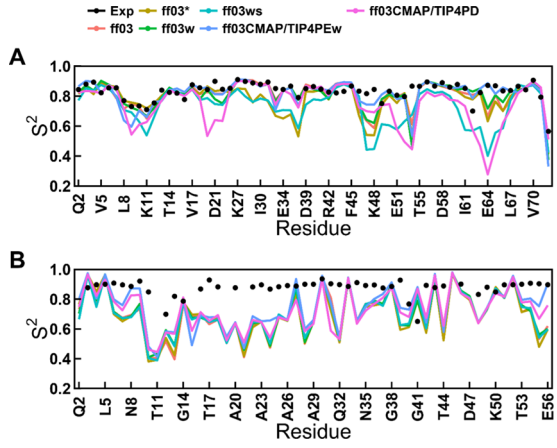


Figure 4. Backbone amide order parameters of six *ff*03-series force fields for ubiquitin (A) and GB3 (B). The RMSDs between the simulation and experiment for ubiquitin for *ff*03, *ff*03*, *ff*03*w*, *ff*03*ws*, *ff*03*CMAP*/*TIP4PEw*, and *ff*03*CMAP*/*TIP4PD* are 0.074, 0.104, 0.081, 0.149, 0.069, and 0.154. The RMSDs between the simulation and experiment for GB3 for *ff*03, *ff*03*, *ff*03*w*, *ff*03*ws*, *ff*03*CMAP*/*TIP4PEw*, and *ff*03*CMAP*/*TIP4PD* are 0.317, 0.324, 0.314, 0.311, 0.290, and 0.301.

summary indicates that *ff*03*CMAP*/*TIP4PEw* performs excellently in reproducing all available experimental measurements, and its FF score is very close to 1. However, the FF score of *ff*03*CMAP*/*TIP4PD* is more than 1.5, which suggests that the TIP4P-D water model is unsuitable for the simulation of folded proteins.

To further evaluate the stability of *ff*03*CMAP* for folded proteins, the dominant conformers of ubiquitin from six *ff*03-

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769−6780

Table 3. RMSDs of Secondary Chemical Shifts, *J*-Coupling Constants, RDCs, S^2 Parameters and FF Scores of Ubiquitin for Six *ff*03-Series Force Fields[a]

	ff03/TIP3P	ff03*/TIP3P	ff03w/TIP4P2005	ff03ws/TIP4P2005	ff03CMAP/TIP4PEw	ff03CMAP/TIP4PD
Cα	0.481	0.716	0.507	0.723	0.485	0.575
Cβ	0.816	0.885	0.825	1.069	0.770	0.860
C	0.607	0.702	0.611	0.789	0.609	0.650
N	2.050	2.450	2.188	2.625	2.142	2.573
HA	0.145	0.185	0.141	0.234	0.106	0.165
HN	0.329	0.349	0.326	0.407	0.298	0.293
$^3J_{HNH\alpha}$	1.341	1.487	1.484	1.597	1.196	1.319
$^3J_{H\alpha C}$	0.684	0.725	0.623	0.986	0.410	0.609
$^3J_{HNC\beta}$	0.643	0.734	0.671	0.701	0.633	0.612
$^3J_{HNC}$	0.681	0.734	0.701	0.852	0.564	0.662
$^1J_{C\alpha N}$	0.465	0.567	0.494	0.588	0.415	0.469
$^1J_{H\alpha C\alpha}$	2.669	2.717	2.628	3.030	2.225	2.476
$^1J_{C\alpha C\beta}$	0.984	1.056	0.977	1.229	0.856	0.966
$^2J_{C\alpha N}$	0.464	0.555	0.437	0.682	0.361	0.477
$^2J_{CC\gamma}$	0.650	0.973	0.717	0.936	0.419	0.775
$^3J_{NC\gamma}$	0.490	0.592	0.501	0.557	0.428	0.496
S_{NH}^2	0.074	0.104	0.081	0.149	0.069	0.154
S_{axis}^2	0.279	0.260	0.247	0.289	0.174	0.279
RDC	0.174	0.249	0.187	0.325	0.166	0.222
CS_{score}	1.092	1.321	1.107	1.512	1.012	1.200
backbone ^3J	1.225	1.359	1.228	1.562	1.004	1.182
side-chain ^3J	1.348	1.853	1.440	1.768	1.000	1.504
backbone S^2	1.070	1.510	1.175	2.156	1.000	2.226
side-chain S^2	1.603	1.493	1.418	1.657	1.000	1.603
backbone RDC	1.048	1.500	1.127	1.958	1.000	1.337
FF_{score}	1.231	1.506	1.249	1.769	1.003	1.509

[a]Chemical shifts are in ppm, *J*-coupling constants and RDC are in Hz, and the scores and S^2 parameters are unitless.

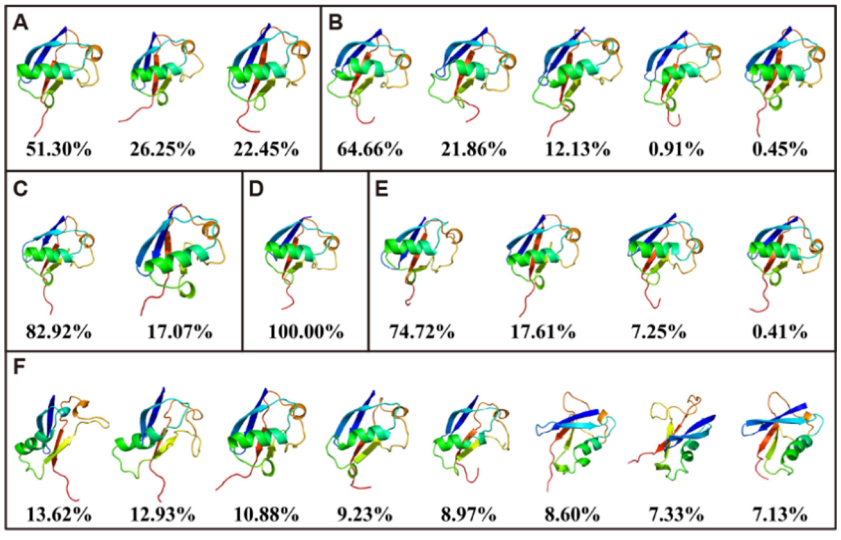


Figure 5. Conformation clustering of simulation for ubiquitin. The figure shows the top eight clusters at most with dominant conformations and percentage for *ff*03 (A), *ff*03* (B), *ff*03*w* (C), *ff*03*CMAP/TIP4PEw* (D), *ff*03*CMAP/TIP4PD* (E), and *ff*03*ws* (F).

series force fields are retrieved and shown in Figure 5. It was found that the top three clusters in the *ff*03 simulation occupy 100.00% of the snapshots. All of the conformers include a high percentage of helical structures. Top 5 clusters in the *ff*03* simulation also occupy 100.00% of the snapshots with partially nonhelical structures. In the *ff*03*w* simulation, only 2 clusters were found, and the conformers are highly structured. In the *ff*03*ws* simulation, the top 8 clusters only occupy 78.70% of the snapshots. In the *ff*03*CMAP/TIP4PEw* simulation, there is only one cluster and 4 clusters in the *ff*03*CMAP/TIP4PD* simulation. Additional conformation clustering was also conducted for the GB3, BPTI, CspTm, and SPR17 simulations (Supporting Information Figures S27−S30). These conformer clusters indicate that *ff*03*CMAP/TIP4PEw*, *ff*03, and *ff*03*w* may be the better *ff*03 choices for folded protein MD simulation.

Intrinsically Disordered Proteins. We tested 9 typical disordered proteins with 19−124 residues. The FF scores of

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769−6780

Table 4. FF Scores of Six *ff*03-Series Force Field for Nine Disordered Proteins

protein	ff03/TIP3P	ff03*/TIP3P	ff03w/TIP4P2005	ff03ws/TIP4P2005	ff03CMAP/TIP4PEw	ff03CMAP/TIP4PD
HEWL19	1.290[a]	1.039	1.089	1.117	1.125	1.092
RS	4.453	1.657	1.600	1.454	1.140	1.105
HIVRev	1.761	1.763	1.677	1.208	1.244	1.295
AB40	2.016	1.756	1.872	1.758	1.124	1.001
AB42	2.106	1.866	1.972	1.858	1.184	1.021
ACTR	1.618	1.474	1.354	1.172	1.245	1.033
IA3	2.111	1.810	1.923	1.677	1.404	1.000
p53N	1.668	1.441	1.537	1.351	1.133	1.019
tauF4	1.587	1.590	1.384	1.301	1.226	1.000

Table 5. Average of RMSDs for Experimental Observables of IDPs[a]

exp. type	ff03/TIP3P	ff03*/TIP3P	ff03w/TIP4P2005	ff03ws/TIP4P2005	ff03CMAP/TIP4PEw	ff03CMAP/TIP4PD
Cα	1.332	0.986	1.001	0.849	0.641	0.535
Cβ	0.545	0.607	0.559	0.566	0.461	0.405
C	0.853	0.745	0.649	0.613	0.553	0.544
N	2.545	2.153	2.275	1.958	1.575	1.208
Hα	0.146	0.132	0.137	0.126	0.118	0.114
HN	0.334	0.311	0.320	0.297	0.262	0.237
$^3J_{HNH\alpha}$	1.322	0.996	1.026	0.920	0.475	0.462
RDC	0.928	0.788	0.764	0.737	0.764	0.722

[a]The experimental measurements are calculated in four different IDP systems at least. Chemical shifts are in ppm, *J*-coupling constants and RDCs are in Hz.

the six *ff*03-series force fields are listed in Table 4. Except for HEWL19 and HIVRev, the FF scores of *ff*03*CMAP/TIP4P-D* simulations are the lowest, and most of them are very close to 1, indicating very good agreement with the experiment. Although the FF scores of *ff*03*CMAP/TIP4PD* for HEWL19 and HIVRev are not the best, the differences with the best performing force fields are not significant. This suggests that *ff*03*CMAP/TIP4P-D* can reproduce good conformers of tested IDPs. The details of the FF score composition for each tested system are shown in Tables S9−S17.

It is interesting to note that all *ff*03 revisions improve over the original *ff*03 in IDP simulations, as they are all designed to reproduce the properties of IDPs. As expected, the CMAP method can provide accurate descriptions of IDPs as in previous developments.[11−13,16,72] In addition, the TIP4P-D water model is demonstrated again to be suitable for IDP simulations.[18] To understand the influence of solvent models, three IDPs were simulated with *ff*03/*TIP4P-D* (Tables S12, S14, and S15). The results show that the TIP4P-D water model indeed partly improves the performance of the tested IDP conformers. However, the results are still much worse than those of *ff*03*CMAP/TIP4P-D*, suggesting that the CMAP improvement in *ff*03*CMAP/TIP4P-D* simulations plays a key role in reproducing the IDP conformers.

To further illustrate the properties of *ff*03*CMAP* force fields, we calculated the average RMSDs for different experimental measurements of IDPs (Table 5). We found that the performance of *ff*03*CMAP/TIP4PD* is the best for all experimental observables, and *ff*03*CMAP/TIP4PEw* also performs reasonably well. It is noticeable that *ff*03*CMAP* significantly improved the quality of simulated Cα and N secondary chemical shifts and $^3J_{HNH\alpha}$ scalar coupling constant, which are closely related to backbone dihedrals. It is clear that the CMAP method can be used to correct the dihedral distributions, and the *TIP4P-D* water model further refines the interactions between protein and water, leading to excellent observed performance in the *ff*03*CMAP/TIP4PEw* simulations for IDPs.

For IDP simulations, underestimation of R_g is a common limitation for generic protein force fields.[9,18] We calculated R_g distributions for all tested IDPs (Figure 6). The four force fields with four-site water models can sample a wider range of R_g distributions and larger R_g mean values, especially for *ff*03*CMAP/TIP4P-D* and *ff*03*ws*. Conformers are more compact in force fields in the TIP3P water model such as *ff*03 and *ff*03*. We compared the experimental R_g and simulated R_g of three IDPs. The analysis indicates that the *ff*03*CMAP* force field and *ff*03*ws* overestimate the R_g of Aβ40 and RS. The average R_g of *ff*03*CMAP/TIP4PEw* is also very small and close to the experimental value. For ACTR, only the R_g in the *ff*03*CMAP/TIP4PD* simulation is located within the experimental range and other force fields significantly underestimate it.

We also calculated residual helicity for ACTR, whose results are shown in Figure 7. The figure indicates that *ff*03*CMAP/TIP4PD* shows the best agreement with the experiment data among these force fields.

Besides the above overall assessments, we next use a classical example of IDPs, RS, in the following discussion to illustrate the performance of these force fields. To evaluate the backbone and side-chain sampling for RS, we compared secondary chemical shifts and scalar coupling constants. Figure 8 shows the secondary chemical shifts and backbone scalar coupling constants for six *ff*03-series force fields. Tables S21 and S22 list detailed data used in analysis. The chemical shifts and scalar coupling constants calculated from all revised *ff*03 force fields are much closer to the experimental data than those from the original *ff*03 force field, and the RMSDs between the simulated and experimental values from *ff*03*CMAP/TIP4P-Ew* and *ff*03*CMAP/TIP4P-D* combinations are smaller than those from the other *ff*03-series force fields (Table 4). We also calculated the backbone N−HN, Cα−Hα, and Cα−C RDCs

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769−6780

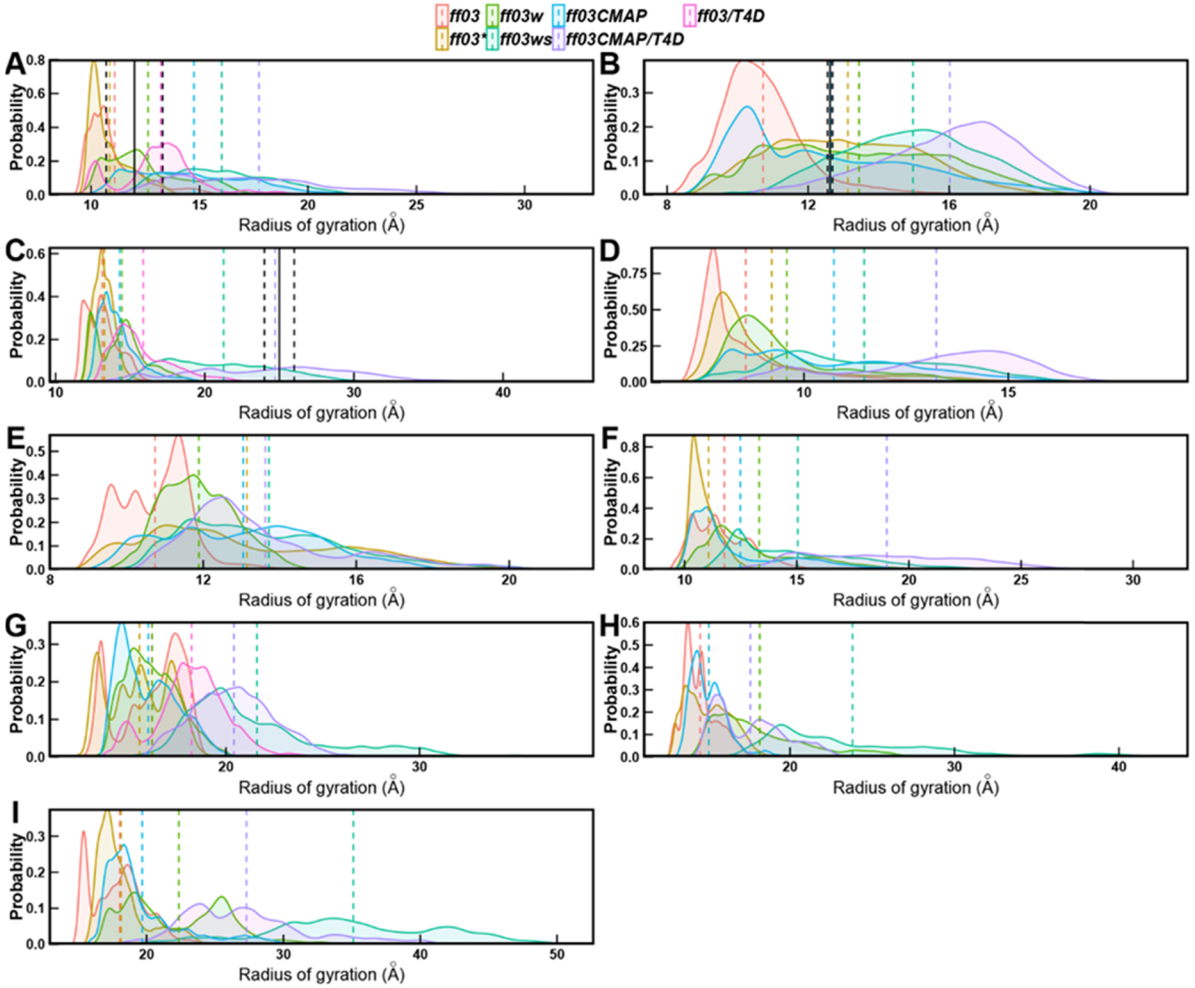


Figure 6. R_g distribution of the simulation and experimental data for disordered proteins. Simulated and experimental R_g distribution for (A) Aβ40, (B) RS, and (C) ACTR, and simulated R_g distribution for (D) HEWL19, (E) HIVRev, (F) Aβ42, (G) IA3, (H) p53N, and (I) tauF4. Mean R_g is shown in dash lines for ff03 (red), ff03* (brown), ff03w (light green), ff03ws (cyan), ff03CMAP (blue), ff03CMAP/T4D (violet), and ff03/T4D (mauve). Experimental values are displayed as black solid lines. The black dash lines mean the standard error of experimental values.

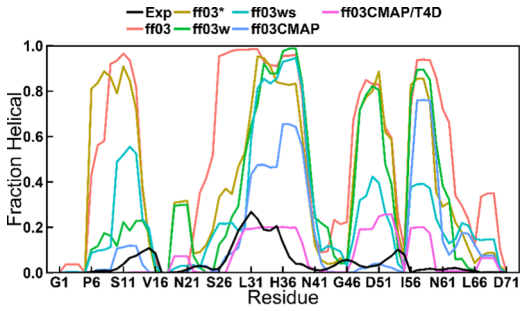


Figure 7. Helical propensities observed in simulations of ACTR. Experimental values are displayed as black lines predicted from experimental NMR chemical shifts using the program δ2d.[73]

with a local alignment window of 15 residues (Figure 9), whose performance is similar to that of chemical shifts and coupling constants, with both *ff*03*CMAP*/*TIP4PEw* and *ff*03*CMAP*/*TIP4PD* giving a lower *Q*-factor of RDCs (calculated by PALES[64]) with smaller standard deviations than the other tested force fields.

Next, FF scores were used to compare the performance of all force fields (listed in Table 6) in RS simulations. The RMSDs of the original *ff*03 force field are the largest and the FF score is larger than 4. The *ff*03*CMAP*/*TIP4PD* combination gives the best agreement with all experimental measurements and the FF scores are around 1.1.

Finally, we computed the ensemble-averaged SAXS curves for RS and fitted with the experimental curve (Figure 10). The χ^2 value was used to evaluate the quality of the fitted result to the given experimental SAXS profile as shown in the literature.[74] Our analysis shows that the χ^2 of *ff*03*CMAP*/*TIP4PEw* is the smallest among 6 tested *ff*03-series force fields. This suggests that *ff*03*CMAP*/*TIP4PEw* can reproduce the SAXS properties for RS, while the *ff*03*CMAP*/*TIP4PD* combination leads to conformers that are too expansive.

To further illustrate the conformer sampling efficiency, kClust was used to cluster conformers according to φ angle and Cα RMSD. Representative conformers and their occupations are shown in Figures S31−S39. The results indicate that both *ff*03*CMAP* and *ff*03*ws* can sample more flexible and diverse disordered conformers, while the representative conformers in the *ff*03 simulation contain several short helixes with tight packing. The convergence of conformer sampling is another important issue for IDP simulations. We used the biphasic decay model to evaluate the convergence time scales for IDP simulations. It is interesting to note that *ff*03*CMAP* simulations have smaller

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769−6780

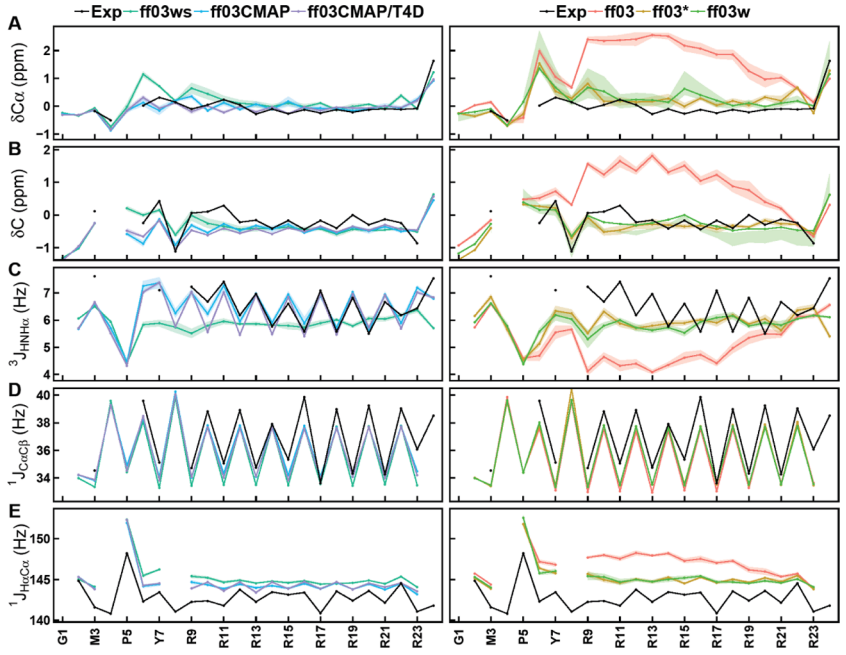


Figure 8. Secondary chemical shifts and backbone scalar coupling constants of simulation and experimental data for RS. Simulated and experimental secondary chemical shifts for (A) Cα and (B) C, backbone scalar coupling constants for (C) $^3J_{HNH\alpha}$, (D) $^1J_{C\alpha C\beta}$, and (E) $^1J_{H\alpha C\alpha}$. Simulated values are shown for *ff*03 (red), *ff*03* (brown), *ff*03*w* (light green), *ff*03*ws* (cyan), *ff*03*CMAP* (blue), and *ff*03*CMAP/T4D* (violet). Experimental values are displayed as black lines. The shadow means the standard error of mean.

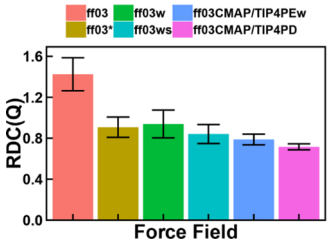


Figure 9. Backbone RDCs of simulation with six *ff*03-series force fields for RS. Error bar means the standard deviations of calculation.

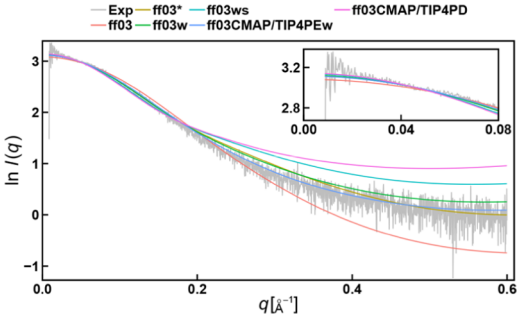


Figure 10. SAXS profiles of simulation and experimental data for RS. Experimental values are displayed as gray drawing. The χ^2 of *ff*03, *ff*03*, *ff*03*w*, *ff*03*ws*, *ff*03*CMAP/TIP4PEw*, and *ff*03*CMAP/TIP4PD* are 5.118, 2.217, 2.418, 10.708, 1.319, and 23.658.

decay half times, which suggests that *ff*03*CMAP* simulations converge earlier than other *ff*03-series simulations.

Ab Initio Folding of Fast-Folding Proteins. We performed REMD for three typical fast-folding proteins, such as 16-residue two β-sheets GB1, small β-hairpin-forming protein CLN025, and helical 15-mer AAQAA3.[36,38,39] The

Table 6. RMSD of Secondary Chemical Shifts, *J*-Coupling Constants, RDC and FF Score of RS for Six *ff*03-Series Force Fields[a]

	ff03/TIP3P	ff03*/TIP3P	ff03w/TIP4P2005	ff03ws/TIP4P2005	ff03CMAP/TIP4PEw	ff03CMAP/TIP4PD
Cα	1.813	0.502	0.484	0.385	0.239	0.193
C	1.199	0.333	0.296	0.307	0.354	0.373
$^3J_{HNH\alpha}$	1.777	0.987	0.974	0.981	0.407	0.491
$^1J_{C\alpha C\beta}$	1.625	1.407	1.420	1.438	1.084	1.195
$^1J_{H\alpha C\alpha}$	4.190	2.462	2.511	2.395	1.774	1.824
$^3J_{CC\gamma}$	0.302	0.314	0.271	0.278	0.330	0.311
$^3J_{NC\gamma}$	0.386	0.254	0.272	0.233	0.230	0.217
RDC	1.425	0.908	0.939	0.841	0.788	0.716
CS_{score}	6.720	1.862	1.753	1.514	1.217	1.128
NMR_{score}	2.186	1.451	1.448	1.394	1.063	1.081
FF_{score}	4.453	1.657	1.600	1.454	1.140	1.105

[a]Chemical shifts are in ppm, *J*-coupling constants, and RDCs are in Hz.

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769−6780

melting curves in the *ff*03*CMAP*/*TIP4PEw* simulations are shown in Figure 11. The melting curves show that GB1 and

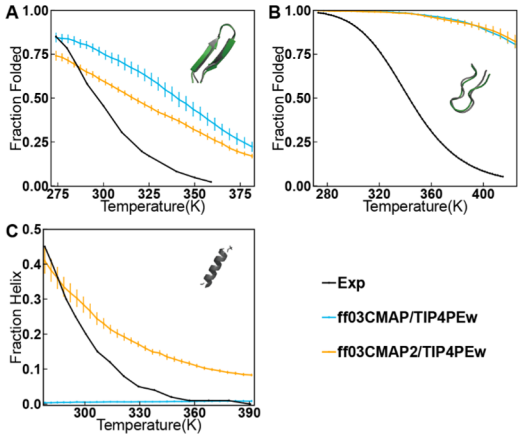


Figure 11. Melting curves for *ff*03*CMAP*/*TIP4PEw* and *ff*03*CMAP*2/*TIP4PEw* with experimental values. (A) Temperature-dependent folded population of GB1. (B) Temperature-dependent folded population of CLN025. (C) Temperature-dependent helix formation of AAQAA3 (the definition of helicity calculation is mention in the SI). Simulated values shown for *ff*03*CMAP*/*TIP4PEw* are in blue lines, *ff*03*CMAP*2/*TIP4PEw* are in orange line, and experimental melting curves are shown in black lines. All simulations initiated from fully unfolded structures and 100 ns (for CLN025 and AAQAA3) and 300 ns (for GB1) for equilibration. Folded structures obtained from REMD simulations (gray) starting from the completely unfolded state are compared with the native structure from the PDB database (green), and the RMSD between folded structure and native structure is 0.728 Å for GB1 and 0.993 Å for CLN025.

CLN025 can be ab initio folded when modeled with *ff*03*CMAP*/*TIP4PEw*. However, a few folded structures were observed in the REMD simulation of AAQAA3. To study whether the ab initio folding of helical structures can be improved by modifying CMAP parameters, we updated a new set of CMAP parameters by only decreasing the parameters in the α_h region with the revised force field termed as *ff*03*CMAP*2. Our REMD simulation shows that *ff*03*CMAP*2/*TIP4PEw* performs significantly better in helix folding. In the meantime, *ff*03*CMAP*2/*TIP4PEw* can maintain almost the same melting curves for sheet and hairpin fast-folding proteins. For IDP systems, *ff*03*CMAP*2/*TIP4PEw* performs slightly worse compared with the previous version but still better than other ff03-driven force fields.

CONCLUSIONS

The backbone dihedral term for all 20 amino acids was optimized to improve the performance of the current force field. TIP4P-Ew and TIP4P-D are combined with the newly developed force field *ff*03*CMAP* to simulate different types of proteins. Extensive tests of the typical short peptide, folded proteins, disordered proteins, and fast-folding proteins show that the simulated chemical shifts, *J*-coupling, order parameters, and RDC with the *ff*03*CMAP* force fields are in quantitative agreement with those from the NMR experiment and are more accurate than other *ff*03-series force fields. The influences of solvent models were also investigated. The results indicate that *ff*03*CMAP*/*TIP4PEw* for folded proteins and *ff*03*CMAP* combined with TIP4P-D was suitable for disordered proteins (*ff*03*CMAP*/*TIP4PEw* also shows good performance in IDPs). Therefore, these findings confirm that the newly developed force field *ff*03*CMAP* can improve the balance and efficiency of conformer sampling between intrinsically folded proteins and disordered proteins. Although the *ff*03*CMAP* force field has the limitation of folding helix structures, this can be improved by adjusting CMAP parameters, which is the *ff*03*CMAP*2.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00623.

Half-time convergence and definition of AAQAA3 helicity are listed in the supplementary file; simulation conditions for all protein or peptide systems tested in this research (Table S1); experimental measurements used in this work (Table S2); the FF score of six ff03-series force fields for 15 tested systems (Table S3); detailed comparison between the simulation and experimental data for each force field for each system we tested (Tables S4−S17); side-chain calculations (coupling constants and order parameters) of ubiquitin and RS (Tables S18−S22); average half-time of IDP simulation for six ff03-series force fields (Table S23); count of amino acids in coil database (Figure S1); RMSp between simulation and benchmark for CMAP optimization of 20 amino acids (Figure S2); the detailed information of secondary chemical shifts, coupling constants, and RDC for each force field for each system we tested (Figures S3−S26); conformation clustering of simulation for each force field for each system we tested (Figure S27−S39); and the biphasic exponential decay model fitting for disordered proteins (Figure S40) (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: haifengchen@sjtu.edu.cn (H.-F.C.).

*E-mail: ray.luo@rayluolab.org. Tel/Fax: +86-21-34204348 (R.L.).

ORCID

Ray Luo: 0000-0002-6346-8271

Hai-Feng Chen: 0000-0002-7496-4182

Author Contributions

‖Y.Z. and H.L. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by Center for HPC at Shanghai Jiao Tong University, the National Natural Science Foundation of China (31770771, 21977068, and 31620103901), the National Key Research and Development Program of China (2018YFC0310803 and 2017YFE0103300), Medical Engineering Cross Fund of Shanghai Jiao Tong University (YG2017MS08), and the National Institutes of Health/NIGMS (GM093040 & GM079383).

REFERENCES

(1) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.;

Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically disordered protein. *J. Mol. Graphics Modell.* **2001**, *19*, 26−59.
(2) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry* **2002**, *41*, 6573−6582.
(3) Babu, M. M. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **2016**, *44*, 1185−1200.
(4) Cheng, Y.; LeGall, T.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* **2006**, *45*, 10448−10460.
(5) Oldfield, C. J.; Dunker, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **2014**, *83*, 553−584.
(6) Bernadó, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. *J. Am. Chem. Soc.* **2005**, *127*, 17968−17969.
(7) Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **2007**, *129*, 5656−5664.
(8) Schneider, R.; Huang, J. R.; Yao, M.; Communie, G.; Ozenne, V.; Mollica, L.; Salmon, L.; Jensen, M. R.; Blackledge, M. Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol. BioSyst.* **2012**, *8*, 58−68.
(9) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E4758−E4766.
(10) Wang, W.; Ye, W.; Jiang, C.; Luo, R.; Chen, H. F. New force field on modeling intrinsically disordered proteins. *Chem. Biol. Drug Des.* **2014**, *84*, 253−269.
(11) Ye, W.; Ji, D.; Wang, W.; Luo, R.; Chen, H. F. Test and Evaluation of ff99IDPs Force Field for Intrinsically Disordered Proteins. *J. Chem. Inf. Model.* **2015**, *55*, 1021−1029.
(12) Song, D.; Wang, W.; Ye, W.; Ji, D.; Luo, R.; Chen, H. F. ff14IDPs force field improving the conformation sampling of intrinsically disordered proteins. *Chem. Biol. Drug Des.* **2017**, *89*, 5−15.
(13) Song, D.; Luo, R.; Chen, H. F. The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. *J. Chem. Inf. Model.* **2017**, *57*, 1166−1178.
(14) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.
(15) Zhou, C. Y.; Jiang, F.; Wu, Y. D. Residue-specific force field based on protein coil library. RSFF2: modification of AMBER ff99SB. *J. Phys. Chem. B* **2015**, *119*, 1035−1047.
(16) Liu, H.; Song, D.; Lu, H.; Luo, R.; Chen, H. F. Intrinsically disordered protein-specific force field CHARMM36IDPSFF. *Chem. Biol. Drug Des.* **2018**, *92*, 1722−1735.
(17) Liu, H.; Song, D.; Zhang, Y. P.; Yang, S.; Luo, R. A.; Chen, H. F. Extensive tests and evaluation of the CHARMM36IDPSFF force field for intrinsically disordered proteins and folded proteins. *Phys. Chem. Chem. Phys.* **2019**, *21*, 21918−21931.
(18) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(19) 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. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(20) Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(21) Best, R. B.; Mittal, J. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **2010**, *114*, 14916−14923.
(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. A.* **1983**, *79*, 926−935.
(23) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(24) Abascal, J. L.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, No. 234505.
(25) Welker, C.; Bohm, G.; Schurig, H.; Jaenicke, R. Cloning, overexpression, purification, and physicochemical characterization of a cold shock protein homolog from the hyperthermophilic bacterium *Thermotoga maritima*. *Protein Sci.* **1999**, *8*, 394−403.
(26) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* **1987**, *194*, 531−544.
(27) Bennett, V.; Gilligan, D. M. The spectrin-based membrane skeleton and micron-scale organization of the plasma membrane. *Annu. Rev. Cell Biol.* **1993**, *9*, 27−66.
(28) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. *J. Am. Chem. Soc.* **2007**, *129*, 1179−1189.
(29) Xiang, S.; Gapsys, V.; Kim, H. Y.; Bessonov, S.; Hsiao, H. H.; Mohlmann, S.; Klaukien, V.; Ficner, R.; Becker, S.; Urlaub, H.; Luhrmann, R.; de Groot, B.; Zweckstetter, M. Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure* **2013**, *21*, 2162−2174.
(30) Tan, R.; Chen, L.; Buettner, J. A.; Hudson, D.; Frankel, A. D. RNA recognition by an isolated α helix. *Cell* **1993**, *73*, 1031−1040.
(31) Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E. The Alzheimer's peptides Abeta40 and 42 adopt distinct conformations in water: a combined MD / NMR study. *J. Mol. Biol.* **2007**, *368*, 1448−1457.
(32) Iesmantavicius, V.; Jensen, M. R.; Ozenne, V.; Blackledge, M.; Poulsen, F. M.; Kjaergaard, M. Modulation of the intrinsic helix propensity of an intrinsically disordered protein reveals long-range helix-helix interactions. *J. Am. Chem. Soc.* **2013**, *135*, 10155−10163.
(33) Li, M.; Phylip, L. H.; Lees, W. E.; Winther, J. R.; Dunn, B. M.; Wlodawer, A.; Kay, J.; Gustchina, A. The aspartic proteinase from *Saccharomyces cerevisiae* folds its own inhibitor into a helix. *Nat. Struct. Biol.* **2000**, *7*, 113−117.
(34) Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5762−5767.
(35) Sillen, A.; Barbier, P.; Landrieu, I.; Lefebvre, S.; Wieruszeski, J. M.; Leroy, A.; Peyrot, V.; Lippens, G. NMR investigation of the interaction between the neuronal protein tau and the microtubules. *Biochemistry* **2007**, *46*, 3055−3064.
(36) Shalongo, W.; Dugad, L.; Stellwagen, E. Distribution of helicity within the model peptide acetyl (AAQAA) 3amide. *J. Am. Chem. Soc.* **1994**, *116*, 8288−8293.
(37) Blanco, F. J.; Rivas, G.; Serrano, L. A short linear peptide that folds into a native stable β-hairpin in aqueous solution. *Nat. Struct. Mol. Biol.* **1994**, *1*, 584−590.
(38) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding dynamics and mechanism of β-hairpin formation. *Nature* **1997**, *390*, 196−199.
(39) Honda, S.; Akiba, T.; Kato, Y. S.; Sawada, Y.; Sekijima, M.; Ishimura, M.; Ooishi, A.; Watanabe, H.; Odahara, T.; Harata, K. Crystal structure of a ten-amino acid protein. *J. Am. Chem. Soc.* **2008**, *130*, 15327−15331.

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769−6780

(40) Case, D. A.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham Iii, T.; Darden, T.; Duke, R.; Gohlke, H. *Amber 14*, 2014.
(41) Lee, T. S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58*, 2043−2050.
(42) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. 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.
(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) Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.
(45) Patriksson, A.; van der Spoel, D. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073−2077.
(46) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.
(47) Joosten, R. P.; Te Beek, T. A.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W.; Schneider, R.; Sander, C.; Vriend, G. A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **2011**, *39*, D411−D419.
(48) Wang, W.; Ye, W.; Jiang, C.; Luo, R.; Chen, H. F. New force field on modeling intrinsically disordered proteins. *Chem. Biol. Drug Des.* **2014**, *84*, 253−269.
(49) 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.; 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. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(50) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(51) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(52) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.
(53) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **2011**, *50*, 43−57.
(54) Vögeli, B.; Ying, J.; Grishaev, A.; Bax, A. Limits on variations in protein backbone dynamics from precise measurements of scalar couplings. *J. Am. Chem. Soc.* **2007**, *129*, 9377−9385.
(55) Hu, J.-S.; Bax, A. Determination of ϕ and χ1 Angles in Proteins from 13C− 13C Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond? *J. Am. Chem. Soc.* **1997**, *119*, 6360−6368.
(56) Hennig, M.; Bermel, W.; Schwalbe, H.; Griesinger, C. Determination of ψ torsion angle restraints from 3 J (Cα, Cα) and 3 J (Cα, HN) coupling constants in proteins. *J. Am. Chem. Soc.* **2000**, *122*, 6268−6277.
(57) Lee, J. H.; Li, F.; Grishaev, A.; Bax, A. Quantitative residue-specific protein backbone torsion angle dynamics from concerted measurement of 3J couplings. *J. Am. Chem. Soc.* **2015**, *137*, 1432−1435.
(58) Ding, K.; Gronenborn, A. M. Protein Backbone 1H(N)-13Calpha and 15N-13Calpha residual dipolar and J couplings: new constraints for NMR structure determination. *J. Am. Chem. Soc.* **2004**, *126*, 6232−6233.
(59) Wirmer, J.; Schwalbe, H. Angular dependence of 1J-(Ni,Calphai) and 2J(Ni,Calpha(i-1)) coupling constants measured in J-modulated HSQCs. *J. Biomol. NMR* **2002**, *23*, 47−55.
(60) Mantsyzov, A. B.; Maltsev, A. S.; Ying, J.; Shen, Y.; Hummer, G.; Bax, A. A maximum entropy approach to the study of residue-specific backbone angle distributions in alpha-synuclein, an intrinsically disordered protein. *Protein Sci.* **2014**, *23*, 1275−1290.
(61) Vuister, G. W.; Delaglio, F.; Bax, A. The use of 1 J CαHα coupling constants as a probe for protein backbone conformation. *J. Biomol. NMR* **1993**, *3*, 67−80.
(62) Schmidt, J. M.; Howard, M. J.; Maestre-Martinez, M.; Perez, C. S.; Lohr, F. Variation in protein C(alpha)-related one-bond J couplings. *Magn. Reson. Chem.* **2009**, *47*, 16−30.
(63) Schmidt, J. M. Asymmetric Karplus curves for the protein side-chain 3J couplings. *J. Biomol. NMR* **2007**, *37*, 287−301.
(64) Zweckstetter, M. NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.* **2008**, *3*, 679−690.
(65) Marsh, J. A.; Baker, J. M.; Tollinger, M.; Forman-Kay, J. D. Calculation of residual dipolar couplings from disordered state ensembles using local alignment. *J. Am. Chem. Soc.* **2008**, *130*, 7804−7805.
(66) Trbovic, N.; Kim, B.; Friesner, R. A.; Palmer, A. G., 3rd Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation. *Proteins* **2008**, *71*, 684−694.
(67) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **2013**, *105*, 962−974.
(68) Feig, M.; Karanicolas, J.; Brooks, C. L., 3rd MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graphics Modell.* **2004**, *22*, 377−395.
(69) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernandez, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.
(70) *The PyMOL Molecular Graphics System*, version 1.8.; Schrodinger, LLC, 2015.
(71) Duong, V. T.; Chen, Z.; Thapa, M. T.; Luo, R. Computational Studies of Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2018**, *122*, 10455−10469.
(72) Guo, X.; Han, J.; Luo, R.; Chen, H. F. Conformation Dynamics of the Intrinsically Disordered Protein c-Myb with the ff99IDPs Force Field. *RSC Adv.* **2017**, *7*, 29713−29721.
(73) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry* **2012**, *51*, 2224−2231.
(74) Schneidman-Duhovny, D.; Hammel, M.; Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **2010**, *38*, W540−W544.

DOI: 10.1021/acs.jctc.9b00623
J. Chem. Theory Comput. 2019, 15, 6769−6780