JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Test and Evaluation of *ff99IDPs* Force Field for Intrinsically Disordered Proteins

Wei Ye,[†] Dingjue Ji,[†] Wei Wang,[†] Ray Luo,*[,‡] and Hai-Feng Chen*[,†,§]

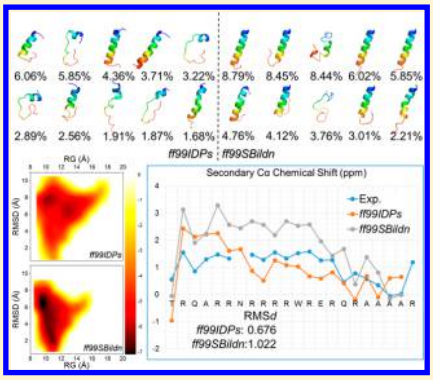
[†]State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, College of Life Sciences and Biotechnology, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai 200240, China

[‡]Department of Molecular Biology and Biochemistry, Department of Chemical Engineering and Materials Science, and Department of Biomedical Engineering, University of California, Irvine, California 92697-3900, United States

[§]Shanghai Center for Bioinformation Technology, 1278 Keyuan Road, Shanghai 200235, China

Supporting Information

ABSTRACT: Over 40% of eukaryotic proteomic sequences have been predicted to be intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) and confirmed to be associated with many diseases. However, widely used force fields cannot well reproduce the conformers of IDPs. Previously the *ff99IDPs* force field was released to simulate IDPs with CMAP energy corrections for the eight disorder-promoting residues. In order to further confirm the performance of *ff99IDPs*, three representative IDP systems (arginine-rich HIV-1 Rev, aspartic proteinase inhibitor IA_3, and α-synuclein) were used to test and evaluate the simulation results. The results show that for free disordered proteins, the chemical shifts from the *ff99IDPs* simulations are in quantitative agreement with those from reported NMR measurements and better than those from *ff99SBildn*. Thus, *ff99IDPs* can sample more clusters of disordered conformers than *ff99SBildn*. For structural proteins, both *ff99IDPs* and *ff99SBildn* can well reproduce the conformations. In general, *ff99IDPs* can successfully be used to simulate the conformations of IDPs and IDRs in both bound and free states. However, relative errors could still be found at the boundaries of ordered residues scattered in long disorder-promoting sequences. Therefore, polarizable force fields might be one of the possible ways to further improve the performance on IDPs.



INTRODUCTION

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) in structured proteins have been confirmed to play key roles in the biological function of proteins. Although they cannot automatically fold into stable and ordered structures, the structural flexibilities still have important functions in many cellular processes, such as molecular recognition, assembly, and post-translational modifications.[1−5] Over 40% of eukaryotic proteomic sequences have been predicted to be disordered with more than 40 consecutive amino acids, indicating that IDPs are widespread.[6,7] Recently, IDPs have been intensively studied because of their association with many diseases, such as cancer,[8] cardiovascular diseases,[9] and neurodegenerative diseases.[10−14] Thus, IDPs obviously provide exciting opportunities and bring huge challenges to the understanding of the protein structure−function paradigm.

Interestingly, IDPs can rapidly fold into structural proteins upon binding with their partners. Typical observations have been found in tumor suppressor p53, which acts as a central hub in multiple signaling pathways[8,15−17] through interactions with hundreds of partners. Since significant conformational adjustments occur as these proteins go from disordered to ordered structures, essential challenges are also brought to molecular dynamics (MD) simulations on IDPs.

Indeed, MD simulations have become a powerful tool to investigate the dynamic motions in macromolecules on long time scales (approximately nanosecond to microsecond).[18−22] However, the accuracy of the force field remains an issue in applications of MD simulations.[23,24] In the case of IDPs, the issue is that most protein force fields are often too stable to model the unstructured proteins on the basis of our tests. To overcome this issue, we developed an AMBER[25] force field, *ff99IDPs*,[26] with the addition of grid-based energy correction maps (CMAPs)[27,28] on the standard *ff99SBildn* force field[29,30] to reproduce the main-chain torsional distributions of eight disorder-promoting residues (A, G, P, R, Q, S, E, and K) that have been reported in the literature.[31] Tests on the apo and bound states of measles virus nucleoprotein (MeV N_{TAIL}) and p53 showed that *ff99IDPs* can reproduce IDPs better than the widely used *ff99SBildn*. In addition, the rigid secondary structures in the bound state could be well-maintained under *ff99IDPs*.

In this study, three IDPs were used to test and evaluate the performance of *ff99IDPs* in both the disordered apo state and the ordered bound state. Arginine-rich HIV-1 Rev is an RNA-binding protein and regulates the HIV-1 replication cycle.[32−34] Another test was conducted on IA_3, the aspartic proteinase

Received: January 26, 2015
Published: April 28, 2015

DOI: 10.1021/acs.jcim.5b00043
J. Chem. Inf. Model. 2015, 55, 1021−1029

inhibitor. Expression of aspartic proteinase has been confirmed in the life cycles of many human infectious agents.[35−37] Among its few natural inhibitors, IA_3 is a typical IDP that can fold into a highly rigid helix upon binding with aspartic proteinase.[38,39] The third test object was α-synuclein. It is an IDP on the cell membrane in substantia nigra pars compacta (SNc), which is crucial in the pathogenesis of Parkinson's disease. It may misfold into highly ordered cross-β fibrils and form Lewis bodies.[18,40−42] Our tests show that *ff99IDPs* can reproduce highly disordered conformations of three tested IDPs and that the predicted chemical shifts are consistent with the experimental observations. Furthermore, the performance of *ff99IDPs* is better than that of *ff99SBildn* in sampling the conformers of IDPs. For the bound state, both force fields reproduced the conformers with stable secondary structures.

As a reference, we also tested the applicability of *ff99IDPs* on the structural proteins lysozyme and ubiquitin.[30] Our data show that the newly developed force field behaves similarly to *ff99SBildn* on the tested structural proteins. However, our test data show that disagreement with experiments still exist for the ordered residues when they are scattered in long disorder-promoting sequences. The possible causes are discussed, and remedies with the deployment of polarizable force fields are also proposed.

MATERIALS AND METHODS

Selection of Test Systems. In order to quantitatively validate the performance of the newly developed AMBER force field *ff99IDPs*, intrinsically disordered proteins were searched in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and the Biological Magnetic Resonance Data Bank (BMRB)[43] with the following criteria: (1) the protein is exactly described as an IDP and (2) thermodynamic data, such as chemical shift and order parameter (S^2), are available. Under these criteria, three representative systems were selected to evaluate the performance of *ff99IDPs*: arginine-rich motif (ARM) of HIV-1 Rev (termed HIVRev; PDB code 1ETF),[44] the aspartic proteinase inhibitor IA_3 (termed IA_3; PDB code 1DP5),[39] and the disordered region of micelle-bound α-synuclein (termed αSyn; PDB code 2KKW),[45]. HIVRev has a very high proportion of disorder-promoting residues, with only four order-promoting ones out of 21 residues. As a distinct contrast, IA_3 has a high proportion of order-promoting residues, with 14 order-promoting residues in the 31-mer polypeptide. Both HIVRev and IA_3 can fold into rigid α-helices in the bound state. αSyn is another typical IDP that consists of a long structured region and a long disordered loop. For HIVRev and IA_3, both the free disordered state and the bound ordered state were tested. Furthermore, we also tested two widely used proteins for the validation of previous force fields,[30] hen egg white lysozyme (HEWL; PDB code 6LYT)[46] and ubiquitin (PDB code 1UBQ),[47] to check whether the IDP-specific force field could be used for structural proteins. All of the simulations were done under *ff99IDPs* or *ff99SBildn*.

Overview of *ff99IDPs* Implementation. The total energy of the IDP-specific force field *ff99IDPs* is the sum of the *ff99SBildn* total energy and the dihedral energy correction term with parameters for the eight disorder-promoting residues, as shown in eq 1:[26]

$$E_{ff99IDPs} = E_{ff99SBildn} + E_{\mathrm{CMAP}} \tag{1}$$

All of the tested structures with PDB file format were first converted into AMBER topology files using the *ff99SBildn* force field. The CMAP parameters were then added using an in-house PERL script.[26]

All of the energy terms of *ff99IDPs* except the dihedral energy term remain the same as a chosen base additive force field of *ff99SBildn*. Furthermore, only the backbone dihedral parameters for the eight disorder-promoting residues were optimized, while the parameters for the other 12 residues were kept the same to minimize the perturbation of folded structure distributions.

CMAP is a matrix of corrections on dihedral grids, with the corrections between grid points calculated using a two-dimensional bicubic interpolation method.[28] The correction matrix for each residue was set up with a dihedral angle grid having a resolution of 15°. Specifically, we used relative conformational free energies ($\Delta G_{i,j}$) converted from φ/ψ distributions for the disordered protein structures to compute the correction matrix according to eq 2:

$$\Delta G_{i,j} = -RT\ \ln(N_{i,j}/N_0) \tag{2}$$

where $N_{i,j}$ is the population of φ/ψ dihedral bin (i, j), N_0 is the population of the most-populated bin, R is the gas constant, and T is the absolute temperature. In this equation, sparsely populated bins could have huge relative free energies, leading to overcorrection. To overcome this limitation, we used an iterative optimization process to determine the CMAP correction matrix self-consistently. Here the CMAP energy terms were calculated at each iteration step using eq 3:

$$E_{i,j}^{\mathrm{CMAP}} = \Delta G_{i,j}^{\mathrm{DB}} - \Delta G_{i,j}^{\mathrm{MD}} \tag{3}$$

where $\Delta G_{i,j}^{\mathrm{DB}}$ and $\Delta G_{i,j}^{\mathrm{MD}}$ are the database and MD simulation converted free energies for φ/ψ dihedral bin (i, j), respectively. The iteration started with a CMAP correction matrix initialized as zero, so the initial $\Delta G_{i,j}^{\mathrm{MD}}$ values were derived from the simulations in the base additive force field *ff99SBildn*. At each iteration step, the CMAP correction matrix derived from the previous step's simulation was added to the base force field *ff99SBildn*. Root-mean-square deviations of population (termed RMSp) among all bins were calculated to quantitatively measure the difference between the MD and database populations until RMSp was less than 0.15%.

Molecular Dynamics Simulations. All of the chosen structures were first minimized in SYBYL-X 2.1.1[48] to eliminate any possible overlaps or clashes. All of the simulations and most of the analysis procedures were performed using the AMBER12 software package.[25] Hydrogen atoms were added using the LEaP module of AMBER12. Counterions were used to maintain system neutrality. All of the systems were solvated in a truncated octahedral box of TIP3P waters with a buffer of 10 Å. Particle mesh Ewald (PME)[49] was employed to treat long-range electrostatic interactions with the default setting in AMBER12. The newly developed *ff99IDPs* was added as described in the previous literature[26] based on *ff99SBildn*, which was also taken as the benchmark to compare with the experimental data. All of the MD simulations were accelerated with the CUDA version of PMEMD[50,51] in NVIDIA Tesla K20 GPU cores. The SHAKE algorithm[52] was used to constrain bonds involving hydrogen atoms. Up to 20000 steps of steepest-descent minimization were performed to relieve any structural clash in the solvated systems. This was followed by a 400 ps heating and a 200 ps equilibration in the *NVT* ensemble at 298 K with PMEMD of AMBER12. Langevin dynamics with a time step of 2 fs was used in the heating and equilibration runs with a friction constant of 1 ps^{-1}.

DOI: 10.1021/acs.jcim.5b00043
J. Chem. Inf. Model. 2015, 55, 1021−1029

Table 1. Simulation Conditions for All of the Models

system	force field	no. of traj.	simulation time (ns)	ions	waters	BMRB accession no.
apo-HIVRev	*ff99IDPs*	5	130	9 Cl^-	4409	18851
	ff99SBildn	5	130			
bound HIVRev	*ff99IDPs*	5	100	23 Na^+	12480	18852
	ff99SBildn	5	100			
apo-IA_3	*ff99IDPs*	5	100	1 Na^+	6434	6078
	ff99SBildn	5	100			
bound IA_3	*ff99IDPs*	3	100	24 Na^+	11781	–
	ff99SBildn	3	100			
αSyn	*ff99IDPs*	5	100	9 Na^+	47408	16302
	ff99SBildn	5	100			
HEWL	*ff99IDPs*	5	100	8 Cl^-	5799	4562, 18304 (S^2)
	ff99SBildn	5	100			
ubiquitin	*ff99IDPs*	5	100	–	4669	5387, 6470 (S^2)
	ff99SBildn	5	100			

To evaluate the performance of *ff99IDPs* and compare it with *ff99SBildn*, five independent trajectories each were simulated for apo- and bound HIVRev, apo-IA_3, and apo-αSyn under *ff99IDPs* and *ff99SBildn*; for bound IA_3, three trajectories were simulated. The simulation time for apo-HIVRev was 130 ns, and the simulation times for the other systems were 100 ns. To show the compatibility of *ff99IDPs* on normal proteins, five 100 ns trajectories were also conducted on HEWL and ubiquitin under both force fields. A total of 6.9 μs of simulations was collected at 298 K, taking about 3500 GPU hours. Detailed simulation conditions are listed in Table 1.

Data Analyses. Root-mean-square deviations (RMSDs) and fluctuations (RMSFs) in MD trajectories were calculated with the PTRAJ module in AMBER12 and AmberTools13.[25] Structural clustering analysis was conducted with the kclust program in the MMTSB toolset[53] on the basis of RMSD with a 5.5 Å cutoff. Cluster numbers of conformers within every 5 ns period along all the IDP trajectories under *ff99IDPs* and *ff99SBildn* were plotted to test whether both force fields can converge to sample the disordered conformers. Secondary structures of all of the snapshots were identified with the Dictionary of Secondary Structure of Proteins (DSSP) algorithm.[54,55] STRIDE[56,57] was also used for HIVRev to compare with DSSP. Experimental Cα chemical shift data for all of the models and N−H order parameters (S^2) for HEWL and ubiquitin were retrieved from the BMRB.[43] The accession numbers are shown in Table 1. Potential of mean force (PMF) free energy landscapes were mapped by calculating normalized probabilities from a histogram analysis and were plotted with Origin 8.5. For each simulation, sampling was conducted every 50 ps (10000 snapshots for five 100 ns simulations). Radius of gyration (RG) and RMSD were both separated into eight bins. The energy landscape was plotted among these 64 (8 × 8) bins. The secondary chemical shift data for the simulated structures were calculated with SPARTA version 1.01.[58] N−H *J*-coupling data of free and bound HIVRev were calculated using the Karplus equation.[59−61] All of the structural visualizations were generated using PyMOL 1.7.[62]

RESULTS AND DISCUSSION

We performed MD simulations on the models mentioned above. Then evaluations were performed through conformation sampling, structural clustering, helicity evolution, secondary Cα chemical shift, and other structural or thermodynamic indices. For each system, the RMSD and Cα fluctuation (RMSF) were first calculated to ensure that *ff99IDPs* could sample rational conformations. In comparison with the initial structures, the RMSD of each trajectory under both *ff99IDPs* and *ff99SBildn* (Figure S1 in the Supporting Information) indicates that 130 ns simulations are sufficient to achieve equilibration of the dynamics at room temperature for apo-HIVRev and that 100 ns simulations are sufficient for the other systems. Furthermore, the RMSDs of free IDPs under *ff99IDPs* were mostly higher than those under *ff99SBildn*. This indicates that *ff99IDPs* might sample more flexible conformations than *ff99SBildn*.

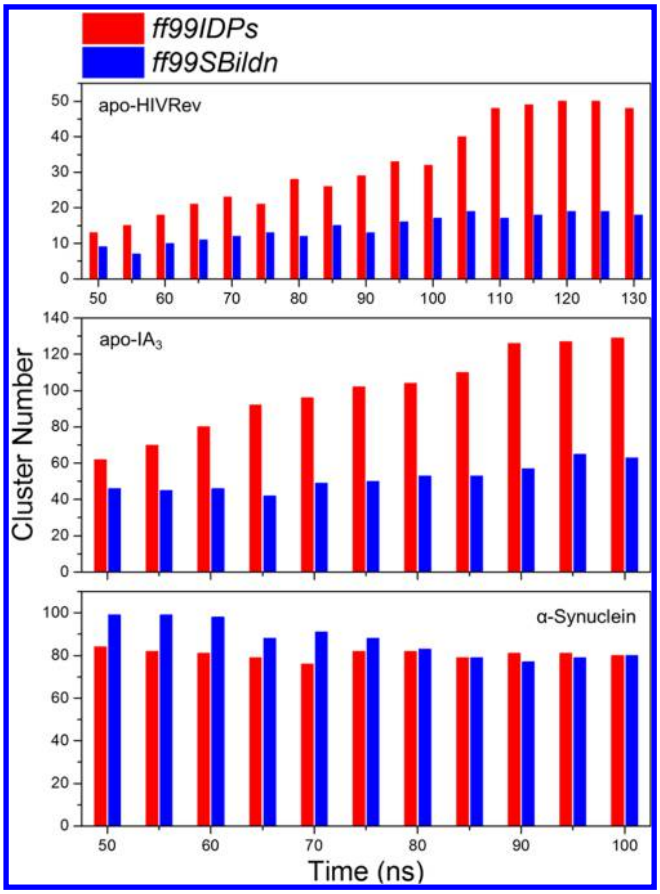


Figure 1. Numbers of clusters that can occupy 70% of the conformations within each 5 ns period under both force fields.

In order to test the convergence of the conformation sampling, the number of clusters within every 5 ns of simulation time was

DOI: 10.1021/acs.jcim.5b00043
J. Chem. Inf. Model. 2015, 55, 1021−1029

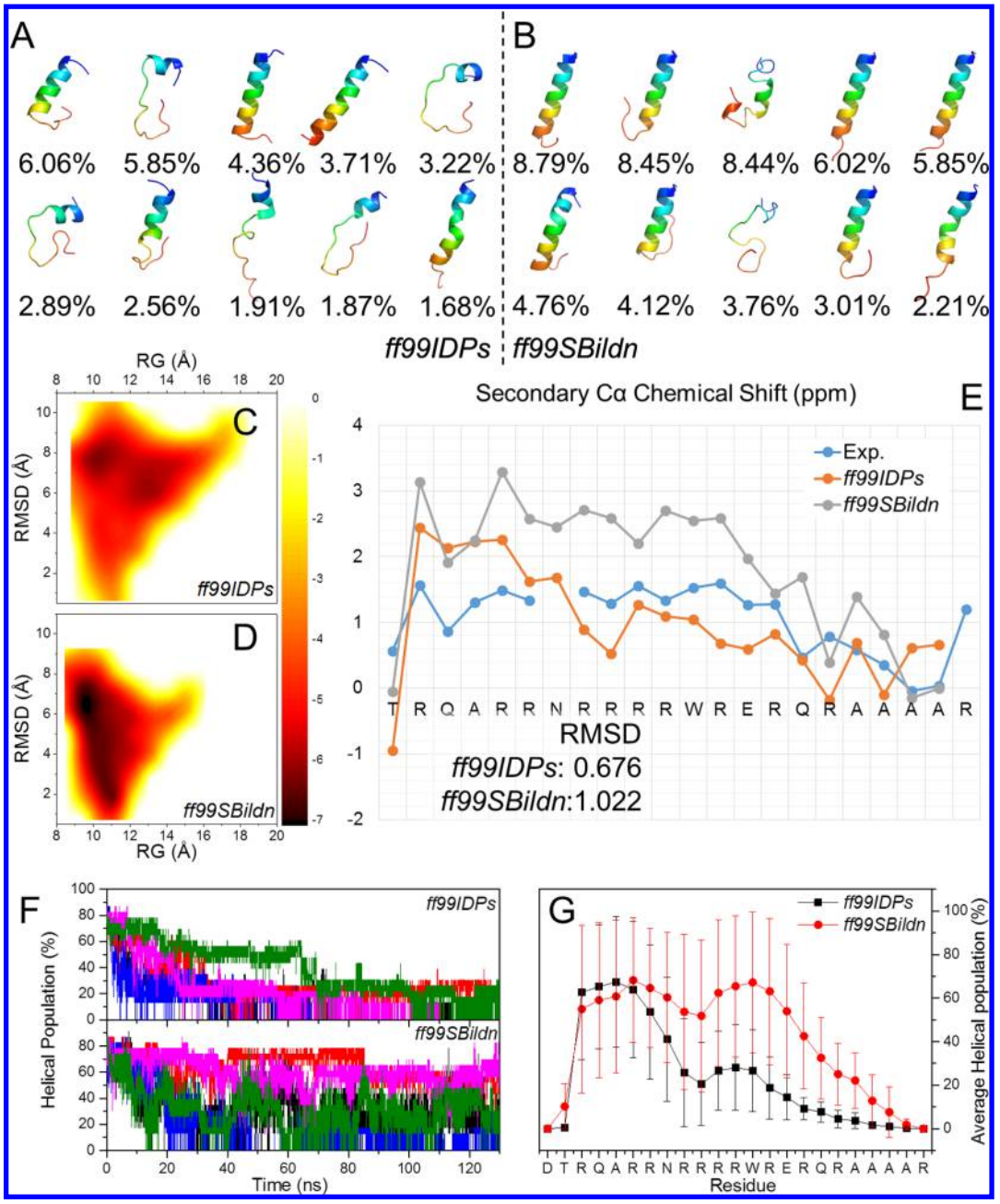


Figure 2. Simulation and thermodynamic data derived from *ff99IDPs* and *ff99SBildn* for apo-HIVRev. (A, B) Representative structures of the top 10 clusters and their occupations. (C, D) PMF free energy landscapes on the 2D space of radius of gyration (RG) and root-mean-square deviation (RMSD), showing that ff99IDPs can sample a wider and more flexible conformation space. (E) Comparison of the secondary chemical shift data. (F, G) Comparison of the average helicities under the two force fields using (F) residual averaging and (G) time averaging.

followed, as shown in Figure 1. For apo-HIVRev, the cluster number of conformers became stable before 130 ns, indicating that 130 ns of simulation is sufficient for the convergence of conformation sampling. For the other IDPs systems, 100 ns of simulation is sufficient for the convergence.

HIV Rev ARM. The secondary Cα chemical shift, clustering, and helicity of apo-HIVRev under *ff99IDPs* and *ff99SBildn* are shown in Figure 2. Apparent differences were found in the structural clustering (Figure 2A,B). The top 10 clusters under *ff99IDPs* occupy 34.11% of the total conformations (top 48 for 70%), many of which have a high ratio of population for disordered structures. However, under *ff99SBildn*, the top 10 clusters occupy up to 55.59% of the total conformations (top 18 for 70%), of which only two conformers are highly disordered. The PMF free energy landscapes as functions of RMSD and RG (Figure 2C,D) show that the distribution of conformers from *ff99IDPs* is located at RMSDs from 1 to 11 Å and RGs from 9 to ~18 Å, while that from *ff99SBildn* is at RMSDs from 1 to 9 Å and RGs from 9 to ~16 Å. This indicates that *ff99IDPs* can sample more flexible conformers than *ff99SBildn*, which is in agreement with the structural clustering. From the clustered representative structures that occupy no less than 70% of the conformations and their occupancies, secondary Cα chemical shifts were calculated and compared with the experimental data (Figure 2E). The full-length RMS*d* between the predicted chemical shifts and the experimental data was 0.676 ppm for *ff99IDPs* and 1.022 ppm for *ff99SBildn*. This suggests that the predicted chemical shifts under *ff99IDPs* are closer to the experimental values than those

DOI: 10.1021/acs.jcim.5b00043
J. Chem. Inf. Model. 2015, 55, 1021−1029

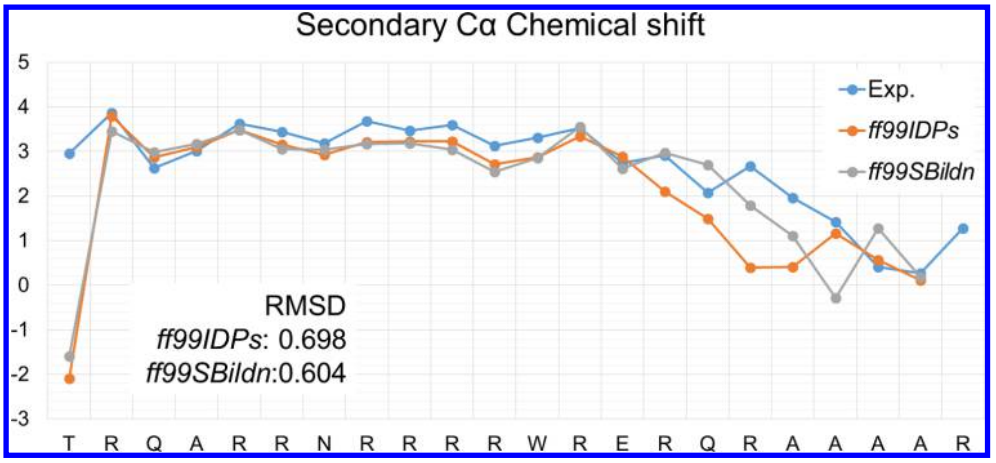


Figure 3. Comparison of the secondary chemical shift data of bound HIVRev under *ff99IDPs* and *ff99SBildn*. RMS*d* values were calculated with the outlier terminal threonine omitted.

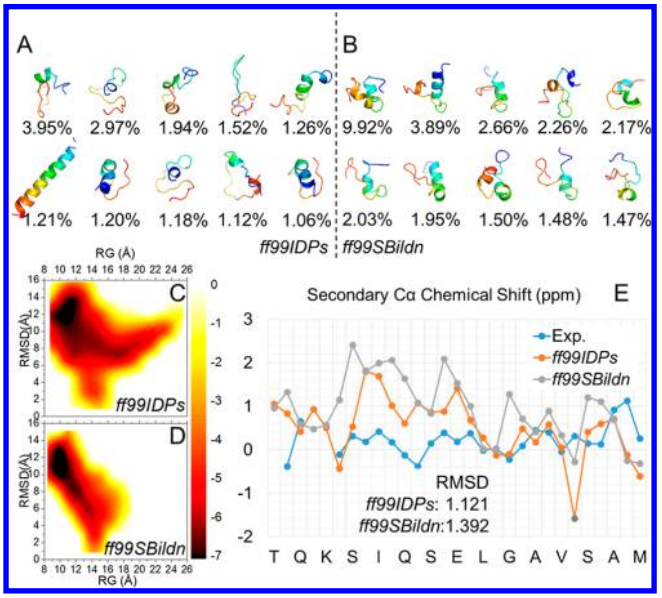


Figure 4. Simulation and thermodynamic data derived from *ff99IDPs* and *ff99SBildn* for apo-IA$_3$. (A, B) Representative structures of the top 10 clusters and their occupations from (A) *ff99IDPs* and (B) *ff99SBildn*. (C, D) PMF free energy landscapes on the 2D space of RG and RMSD. (E) Comparison of the secondary chemical shift data.

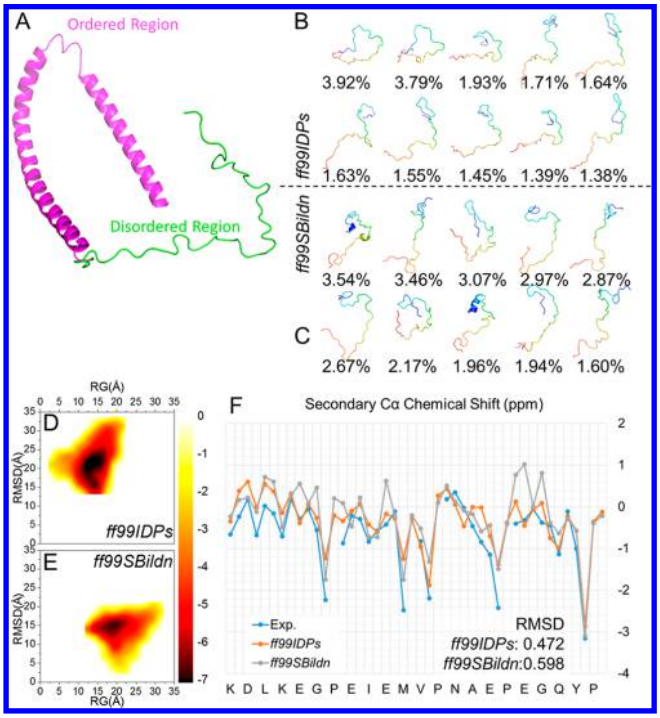


Figure 5. Simulation and thermodynamic data derived from *ff99IDPs* and *ff99SBildn* for apo-αSyn. (A) Cartoon representation of α-synuclein, with the ordered and disordered parts labeled. (B, C) Representative structures of the top 10 clusters and their occupations from (B) *ff99IDps* and (C) *ff99SBildn*. (D, E) PMF free energy landscapes on the 2D space of RG and RMSD. (F) Comparison of the secondary chemical shift data.

under *ff99SBildn*. The calculated *J*-coupling values are shown in Figure S2 in the Supporting Information. Because of the lack of exact experimental data, the full-length RMS*d* was not calculated. However, similar to the prediction of chemical shift, the predicted *J*-coupling values of *ff99IDPs* are better than those of *ff99SBildn*. The HIVRev helicity data are shown in Figure 2F,G. These results (based on two different statistics strategies) indicate that *ff99IDPs* reproduces significantly fewer helical conformations than *ff99SBildn*. We also used STRIDE to yield the secondary structure for HIVRev, as shown in Figure S3 in the Supporting Information. The results are consistent with those from DSSP. This indicates the reliability and reproducibility of the DSSP algorithm. A detailed representation of the time evolution of the secondary structure of apo-HIVRev under the two force fields can be found in Figure S4 in the Supporting Information (the color labels for different secondary structures are defined in Figure S11).

Our tests so far show that *ff99IDPs* performs better than *ff99SBildn* for simulation of IDPs. The clustering analysis shows a higher structural flexibility derived using *ff99IDPs*. The structure trinity of IDPs, i.e., ordered, molten globular, and disordered structures,[3] could also be observed in the *ff99IDPs*-derived structure representatives, especially in the middle region and the C-terminus. In the N-terminus, the performance of *ff99IDPs* is similar to that of *ff99SBildn* for chemical shift or *J*-coupling prediction. We detected internal interactions in apo-HIVRev under both force fields. Long-range interactions, other than short-range hydrogen bonds and hydrophobic interactions, play key roles in the internal contacts. Under *ff99IDPs*, hydrogen

DOI: 10.1021/acs.jcim.5b00043
J. Chem. Inf. Model. 2015, 55, 1021−1029

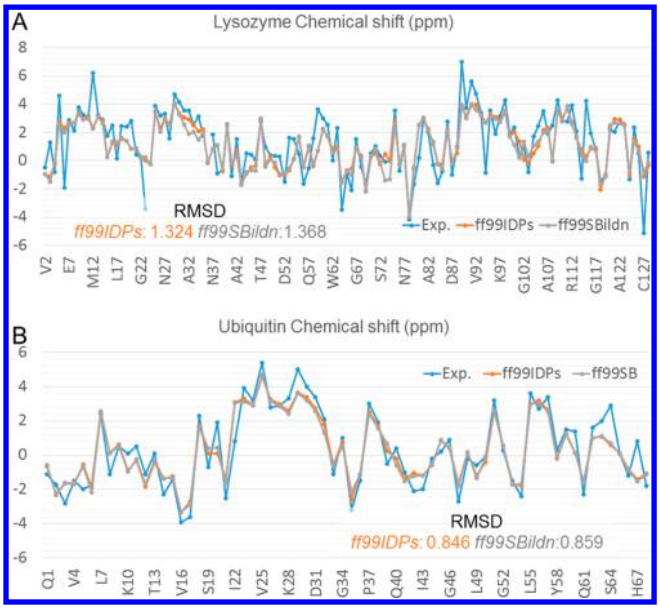


Figure 6. Comparison of secondary chemical shift data for (A) lysozyme and (B) ubiquitin.

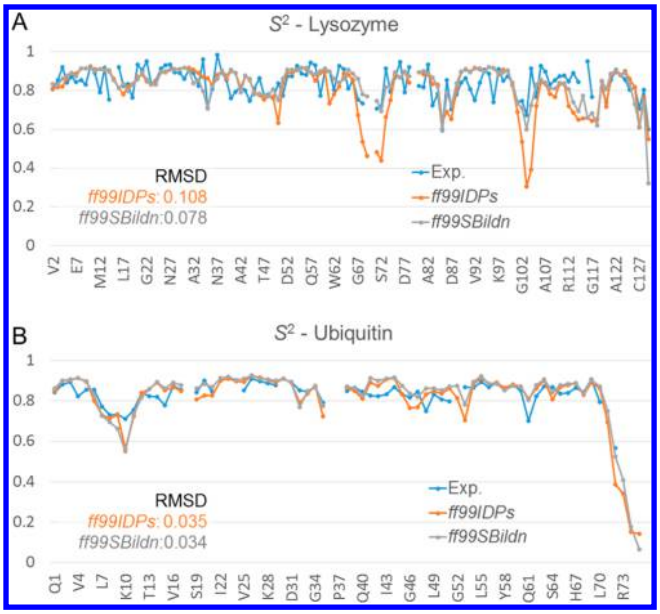


Figure 7. Comparison of order parameter (S^2) data for (A) lysozyme and (B) ubiquitin.

bonds and electrostatic interactions were found only in the N-terminus (A-T-R-Q), as these residues have long side chains that can tangle with charge−charge interactions; under *ff99SBildn*, these interactions were also found in the middle region. The overstability in the N-terminus under *ff99IDPs* may be caused by the consideration of corrections only on the backbone dihedrals, while the charges and electrostatic interactions should be considered as another important aspect in the IDPs modeling. This could be implemented to improve the force field in the next version.

For DNA-bound HIVRev, both *ff99IDPs* and *ff99SBildn* exhibit good performance in reproducing the conformation of structured HIVRev. The calculated secondary Cα chemical shifts of bound HIVRev under both force fields are shown in Figure 3 and compared with experimental data. The RMSD was 0.698 ppm under *ff99IDPs* and 0.604 ppm under *ff99SBildn*. As shown in Figure S12 in the Supporting Information, the helical structures of HIVRev were very stable upon DNA binding, indicating that *ff99IDPs* can also be used for the ordered protein. The time evolution of the secondary structure of bound HIVRev under both force fields is shown in Figure S5.

Aspartic Proteinase Inhibitor. IA_3 is another intrinsically disordered protein that folds rigidly upon binding with aspartic proteinase. Similar to HIVRev, the structural cluster also indicates that *ff99IDPs* reproduces more disordered and flexible conformations than *ff99SBildn* (Figure 4A,B). The top 10 clusters under *ff99IDPs* occupy only 17.41% of the total conformations (top 129 for 70%), while under *ff99SBildn* the top 10 clusters occupy 29.33% of the total conformations (top 63 for 70%). Therefore, the disorder level of conformers from *ff99IDPs* was also higher than that from *ff99SBildn*. The full-length RMS*d* between the predicted secondary Cα chemical shifts and the experimental data was 1.121 ppm for *ff99IDPs* (V25 as an outlier) and 1.392 ppm for *ff99SBildn*. The time evolution of the secondary structure for apo-IA_3 under both force fields is shown in Figure S6 in the Supporting Information. This also shows that the performance of *ff99IDPs* is better than that of *ff99SBildn* to simulate IDPs. However, neither force field could well reproduce the disordered conformers for the middle region of apo-IA_3 because of overstabilized α-helical structures. The CMAP corrections are added only onto disorder-promoting residues, which might cause an energy gap between order- and disorder-promoting residues. As a result, the most significant disparities in IA_3 were found at the boundaries between order- and disorder-promoting residues. This might be eliminated by using a polarizable charge model.

The structures of aspartic proteinase-bound IA_3 modeled under the two force fields are similar. Because of the lack of experimental data, we cannot quantitatively evaluate the performance. Chemical shift data for bound IA_3 can be found in Figure S13 in the Supporting Information.

α-Synuclein. The NMR structures of αSyn have two long α-helices linked with a short turn (residues 1−94; Figure 5A). In the crystal experiment, this region binds with vesicle and micelle.[45] In this test, we focused on the disordered region (residues 95−140) and compared its thermodynamic data with the experimental values. Therefore, all analysis results are focused on this IDR. Structural clustering results are shown in Figure 5B,C, and the secondary structure evolution is presented in Figure S8 in the Supporting Information. The conformers of the top 10 clusters occupy 20.39% (*ff99IDPs*) and 26.25% (*ff99SBildn*), respectively. As shown in Figure 5, the whole conformers are highly disordered under *ff99IDPs*; however, the conformers evolve some helical structure under *ff99SBildn*. This shows that the disordered ratio of the population is larger under *ff99IDPs* than under *ff99SBildn*. The full-length RMS*d* between the predicted secondary Cα chemical shifts and the experimental data was 0.472 ppm for *ff99IDPs* and 0.598 ppm for *ff99SBildn*. This also shows that the performance of *ff99IDPs* is better than that of *ff99SBildn* for simulation of IDPs.

Lysozyme and Ubiquitin. Lysozyme and ubiquitin have been used to evaluate previous force fields many times.[30,63] Here these two intensively studied proteins were also employed to evaluate the performance of *ff99IDPs*. Figures 6 and 7 show the secondary Cα chemical shifts and the order parameters (S^2) derived from *ff99IDPs* and *ff99SBildn* and their comparison to the experimental data. Both force fields perform well, showing that the newly developed *ff99IDPs* is also suitable for the

DOI: 10.1021/acs.jcim.5b00043
J. Chem. Inf. Model. 2015, 55, 1021−1029

structured proteins. The calculated chemical shift results from the two force fields are similar and approach the experimental values. It is interesting to note that *ff99IDPs* leads to more disordered properties in the loop and turn regions, as shown in the S^2 values, which is consistent with the destabilization of the CMAP energy term. Their secondary structure evolution is shown in Figures S9 and S10 in the Supporting Information.

CONCLUSION AND PERSPECTIVES

In this study, we have validated the performance of the newly developed *ff99IDPs* force field with three intrinsically disordered proteins (HIVRev, IA_3, and αSyn) and two ordered proteins (lysozyme and ubiquitin) by means of extensive all-atom molecular dynamics simulations. Thermodynamic data were calculated and compared with experimental data. Overall, *ff99IDPs* leads to more consistent chemical shifts than *ff99SBildn* and reproduces more flexible disorder conformations for IDPs. Thus, the validation data indeed show that *ff99IDPs* exhibits better performance than *ff99SBildn* in simulation of IDPs. At the same time, *ff99IDPs* can also be used to simulate structural proteins.

Nevertheless, some limitations do exist. *ff99IDPs* could not perfectly reproduce the full length of the intrinsically disordered proteins, e.g., the N-terminus of HIVRev and the middle region of IA_3. Stable secondary structures (α-helices in most cases) in these regions were observed in the simulations, which may be the cause of the high Cα chemical shifts. The electrostatic and hydrogen-bonding interactions in these regions might play key roles in the structural stabilization, rather than the dihedral energy corrections in the CMAP energy term. Thus, the next step for improving the performance and accuracy of the IDP-specific force field might be focused on the polar interactions. Indeed, the charge distribution of a residue should be perturbed by its neighboring residues, which is the principle of polarizable force fields. Therefore, our next step will be to explore how to improve the accuracy of the inter-residue interactions in the disordered regions with strategies based on polarizable force fields and incorporate these to improve the IDP force fields.

ASSOCIATED CONTENT

Supporting Information

RMS*d* time evolution for all the test systems (Figure S1); calculated *J*-coupling of apo- and bound HIVRev (Figure S2); comparisons of the helices of HIVRev calculated using DSSP and STRIDE (Figure S3); time evolution of the secondary structure for all of the test systems (Figures S4−S11); structural clusters of bound HIVRev (Figure S12); predicted secondary chemical shift of bound IA_3 under *ff99IDPs* and *ff99SBildn* (Figure S13). The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00043.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: haifengchen@sjtu.edu.cn. Tel: 86-21-34204348. Fax: 86-21-34204348.

*E-mail: ray.luo@uci.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Center for HPC at Shanghai Jiaotong University by grants from the Ministry of Science and Technology of China (2012CB721003), the National High-Tech R&D Program of China (863 Program) (2014AA021502), the National Natural Science Foundation of China (J1210047 and 31271403), the Innovation Program of the Shanghai Education Committee (12ZZ023), and the Medical Engineering Cross Fund of Shanghai Jiaotong University (YG2013MS68 and YG2014MS47).

REFERENCES

(1) Dunker, A. K.; Babu, M. M.; Barbar, E.; Blackledge, M.; Bondos, S. E.; Dosztányi, Z.; Dyson, H. J.; Forman-Kay, J.; Fuxreiter, M.; Gsponer, J.; et al. What's in a Name? Why These Proteins Are Intrinsically Disordered. *Intrinsically Disord. Proteins* **2013**, *1*, No. e24157.

(2) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annu. Rev. Biophys.* **2008**, *37*, 215−246.

(3) 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.; et al. Intrinsically Disordered Protein. *J. Mol. Graphics Modell.* **2001**, *19*, 26−59.

(4) Gough, J.; Dunker, A. K. Sequences and Topology: Disorder, Modularity, and Post/Pre Translation Modification. *Curr. Opin. Struct. Biol.* **2013**, *23*, 417−419.

(5) Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **2014**, *83*, 553−584.

(6) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. Intrinsic Protein Disorder in Complete Genomes. *Genome Inf. Ser.* **2000**, 161−171.

(7) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. Intrinsic Disorder and Protein Function. *Biochemistry* **2002**, *41*, 6573−6582.

(8) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradović, Z.; Dunker, A. K. Intrinsic Disorder in Cell-Signaling and Cancer-Associated Proteins. *J. Mol. Biol.* **2002**, *323*, 573−584.

(9) 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.

(10) Wenning, G. K.; Jellinger, K. A. The Role of α-Synuclein and Tau in Neurodegenerative Movement Disorders. *Curr. Opin. Neurol.* **2005**, *18*, 357−362.

(11) Pawar, A. P.; Dubay, K. F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C. M. Prediction of "Aggregation-Prone" and "Aggregation-Susceptible" Regions in Proteins Associated with Neurodegenerative Diseases. *J. Mol. Biol.* **2005**, *350*, 379−392.

(12) Hampel, H.; Blennow, K.; Shaw, L. M.; Hoessler, Y. C.; Zetterberg, H.; Trojanowski, J. Q. Total and Phosphorylated Tau Protein as Biological Markers of Alzheimer's Disease. *Exp. Gerontol.* **2010**, *45*, 30−40.

(13) Goedert, M.; Spillantini, M.; Jakes, R.; Rutherford, D.; Crowther, R. Multiple Isoforms of Human Microtubule-Associated Protein Tau: Sequences and Localization in Neurofibrillary Tangles of Alzheimer's Disease. *Neuron* **1989**, *3*, 519−526.

(14) Gamblin, T. C.; Berry, R. W.; Binder, L. I. Modeling Tau Polymerization in Vitro: A Review and Synthesis. *Biochemistry* **2003**, *42*, 15009−15017.

(15) Tsai, C.-J.; Ma, B.; Nussinov, R. Protein−Protein Interaction Networks: How Can a Hub Protein Bind So Many Different Partners? *Trends Biochem. Sci.* **2009**, *34*, 594−600.

(16) Collavin, L.; Lunardi, A.; Del Sal, G. P53-Family Proteins and Their Regulators: Hubs and Spokes in Tumor Suppression. *Cell Death Differ.* **2010**, *17*, 901−911.

(17) Chouard, T. Structural Biology: Breaking the Protein Rules. *Nature* **2011**, *471*, 151−153.

(18) Ye, W.; Wang, W.; Jiang, C.; Yu, Q.; Chen, H. Molecular Dynamics Simulations of Amyloid Fibrils: An in Silico Approach. *Acta Biochim. Biophys. Sin.* **2013**, *45*, 503−508.

DOI: 10.1021/acs.jcim.5b00043
J. Chem. Inf. Model. 2015, 55, 1021−1029

(19) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(20) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(21) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. ROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(22) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(23) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-Timescale Molecular Dynamics Simulations of Protein Structure and Function. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120−127.
(24) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(25) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. *AMBER 12*; University of California: San Francisco, 2012.
(26) 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.
(27) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(28) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. 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.
(29) Sorin, E. J.; Pande, V. S. Exploring the Helix−Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophys. J.* **2005**, *88*, 2472−2493.
(30) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(31) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. Sequence Complexity of Disordered Protein. *Proteins: Struct., Funct., Bioinf.* **2001**, *42*, 38−48.
(32) Tan, R.; Chen, L.; Buettner, J. A.; Hudson, D.; Frankel, A. D. RNA Recognition by an Isolated α Helix. *Cell* **1993**, *73*, 1031−1040.
(33) Calnan, B. J.; Biancalana, S.; Hudson, D.; Frankel, A. D. Analysis of Arginine-Rich Peptides from the HIV Tat Protein Reveals Unusual Features of RNA−Protein Recognition. *Gene Dev.* **1991**, *5*, 201−210.
(34) Lazinski, D.; Grzadzielska, E.; Das, A. Sequence-Specific Recognition of RNA Hairpins by Bacteriophage Antiterminators Requires a Conserved Arginine-Rich Motif. *Cell* **1989**, *59*, 207−218.
(35) Craig, C.; Race, E.; Sheldon, J.; Whittaker, L.; Gilbert, S.; Moffatt, A.; Rose, J.; Dissanayeke, S.; Chirn, G.-W.; Duncan, I. B.; Cammack, N. HIV Protease Genotype and Viral Sensitivity to HIV Protease Inhibitors following Saquinavir Therapy. *AIDS* **1998**, *12*, 1611−1618.
(36) Monod, M.; Togni, G.; Hube, B.; Sanglard, D. Multiplicity of Genes Encoding Secreted Aspartic Proteinases in *Candida* Species. *Mol. Microbiol.* **1994**, *13*, 357−368.
(37) Dame, J. B.; Reddy, G. R.; Yowell, C. A.; Dunn, B. M.; Kay, J.; Berry, C. Sequence, Expression and Modeled Structure of an Aspartic Proteinase from the Human Malaria Parasite *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **1994**, *64*, 177−190.
(38) Dreyer, T.; Valler, M.; Kay, J.; Charlton, P.; Dunn, B. The Selectivity of Action of the Aspartic-Proteinase Inhibitor IA_3 from Yeast (*Saccharomyces cerevisiae*). *Biochem. J.* **1985**, *231*, 777−779.
(39) 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. Mol. Biol.* **2000**, *7*, 113−117.
(40) Auluck, P. K.; Caraveo, G.; Lindquist, S. α-Synuclein: Membrane Interactions and Toxicity in Parkinson's Disease. *Annu. Rev. Cell Dev. Biol.* **2010**, *26*, 211−233.
(41) Goldberg, M. S.; Lansbury, P. T., Jr. Is There a Cause-and-Effect Relationship between α-Synuclein Fibrillization and Parkinson's Disease? *Nat. Cell Biol.* **2000**, *2*, E115−E119.
(42) Makin, O. S.; Atkins, E.; Sikorski, P.; Johansson, J.; Serpell, L. C. Molecular Basis for Amyloid Fibril Formation and Stability. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 315−320.
(43) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; et al. BioMagResBank. *Nucleic Acids Res.* **2008**, *36* (Suppl. 1), D402−D408.
(44) Casu, F.; Duggan, B. M.; Hennig, M. The Arginine-Rich RNA-Binding Motif of HIV-1 Rev Is Intrinsically Disordered and Folds upon RRE Binding. *Biophys. J.* **2013**, *105*, 1004−1017.
(45) Rao, J. N.; Jao, C. C.; Hegde, B. G.; Langen, R.; Ulmer, T. S. A Combinatorial NMR and EPR Approach for Evaluating the Structural Ensemble of Partially Folded Proteins. *J. Am. Chem. Soc.* **2010**, *132*, 8657−8668.
(46) Young, A.; Dewan, J. C.; Nave, C.; Tilton, R. Comparison of Radiation-Induced Decay and Structure Refinement from X-ray Data Collected from Lysozyme Crystals at Low and Ambient Temperatures. *J. Appl. Crystallogr.* **1993**, *26*, 309−319.
(47) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **1987**, *194*, 531−544.
(48) *SYBYL-X Suite*, version 2.1.1; Certara: Princeton, NJ, 2012.
(49) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: an $N\cdot\log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(50) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542−1555.
(51) Götz, A. W.; Salomon-Ferrer, R.; Poole, D.; Grand, S.; Walker, R. Routine Microsecond Molecular Dynamics Simulations with AMBER. Part II: Particle Mesh Ewald. *J. Chem. Theory. Comput.* **2013**, *9*, 3878−3888.
(52) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Chem. Phys.* **1977**, *23*, 327−341.
(53) Feig, M.; Karanicolas, J.; Brooks, C. L., III. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graphics Modell.* **2004**, *22*, 377−395.
(54) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.
(55) 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* (Suppl.1), D411−D419.
(56) Frishman, D.; Argos, P. Knowledge-Based Protein Secondary Structure Assignment. *Proteins* **1995**, *23*, 566−579.
(57) Heinig, M.; Frishman, D. STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. *Nucleic Acids Res.* **2004**, *32*, W500−W502.
(58) Shen, Y.; Bax, A. Protein Backbone Chemical Shifts Predicted from Searching a Database for Torsion Angle and Sequence Homology. *J. Biomol. NMR* **2007**, *38*, 289−302.
(59) Bystrov, V. F. Spin−Spin Coupling and the Conformational States of Peptide Systems. *Prog. Nucl. Magn. Reson. Spectrosc.* **1976**, *10*, 41−82.
(60) Pardi, A.; Billeter, M.; Wüthrich, K. Calibration of the Angular Dependence of the Amide Proton−C^{α} Proton Coupling Constants,

DOI: 10.1021/acs.jcim.5b00043
J. Chem. Inf. Model. 2015, 55, 1021−1029

$^3J_{HN\omega}$ in a Globular Protein: Use of $^3J_{HN\alpha}$ for Identification of Helical Secondary Structure. *J. Mol. Biol.* **1984**, *180*, 741−751.
(61) Ludvigsen, S.; Andersen, K. V.; Poulsen, F. M. Accurate Measurements of Coupling Constants from Two-Dimensional Nuclear Magnetic Resonance Spectra of Proteins and Determination of φ-Angles. *J. Mol. Biol.* **1991**, *217*, 731−736.
(62) *PyMOL*, version 1.7; Schrodinger, LLC: New York, 2014.
(63) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D., Jr. Importance of the CMAP Correction to the CHARMM22 Protein Force Field: Dynamics of Hen Lysozyme. *Biophys. J.* **2006**, *90*, L36−L38.

DOI: 10.1021/acs.jcim.5b00043
J. Chem. Inf. Model. 2015, 55, 1021−1029