JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment

Sarah Rauscher,*[,†] Vytautas Gapsys,[†] Michal J. Gajda,[‡] Markus Zweckstetter,[‡,§,∥] Bert L. de Groot,[†] and Helmut Grubmüller[†]

[†]Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen 37077, Germany

[‡]Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Göttingen 37077, Germany

[§]German Center for Neurodegenerative Diseases (DZNE), Göttingen 37077, Germany

[∥]Center for Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB), University Medical Center, Göttingen 37073, Germany

Supporting Information

ABSTRACT: Intrinsically disordered proteins (IDPs) are notoriously challenging to study both experimentally and computationally. The structure of IDPs cannot be described by a single conformation but must instead be described as an ensemble of interconverting conformations. Atomistic simulations are increasingly used to obtain such IDP conformational ensembles. Here, we have compared the IDP ensembles generated by eight all-atom empirical force fields against primary small-angle X-ray scattering (SAXS) and NMR data. Ensembles obtained with different force fields exhibit marked differences in chain dimensions, hydrogen bonding, and secondary structure content. These differences are unexpectedly large: changing the force field is found to have a stronger effect on secondary structure content than changing the entire peptide sequence. The CHARMM 22* ensemble performs best in this force field comparison: it has the lowest error in chemical shifts and J-couplings and agrees well with the SAXS data. A high population of left-handed α-helix is present in the CHARMM 36 ensemble, which is inconsistent with measured scalar couplings. To eliminate inadequate sampling as a reason for differences between force fields, extensive simulations were carried out (0.964 ms in total); the remaining small sampling uncertainty is shown to be much smaller than the observed differences. Our findings highlight how IDPs, with their rugged energy landscapes, are highly sensitive test systems that are capable of revealing force field deficiencies and, therefore, contributing to force field development.

INTRODUCTION

Intrinsically disordered proteins (IDPs) carry out crucial biological functions in all kingdoms of life.[1] The human proteome is estimated to contain approximately a million disordered motifs, which often act as signals in cellular pathways, including protein degradation, trafficking, and targeting.[2] IDP aggregation is involved in diverse cellular functions including the selective passage of material through the nuclear pore complex[3] and the segregation of materials in membrane-less organelles via intracellular phase separation.[4,5] A fundamental understanding of the structural properties of IDPs is crucial to understanding the wide range of cellular functions relying on protein disorder.

An IDP by definition cannot be described by a single average structure but instead must be described as an ensemble of interconverting conformations. Obtaining accurate structural ensembles of IDPs is the aim of many recent studies, both experimental and computational.[6,7] The protein ensemble database (pE-DB),[8] which is analogous to the PDB for structures of folded proteins, contains a growing collection of IDP ensembles. On the experimental side, nuclear magnetic resonance (NMR), small-angle X-ray scattering (SAXS), and single-molecule spectroscopy have emerged as highly useful and complementary methods for obtaining structural information.[6,9] Fluorescence resonance energy transfer (FRET), fluorescence correlation spectroscopy (FCS), and SAXS provide measurements of overall chain dimensions. NMR spectroscopy provides site-specific information, for example, on secondary structure content and distances between labeled sites as well as measurements of hydrodynamic radius using pulsed field gradient NMR (PFG-NMR).[10]

On the theory side, a variety of computational methods have been developed to obtain structural ensembles of IDPs. These

Received: August 3, 2015
Published: October 9, 2015

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524

methods can be broadly classified into two types: (1) those that use experimental data to guide ensemble generation or selection, and (2) those that generate ensembles of IDPs de novo, that is, without using experimental data as an input.

An example of the first type of method is the use of experimental data as restraints in simulations. For instance, NMR chemical shift restraints and distance restraints based on paramagnetic relaxation enhancement (PRE) measurements were used in molecular dynamics (MD) simulations of the denatured state of ACBP and α-synuclein, respectively.[11−13] Other computational methods, such as ENSEMBLE[14] and ASTEROIDS,[15] use experimental data to select ensembles from pregenerated pools of conformations. Ensembles consistent with experimental data have also been selected from conformations obtained using MD simulations.[16−18] Ball et al. compared knowledge-based ensemble selection approaches to ensembles obtained using de novo MD simulations.[17] Because IDP ensembles are severely underdetermined (that is, there are many degrees of freedom and relatively few experimental observables), cross-validation and care in avoiding overfitting are essential to these approaches.[16,19,20]

Computational methods of the second type have been used extensively to obtain ensembles of IDPs de novo. A variety of simulation methods (MD, Monte Carlo, metadynamics, replica exchange) and different levels of representation (coarse-grained, implicit solvent, all-atom with explicit water) have been used to obtain IDP ensembles.[7,21−29] There are two main challenges encountered in de novo simulations. First, extensive simulations are needed to ensure that relevant regions of conformational space are adequately sampled. Although this requirement applies to all biomolecular simulations, it presents a particularly formidable challenge in the case of IDPs due to their high conformational heterogeneity. Second, and more importantly, the accuracy of modern force fields for IDP simulations is not well-characterized.

MD simulations have been used to study the structure and dynamics of folded proteins for decades.[30] During this time, substantial effort has been put into the development and improvement of empirical force fields. The accuracy of the description of the structure and dynamics of globular proteins, as well as that of the relative stabilities of different types of secondary structure, were improved in the CHARMM and Amber force fields.[31−34] Systematic force field comparison studies have shown that, overall, force field modifications indeed tend to be improvements: simulations with more recently developed force fields produce more accurate ensembles of globular proteins compared to the older force fields on which they are based.[35,36]

Some modern force fields describe small, globular proteins quite well: NMR observables computed from these ensembles agree with experimental values within the error expected for the calculation of these observables.[36,37] In a recent comparison of force fields for folded proteins, Amber ff99sb*-ildn and CHARMM22* were the only two force fields consistent with experimental data.[35] In contrast, force fields have been shown to differ significantly in their ability to fold proteins,[38] especially in the challenging task of folding proteins from multiple structural classes.[35] A study of villin headpiece aggregation using different force fields and solution conditions suggested that protein−protein interactions generally tend to be overestimated: all of the force fields in this study showed aggregation and are therefore inconsistent with experimental evidence indicating no aggregation.[39]

Many force field modifications have been directed at improving the accuracy of conformational ensembles of globular proteins. In contrast, most force fields have not been developed for simulations of IDPs. Several recent force field modifications have improved the balance of secondary structure propensities to be able to fold proteins of multiple structural classes. It would in fact be somewhat surprising if accurate IDP ensembles could be obtained using a force field optimized only for folded proteins. Nevertheless, all-atom simulations are increasingly being used to obtain ensembles of IDPs (see recent studies 23, 24, and 40−43).

It is difficult to draw conclusions on the accuracy of IDP simulations from the contradictory findings reported so far. On the one hand, good agreement between computed and measured experimental observables was observed in some IDP simulations.[23−25,27,28,42,44−46] On the other hand, the accuracy of unfolded state and IDP structural ensembles obtained using several widely used force fields has been called into question.[25−27,43,46−53] Piana et al. suggested that modern force fields, including Amber ff99sb*-ildn and CHARMM 22* in particular, produce IDP and unfolded state ensembles that are on average too compact.[48] It has also been suggested that Amber force fields systematically underestimate chain dimensions of IDPs and unfolded states.[51] To address these force field deficiencies, a new water model, TIP4P-D,[26] and new force field, Amber 03ws,[51] were recently introduced.

While these recent studies demonstrate a substantial interest in obtaining accurate ensembles of IDPs using all-atom simulations, there is currently no consensus on the most accurate force field or the suitability of any force field for this purpose. Information on the accuracy of IDP ensembles is sparse, anecdotal, and contradictory, which may be due a combination of multiple factors: (1) inadequate conformational sampling of IDP ensembles, (2) comparisons to models derived from primary experimental data rather than the primary data itself, and (3) comparisons to relatively few (and sometimes only one) observables. In the absence of a comprehensive comparison, it is not surprising that a consensus is currently lacking concerning the accuracy of IDP simulations.

Here, we aim to evaluate the accuracy of IDP ensembles obtained using de novo molecular simulations. Toward this aim, we compared ensembles obtained using eight all-atom empirical force fields (Table 1) to primary SAXS and NMR data. For reasons of computational feasibility, we included only eight force fields. We addressed the sampling problem using temperature replica exchange[54] as well as extensive sampling, accumulating a total sampling time of 964 μs (a detailed list of simulations is provided in Table S1).

Table 1. Force Fields Included in the Comparison

force field (abbreviation)	peptide force field	water model
Amber ff99sb*-ildn (a99sb)	Amber ff99sb*-ildn[33]	TIP3P[55]
Amber ff03w (a03w)	Amber ff03w[34]	TIP4P-2005[56]
Amber ff03ws (a03ws)	Amber ff03ws[51]	TIP4P-2005[56]
ABSINTH (ABS)	OPLS-AA/L[57]	ABSINTH implicit solvent[21]
CHARMM 22* (c22*)	CHARMM 22*[33]	charmm-modified TIP3P[31]
CHARMM 22* (c22*/D)	CHARMM 22*[33]	TIP4P-D[26]
CHARMM 36[1] (c36[1])	CHARMM 36[58]	TIP3P[55]
CHARMM 36[2] (c36[2])	CHARMM 36[58]	charmm-modified TIP3P[31]

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524

For such a comparative study, the choice of model system is crucial. The model system must be small enough to make it possible to obtain adequate sampling for each force field. In addition, and more importantly, experimental data characterizing both local and global structural properties are needed. We chose a set of sequences that cover a broad spectrum of sequence properties from highly charged in the case of a disordered arginine/serine (RS) peptide to uncharged and enriched in polar and hydrophobic residues in the case of FG-nucleoporin peptides.

The RS peptide is a well-suited IDP for this comparison, it has previously been extensively characterized by multiple NMR experiments by Xiang et al.[16] as well as by SAXS experiments in this work. Its small size (24 residues) makes it feasible to perform extensive simulations using multiple force fields with current computational capabilities. The RS peptide undergoes a phosphorylation-induced reduction in conformational entropy, which is thought to be important in modulating protein-RNA interactions.[16] Here, we carried out simulations of only the unphosphorylated form. Two FG-nucleoporin peptides with sequences based on the sequence of yeast Nsp1p[59] were also studied. These peptides differ in length (16 residues and 50 residues). FG-nucleoporins are a well-studied class of IDPs that are characterized by the presence of FG motifs and are essential for the selectivity of the nuclear pore complex.[60] We also studied $(AAQAA)_3$ and the HEWL peptide, which is a 19-residue sequence derived from hen egg white lysozyme.[58]

METHODS

Simulations. The RS peptide was built in a fully extended configuration in PYMOL with protonation states to match experimental conditions (arginine residues as well as the N- and C-termini were simulated in their charged states). The simulation system consisted of the peptide in a rhombic dodecahedral box with water molecules and 0.15 M NaCl for a total of ∼42000 atoms. GROMACS version 4.5.4[61] was used for all simulations. Prior to the production runs, energy minimization with the steepest descent algorithm was performed. The lengths of bonds with hydrogen atoms were constrained using the LINCS algorithm.[62] An integration time step of 2 fs was used. A cutoff of 0.95 nm was used for the Lennard-Jones interactions and short-range electrostatic interactions. Long-range electrostatic interactions were calculated by particle-mesh Ewald summation with a grid spacing of 0.12 nm and a fourth order interpolation.[63] The velocity rescaling thermostat was used for all simulations.[64] Equilibration was performed at 298 K for 1 ns using Berendsen pressure coupling[65] followed by 5 ns of simulation in the NPT ensemble using the Parrinello−Rahman algorithm.[66] The configuration from this simulation with a volume closest to the average volume was then used for subsequent replica exchange (RE)[34] simulations in the canonical ensemble. A total of 97 temperatures between 298 and 450 K were used with a mean acceptance ratio of 0.35. Temperature exchanges were attempted every 2 ps for a total of 97 μs (1 μs of simulation per temperature). Coordinates were stored before each temperature exchange; a total of 0.5×10^6 configurations were collected per temperature. Simulations with the TIP4P-D water model[26] were carried out without replica exchange, consistent with Piana et al.,[26] as conformational sampling in this force field is fast, and RE was not needed to obtain sufficient sampling, comparable to the other force fields. Simulations with the ABSINTH[21] implicit solvent model were carried out with the CAMPARI software following a similar protocol to Das and Pappu.[67] The CAMPARI parameter set abs3.2_opls.prm was used for the peptide, and the ion parameters of Mao and Pappu were used.[68] In Table S2, we provide the details of the Monte Carlo move set used. Visual molecular dynamics (VMD) was used for all molecular visualizations.[69]

Analysis of Structural Ensembles. The GROMACS utilities g_gyrate and g_hbond were used to calculate the radius of gyration, R_g, and number of hydrogen bonds, respectively. Secondary structure was assigned according to the DSSP algorithm.[70] Contacts between residue pairs were defined if any two atoms were within a cutoff of 4.5 Å. Left handed α-helix was defined as three or more consecutive residues in the αL basin of the Ramachandran plot. Standard error of the mean was computed using a blocking procedure.[71] For each simulation, an equilibration period was delineated on the basis of both R_g and hydrogen bonds; this initial collapse of the peptide to more compact conformations was excluded from analysis.

Computing Experimental Observables. SAXS scattering curves were computed individually for every conformation in each ensemble using two different approaches, CRYSOL[72] and FOXS.[73] Ensemble-averaged scattering curves were computed for each force field. Each of these curves were fit to the experimental curve following the fitting procedure used by Chen et al.[74] The software PRIMUS[75] was used to compute the R_g from each scattering curve using Guinier analysis. The hydrodynamic radius, R_h, was computed for each configuration using HYDROPRO[76] with the parameters of Mao et al.[45] Comparison is made to the hydrodynamic radius measured using pulse-field-gradient NMR by Xiang et al.[16] The Karplus equation parameters used to calculate scalar couplings are given in Table S3.[77−81] RMS errors for the calculation of J couplings are 0.73 Hz ($^3J_{HNH\alpha}$), 0.5 Hz ($^1J_{C\alpha C\beta}$), 2 Hz ($^1J_{C\alpha H\alpha}$), 0.38 Hz ($^3J_{NC\gamma}$ and $^3J_{CC\gamma}$).[77−81] Chemical shifts were calculated using both SHIFTX2[82] and SPARTA+.[83] The reported errors for SHIFTX2 are the lowest of any chemical shift predictor (0.4412 ppm for Cα and 0.5330 ppm for C′).[82] The RMS errors for SPARTA+ are 0.94 ppm for Cα and 1.09 ppm for C′.[83] The errors may be higher when applied to IDPs. All of the computed NMR observables are compared to those reported by Xiang et al. for the unphosphorylated RS peptide.[16]

Experimental Methods. Small angle scattering data was gathered on the X33 beamline at the EMBL Outstation using the DORIS synchrotron source located in DESY, Hamburg.[84] Samples at concentrations of 2 and 0.33 mg/mL or 0.67 mL in buffer at 25 °C were loaded into sample cell using X33 automated sample changer[85] and exposed for four frames, 30 s each. The buffer was 50 mM Na-phosphate buffer, 100 mM NaCl, and pH 7.0, which was also used for all RS peptide NMR experiments. Data was recorded on a Pilatus 1 M photon counting detector. After comparing frames for radiation damage, each frame was radially averaged and then 1d-averaged using the AutoPilatus software. I(0) and Rg analysis was done manually to ensure quality in the presence of high noise, using the PRIMUS program, and verified by visual checking of used data ranges.[75,86]

RESULTS AND DISCUSSION

Chain Dimensions Depend Strongly on Force Field. Conformational ensembles of the RS peptide were obtained using eight different force fields (Table 1). As can be seen in Figure 1, peptide chain dimensions depend strongly on the

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524

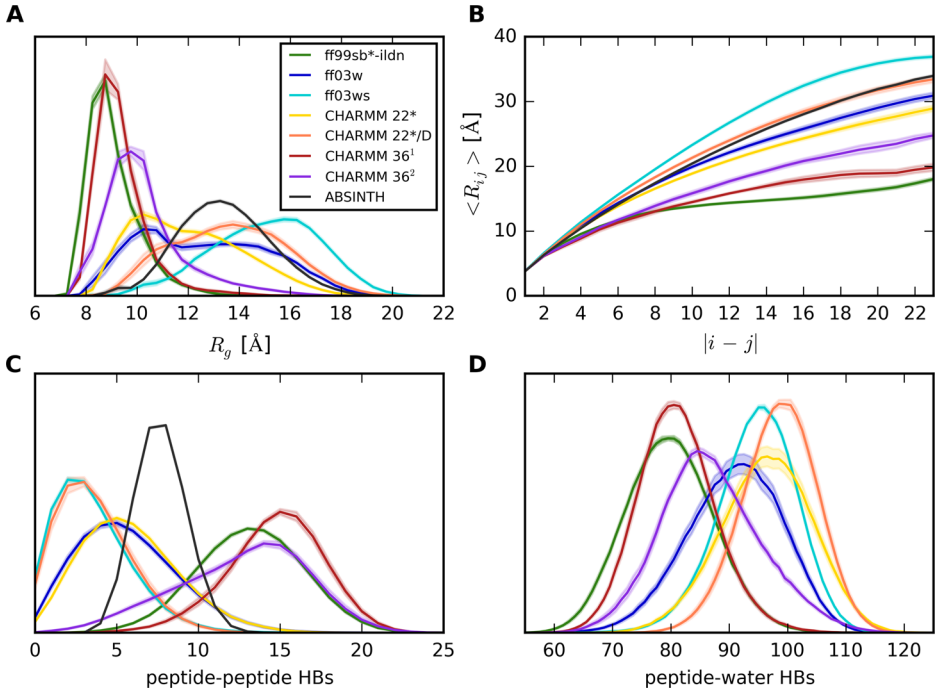

Figure 1. Chain dimensions and hydrogen bonding in different force fields. (A) Histograms of the radius of gyration, R_g, for the structural ensembles obtained in each force field (the legend applies to all figure panels). CHARMM 22* and CHARMM 22*/D refer to simulations with charmm-modified TIP3P and TIP4P-D, respectively. CHARMM 36[1] and CHARMM 36[2] refer to simulations with TIP3P and charmm-modified TIP3P, respectively. (B) Ensemble-averaged distance, $<R_{ij}>$, between the α-carbon atoms of residue pairs, i and j, vs sequence separation, $|i - j|$. (C) Histograms of the number of intrapeptide hydrogen bonds. (D) Histograms of the number of peptide-water hydrogen bonds. Shading indicates statistical uncertainty.

force field. This dependence is evident in both the radius of gyration, R_g (Figure 1A), and mean separation distance between residue pairs (Figure 1B). The ensemble obtained with Amber ff99sb*-ildn is the most compact, resembling a collapsed globule-like ensemble. The most expanded ensemble is obtained with Amber ff03ws. The other six force fields generated a broad spectrum of chain dimensions that fall between these two extreme cases. For illustrating these differences, a selection of structures from each force field are shown in Figure S1. The small statistical uncertainties reported for each structural property and each force field in Figure 1 indicate that sufficient sampling of conformational space was obtained. This is further confirmed by analysis of the contact maps in the first and second halves of the simulations, which are remarkably similar (Figure S2), indicating that the conformational ensembles are well-sampled in all cases. The observed differences therefore must be attributed to inherent force field differences and not to inadequate conformational sampling.

We also assessed the effect of the chosen water model. For three of the force fields, we have therefore kept the same force field for the peptide and used two different water models. In each of these cases, the difference in water model has a significant effect on chain compactness (Figure 1A and B). With CHARMM 22*, the TIP4P-D water model produces a more expanded ensemble than charmm-modified TIP3P. The relative expansion of the ensemble in TIP4P-D compared to charmm-modified TIP3P is consistent with results reported for other IDPs.[26] Amber ff03ws produces a more expanded ensemble than Amber ff03w; this is also to be expected given that the depth of the Lennard-Jones potential, ϵ, between protein atoms and water oxygen atoms is scaled by a factor of 1.1 in Amber ff03ws compared to ff03w.[51] With CHARMM 36, charmm-modified TIP3P produces a significantly more expanded ensemble than TIP3P. Both of these water models were included in the development of CHARMM 36;[58] here, they lead to significant differences in compactness. Because of the strong dependence of chain dimensions on the water model, no simple grouping of Amber- and CHARMM-based force fields with respect to compactness is evident.

The differences in chain dimensions correlate with marked differences in the balance between chain−chain and chain-water interactions (Figure 1C and D). Conformations in the CHARMM 36[1] ensemble have, on average, the highest number of intrapeptide hydrogen bonds. Conformations in the Amber ff99sb*-ildn ensemble have, on average, the fewest hydrogen bonds to water molecules. Intrachain contacts in this ensemble are formed between residues close in sequence (turns) as well as long-range contacts (Figure 2A and Figure S3), consistent with the behavior of a collapsed globule. The ensembles obtained using the other force fields have intrapeptide hydrogen bonds that occur primarily in the form of local turns and helices with relatively few nonlocal contacts (Figure 2A and Figure S3).

Comparing the chain dimensions (Figure 1A and B) of ensembles obtained using different force fields to hydrogen bonding (Figure 1C and D), it can be seen that these structural properties are not perfectly correlated. Specifically, ensembles with significantly different structural properties may have the same mean dimensions. For instance, the chain dimensions CHARMM 22*/TIP4P-D ensemble are nearly the same as those of the ABSINTH ensemble (Figure 1B), but the ensemble with ABSINTH has a significantly higher population of intrapeptide hydrogen bonds (Figures 1C and 2A). The

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524

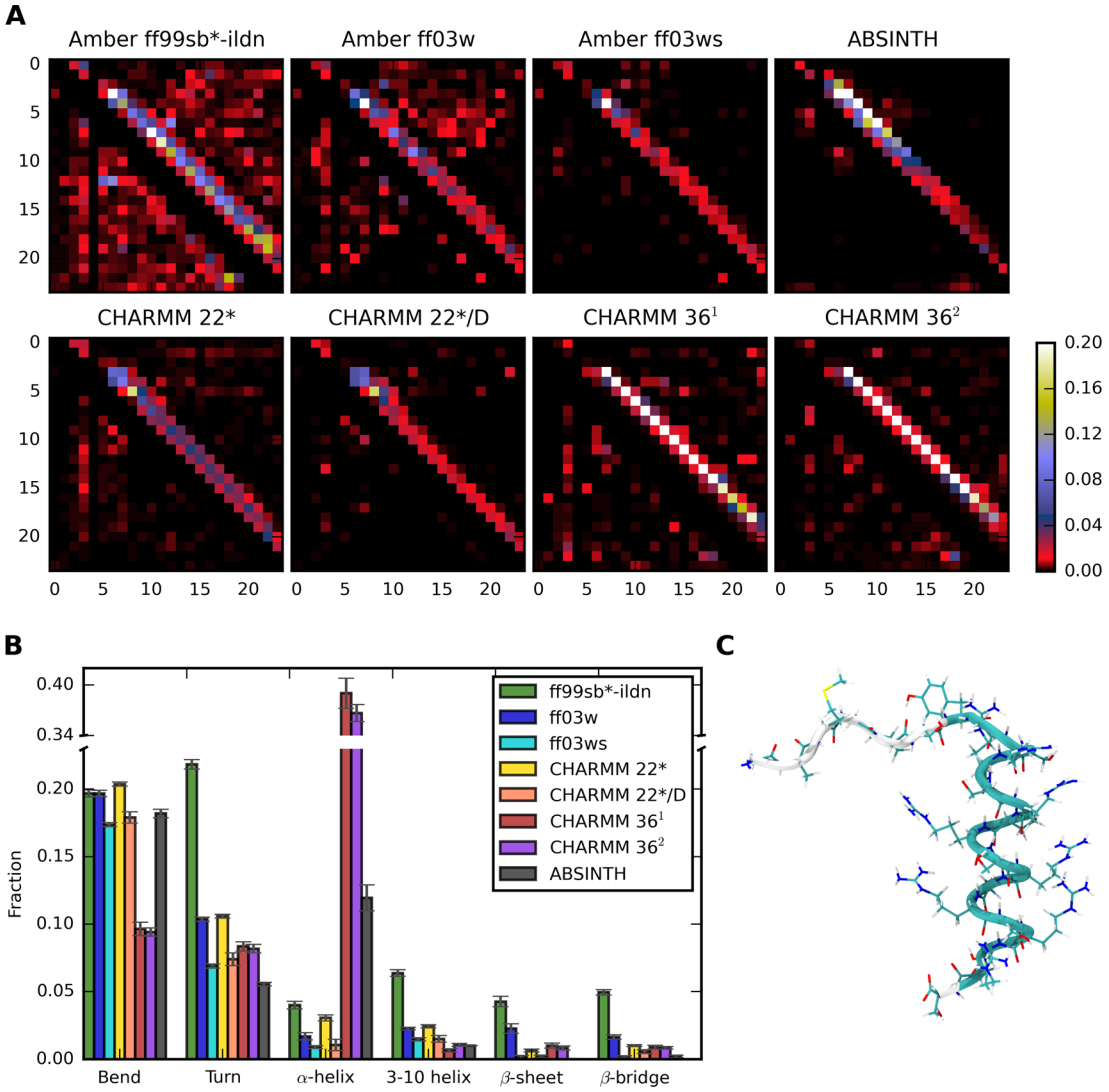

Figure 2. Secondary structure content in different force fields. (A) Hydrogen-bond contact maps for the ensemble in each force field. Shown are hydrogen bonds between the carbonyl oxygen and amide nitrogen of the polypeptide backbone, which are colored according to their population in the ensemble from black (not observed) to white (present in >20% of configurations). (B) The fraction of residues assigned a particular type of secondary structure according to the DSSP[70] algorithm. (C) A representative structure from the CHARMM 36 ensemble containing an extended, left-handed α-helix.

ABSINTH implicit solvent ensemble is unique in another respect: the low R_g peak (centered at 9 Å, Figure 1A) is due to the presence of an arginine claw-type structure (a representative conformation is provided in Figure S4; low Rg (<9.6 Å) conformations were observed in every replica exchange run carried out with ABSINTH). This conformation is reminiscent of the arginine claw observed previously in implicit solvent simulations of a phosphorylated RS peptide in which the phosphoserine residues were pointing outward into the solvent.[87,88] The claw-like conformation in the ABSINTH ensemble is actually an inverse arginine claw in which the arginine residues point outward and the serine residues point inward.

Secondary Structure Content Depends Strongly on Force Field. Like chain dimensions and hydrogen bonding, secondary structure content also depends strongly on the force field for both the RS peptide (Figure 2B) and the FG-nucleoporin peptide (Figure S5). We compared the secondary structure content of ensembles obtained using different force fields for the same peptide sequence. The RMS difference in secondary structure content between ensembles with the same sequence and different force fields was calculated to be 0.1. We also compared the ensembles of the RS peptide and FG-nucleoporin peptide obtained using the same force field. The RMS difference in secondary structure content between ensembles obtained using the same force field and different sequences was calculated to be 0.06. Overall, we find that a change in force field has a stronger effect on secondary structure content than changing the entire peptide sequence. For the two peptides compared here, the change in sequence is quite large. The RS peptide is highly charged, whereas the FG peptide is composed entirely of polar and hydrophobic residues. This result demonstrates how dramatically sensitive IDPs are to force field selection. Their rugged energy landscapes with many iso-energetic minima separated by low barriers make IDPs highly sensitive test systems to assess force field accuracy.

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524

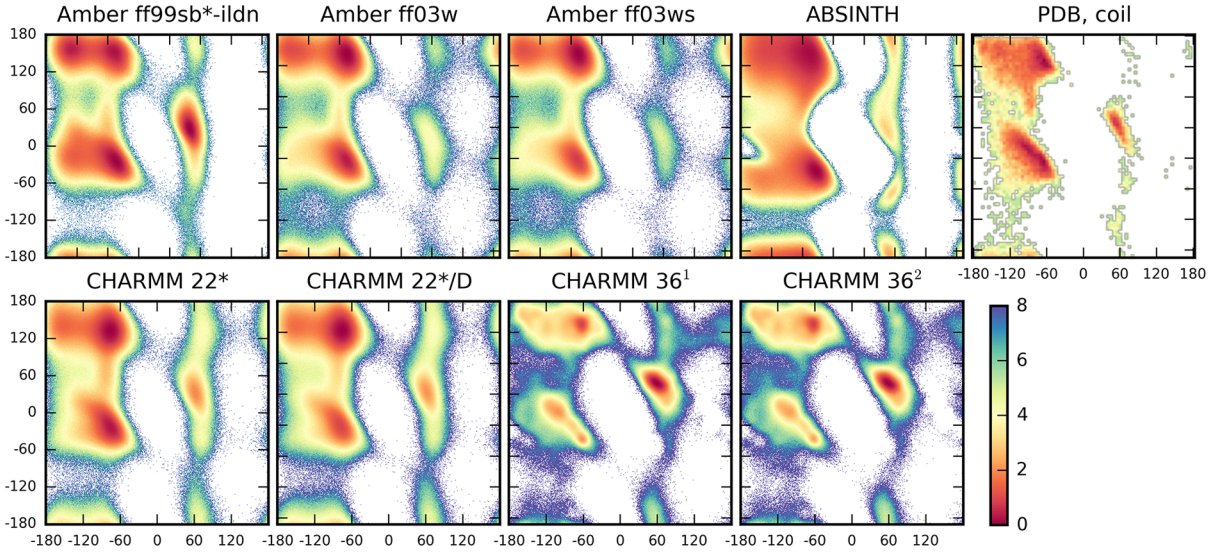

Figure 3. Force field differences in backbone dihedral angles. The potential of mean force (PMF), $-\log(P(\varphi,\psi))$, with color scale in kT, is shown for each force field. The upper right graph (PDB, coil) corresponds to all non-Pro, non-Gly residues that are not in helix or sheet secondary structure in the TOP500[90] structures (a set of nonredundant protein structures from the PDB). The population in the αL basin in each case is Amber ff99sb*-ildn = 0.12, Amber ff03w = 0.01, Amber ff03ws = 0.01, ABSINTH = 0.02, CHARMM 22* = 0.05, CHARMM 22* (with TIP4P-D) = 0.04, CHARMM 36 (with TIP3P) = 0.42, CHARMM 36 (with charmm-modified TIP3P) = 0.41, and PDB, coil = 0.06. More than 85% of conformations in the CHARMM 36 ensemble contain left-handed α-helix. Refer to Figures S6 and S7 for the same analysis for arginine and serine residues separately.

Ensembles Obtained Using CHARMM 36 Have a Bias toward a Left-Handed α-Helix. As can be seen in Figure 2B, the ensembles in most force fields have substantial content of bends and turns and relatively low α-helix and β-sheet content. There are two notable exceptions: CHARMM 36 with TIP3P and charmm-modified TIP3P; both of these ensembles have an exceptionally high α-helix content. Unexpectedly, this helix does not have the usual handedness expected for proteins; it is left-handed (a representative structure is shown in Figure 2C). Left-handed α-helices are exceedingly rare in structured proteins. Only 31 left-handed α-helices were found in a survey of 7984 structures from the PDB, and only 10 contained no glycines.[89] The maximum length of any of these was 6 residues; these helices are actually single helical turns, rather than extended helices. Longer left-handed α-helices, such as those seen in the ensembles with CHARMM 36, are essentially absent from structured proteins, based on known protein structures in the PDB.

Figure 3 compares the Ramachandran maps of all eight force fields to that of nonproline, nonglycine coil residues in the Top500[90] set of proteins (a nonredundant set of protein structures from the PDB). The population in the αL basin in the Top500 coil set is low (6%). Consistent with this population, most of the force fields also have low populations in αL, except for the CHARMM 36 ensembles, both of which have populations of more than 40%. No other force field has such a strong αL bias. In fact, more than 85% of the conformations in the CHARMM 36 ensemble contain a left-handed α-helix. Ramachandran maps of ensembles obtained using the same force field for the peptide, but which differ in water model used, are very similar (Figure 3). This result provides further independent support that sufficient conformational sampling has been achieved.

Structural Ensembles Obtained Using Different Force Fields Exhibit High Variability. Taken together, the results presented so far indicate the strong dependence of secondary structure and compactness on the choice of force field. For example, the ensemble in Amber ff99sb*-ildn has nearly four times as many intrapeptide hydrogen bonds compared to the Amber ff03ws ensemble. The choice of water model also has a significant effect on compactness, consistent with other studies.[26,51] The ensembles with CHARMM 36 are unique in their exceptionally high left-handed α-helix population. Although it is to be expected that the conformational ensemble of an IDP will depend, to some extent, on the chosen force field, the magnitude of the difference between force fields seen for the RS peptide is remarkable, spanning a range from globular to highly expanded. This finding is particularly surprising in light of the fact that many of these force fields have performed well in other benchmark studies.[26,35,51] Therefore, we next asked how well these ensembles compare to a diverse set of experimental data.

Assessing the Accuracy of Chain Dimensions: Comparison to Primary Experimental Data. For this aim, we measured SAXS data for the RS peptide, as described in the Experimental Methods. Figure 4a shows the measured SAXS scattering curve, along with the scattering curve calculated from the MD simulations using each force field. The curves of two ensembles, those obtained using CHARMM 22* and Amber ff03w, agree within error with the experimental data. The mean R_g and R_H of each ensemble are shown in Figure 4B and C, respectively. The mean R_g of the CHARMM 22* ensemble (12.65 ± 0.07 Å) is slightly smaller than that of the Amber ff03w ensemble (13.3 ± 0.1 Å) and agrees best with the measured R_g (12.62 ± 0.07 Å). These results are similar for both approaches to computing scattering curves from structures (results for both CRYSOL and FOXS are shown in Figure S8 and S9). Consistent with the results for radius of gyration, the CHARMM 22* ensemble is in closest agreement with the experimental hydrodynamic radius (11.95 ± 0.01 Å compared to the experimentally measured value of 11.9 ± 0.1 Å).

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524

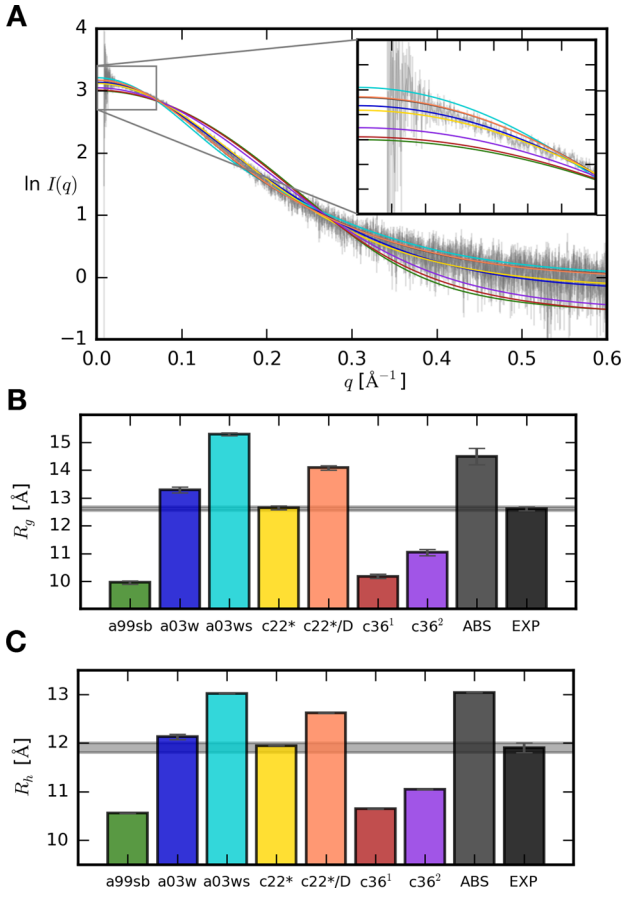

Figure 4. Comparison to chain dimensions measured by SAXS and PFG-NMR. (A) Shown are ensemble-averaged scattering curves for each force field (with the same color scheme as Figure 1). The experimental curve is shown with error in gray. The ensembles in two force fields agree with the experimental scattering curve within error: Amber ff03w and CHARMM 22*. (B) The radius of gyration is shown for each force field and the experimental data (black). (C) The hydrodynamic radius is shown for each force field, and the experimentally measured value is shown in black. The shaded gray lines in (B) and (C) indicate the experimental error.

The collapsed globule ensemble of Amber ff99sb*-ildn and the predominantly left-handed α-helix ensemble of CHARMM 36 are both too compact and inconsistent with the SAXS and PFG-NMR data. These results suggest that chain-chain interactions are too favorable compared to chain-water interactions in these force fields, or, alternatively, that the hydrophobic effect is too strong. A combination of these effects is likely, but our results do not allow us to assess the relative contributions of these effects.

In contrast, ABSINTH, Amber ff03ws, and CHARMM 22* with the TIP4P-D water model generate ensembles that are too expanded compared to the experimental R_g and R_H, suggesting that the balance is shifted too far in favor of chain-water interactions in these force fields. The implicit solvent model, ABSINTH, has previously been found to agree with measurements of R_H for other IDPs.[45] Amber ff03ws and CHARMM 22* with TIP4P-D were both developed with the specific aim of matching the experimental chain dimensions of IDPs.[26,51] The TIP4P-D water model has a Lennard-Jones C_6 parameter 50% larger than that of other water models and thus has increased dispersion interactions.[26] The fact that these force fields result in ensembles that are too expanded in comparison to experimental data for the RS peptide suggests that the problem of obtaining an accurate force field for IDPs may not be solved simply by modified water models parametrized specifically for IDPs.

Assessing the Accuracy of Secondary Structure Content: Comparison to $^3J_{HNH\alpha}$ Couplings. The CHARMM 22* and Amber ff03w ensembles agree well with SAXS and PFG-NMR data. These two ensembles are similar in compactness (Figure 1), but they differ significantly in secondary structure content (Figure 2). To assess the accuracy of secondary structure content, we compare to measured scalar couplings, which report on ensemble-averaged backbone and side chain dihedral angles. Figure 5 shows $^3J_{HNH\alpha}$ scalar

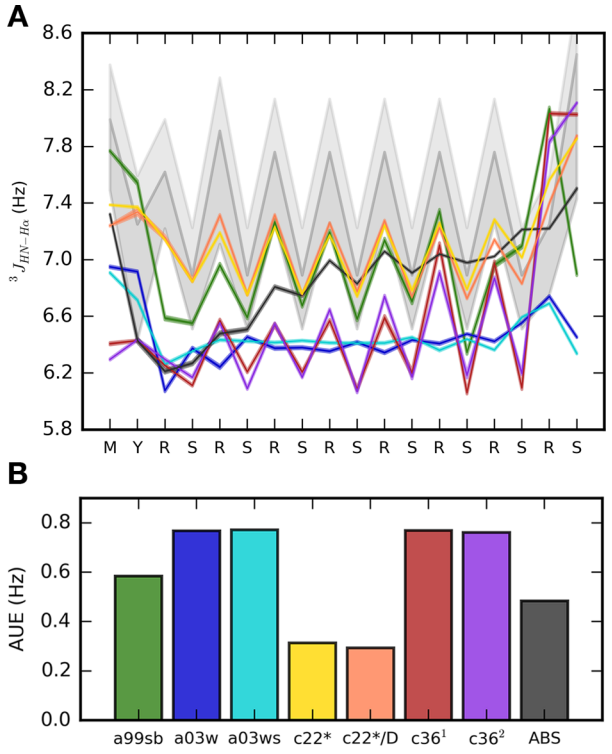

Figure 5. Comparison to measured $^3J_{HN\text{-}H\alpha}$ couplings. (A) The $^3J_{HN\text{-}H\alpha}$ couplings are shown for the ensemble obtained using each force field (the color scheme is consistent with Figures 1 and 4, shading indicates statistical uncertainty, which is on the order of the line thickness, ~0.01 Hz). The experimental $^3J_{HN\text{-}H\alpha}$ couplings (with error) are shown in gray shading; arginine and serine residues in the RS repeats have one measured $^3J_{HN\text{-}H\alpha}$ coupling. A correction to the $^3J_{HNH\alpha}$ scalar couplings has been suggested in the literature.[92] Such a correction would shift the true values of the scalar couplings to higher values compared to the experimentally measured values.[77] This additional uncertainty in the experimental values (estimated to be up to 5%[17]) is indicated in lighter gray. (B) The average unsigned error for the ensemble obtained using each force field is shown.

couplings, which report on the φ backbone dihedral angle and are the most-informative (and most-commonly used) for backbone conformation.[91] As can be seen, two force fields agree well with the measured couplings, CHARMM 22* with charmm-modified TIP3P and TIP4P-D water models. The $^3J_{HNH\alpha}$ scalar couplings for both CHARMM 36 ensembles are outside of experimental error. Because these couplings report on the φ dihedral angle, this finding also implies that the high left-handed α-helix population in CHARMM 36 is not consistent with the experimental data. For all three force fields

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524

for which different water models are tested (CHARMM 22*, Amber ff03w, and CHARMM 36), changing the water model has a relatively small impact on the $^3J_{HNH\alpha}$ scalar couplings. Obviously, the water model has a negligible effect on φ backbone dihedral preferences, which can also be seen in the Ramachandran plots in Figure 3.

Next, we compared the RS peptide ensembles to a large set of scalar couplings and chemical shifts as summarized in Figure 6 (raw data for each observable is provided in Figures S10–S18). Two more scalar couplings ($^1J_{C\alpha C\beta}$ and $^1J_{C\alpha H\alpha}$) report on the φ backbone dihedral angle. However, these scalar couplings are less informative than $^3J_{HNH\alpha}$. $^1J_{C\alpha C\beta}$ coupling constants have been found to depend more on amino acid composition than secondary structure preferences.[79] Nevertheless, as seen with the $^3J_{HNH\alpha}$ scalar couplings, the ensembles with CHARMM 22* are in closest agreement, and the CHARMM 36 ensembles do not agree well for both $^1J_{C\alpha C\beta}$ and $^1J_{C\alpha H\alpha}$. Comparison to side chain scalar couplings are also reported (Figures S12 and S13). The CHARMM 36 ensembles again show the largest errors, whereas the average unsigned errors for the other force fields are low (less than 0.4 Hz, the RMS prediction error). In general, the comparison to the experimental scalar couplings is limited primarily by the accuracy of computing scalar couplings with Karplus relations. The typical statistical uncertainty in the computed scalar couplings is only ~0.01 Hz (smaller than the line thickness in Figure 5) because of the extensive conformational sampling, which is much smaller than the RMS prediction errors. Thus, the observed deviations do not necessarily imply force field artifacts.

With respect to scalar couplings, CHARMM 36 ensembles deviate most from the experimental data, which is likely due to the high left-handed α-helix population. To investigate this further, we calculated the scalar couplings for a subensemble that was generated by selecting all conformations in the CHARMM 36 ensemble that contained no left-handed α-helix (Figure S15). This subensemble has significantly lower error compared to the experimental backbone scalar couplings, providing further evidence that the left-handed α-helix is inconsistent with the experimental data.

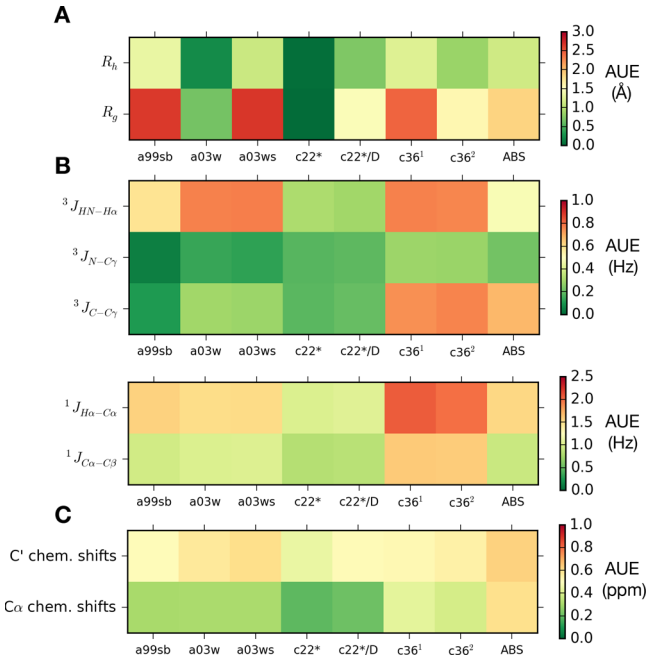

Figure 6. Comparison to NMR and SAXS experimental data: Summary. Average unsigned error (AUE) compared to experimental data in (A) the hydrodynamic radius, R_h, and the radius of gyration, R_g, (in Å), (B) J-couplings (in Hz), and (C) chemical shifts (in ppm) for each force field.

Comparison to Chemical Shifts. Carbonyl carbon and α-carbon chemical shifts have also been measured for the RS peptide.[16] Both of these types of chemical shifts are very sensitive to secondary structure.[93] The comparison to these chemical shifts is primarily limited by the expected accuracy of currently available chemical shift predictors. Within the expected error of SPARTA+, the chemical shifts of all of the ensembles agree with the experimental shifts. Chemical shifts calculated with SHIFTX2 have significantly smaller reported RMS errors than SPARTA+.[82] We report chemical shifts computed using both approaches as secondary chemical shifts (with the random coil values subtracted) in Figures S16–S18.

Only the ensemble obtained with CHARMM 22* (and charmm-modified TIP3P) agrees well with all available experimental data, except for residual dipolar couplings (RDCs). Consistent with earlier reports,[17,94,95] we find that the ensembles in all force fields are in poor agreement with RDCs (see Figure S19). It has been suggested that the conformational ensembles of IDPs may be affected by the presence of alignment media used in measurements of RDCs.[96] It is also well-established that IDPs are highly sensitive to solution conditions.[97,98] Therefore, it is still unclear from the present set of evidence whether the disagreement with RDCs comes from an effect of the alignment media on the ensembles or due to the force field. We are therefore carrying out a more detailed investigation into these effects. This will be published in a further study, as it goes significantly beyond the scope of the present work.

Generality and Causes of the Left-Handed Helix Propensity of CHARMM 36. We next investigated whether only the RS peptide forms a left-handed α-helix with CHARMM 36, or whether this force field has a general propensity for αL. To this end, simulations with $(A)_3$, $(AAQAA)_3$, and the HEWL peptide were performed. The Ramachandran plots for $(A)_3$ and $(AAQAA)_3$ (Figure S20) show a 6-fold lower αL propensity than seen for the RS peptide and are consistent with those reported by Best et al.[58] We also carried out microsecond simulations of the HEWL peptide. Approximately 30% of the ensemble of this peptide contains left-handed α-helix (Figure S21). At least 200 ns of simulation were needed before the left-handed helix formed, which is longer than earlier simulations of this peptide.[58] Thus, even though these proteins do not fold, long simulation times may be needed to find all highly populated conformational states.

The 16-residue FG-nucleoporin peptide (Figure 7a), as well as the longer 50-residue FG-nucleoporin peptide (Figure S22a), also showed a high population in the αL basin for the CHARMM 36 force field. A conformation of the FG-nucleoporin peptide containing helices of both handedness is shown (Figure S22b). Thus, the bias of CHARMM 36 toward left-handed α-helix is not limited to the RS peptide but rather seems to be quite general. This leads directly to the question: what causes the CHARMM 36 bias toward left-handed α-helix?

Left-handed α-helix is sterically disfavored by the close proximity of the side chain Cβ atom with the carbonyl C′ atom of the previous residue.[99] Comparing CHARMM 36 to its predecessor, CHARMM 22/CMAP, one of the modifications is

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513–5524

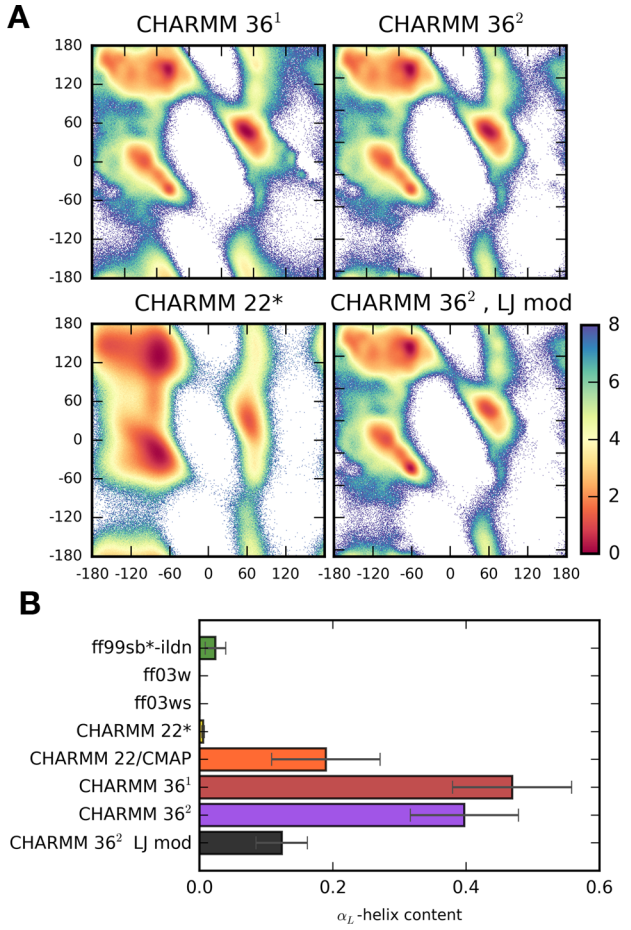

Figure 7. Modified aliphatic LJ parameters significantly reduce α_L-helix in CHARMM 36. (A) The potential of mean force (PMF), $-\log(P(\varphi,\psi))$, with color scale in kT, is shown for the FG-nucleoporin peptide in each force field. The population in the αL basin in each case is CHARMM 36[1] = 0.26; CHARMM 36[2] = 0.23; CHARMM 22* = 0.09; and CHARMM 36, LJ mod = 0.13. (B) The fraction of the ensemble containing left-handed α-helix is shown for each force field.

a change in the Lennard-Jones parameters of aliphatic hydrogen and carbon atoms, including the side chain Cβ atom.[58,100] In particular, the Lennard-Jones σ parameter was decreased. Because it is precisely a steric clash between this atom and the carbonyl C′ atom that precludes the αL conformation in nature, we tested the effect of using the aliphatic LJ parameters of the CHARMM 22/CMAP force field in CHARMM 36, which is here referred to as CHARMM 36, LJ mod.

In Figure 7a, Ramachandran plots are shown for the FG-nucleoporin peptide with CHARMM 36 as well as CHARMM 36, LJ mod. Indeed, the population in the αL basin is significantly reduced by this modification. Furthermore, the fraction of the ensemble containing left-handed α-helix is reduced from over 40% with CHARMM 36 to below 15% with CHARMM 36, LJ mod, which is even less left-handed α-helix than the CHARMM 22/CMAP ensemble (Figure 7b). It is therefore likely that a combination of a CMAP correction rendering the αL basin too favorable, combined with a decrease in the LJ σ parameter of the side chain Cβ atom, promotes αL formation. The CHARMM 36 subensemble containing no left handed α-helix has nearly the same radius of gyration as the entire ensemble (Figure S23). This result suggests that further modifications to the CHARMM 36 force field would be needed beyond adjusting the aliphatic LJ parameters; we did not pursue this direction further.

CONCLUSIONS

Obtaining accurate descriptions of IDPs by means of MD simulations is quite challenging both due to their sensitivity to force field inaccuracies as well as the need for extensive sampling. In addition, it is more demanding to obtain experimental information characterizing the entire ensemble of an IDP than the average structure of a folded protein because more experimental information is needed in the case of IDPs due to their high conformational heterogeneity. Furthermore, most experimental measurements obtained for IDPs are ensemble-averages, which only increases the challenge faced in characterizing IDPs.

Here, we assessed the accuracy of ensembles obtained from de novo simulations without the use of experimental information to guide ensemble selection or generation. Experimental data from SAXS and NMR were then used to evaluate the accuracy of the ensembles obtained using eight state-of-the-art force fields. Note that this study is not a comprehensive test of all available force field combinations. Sufficient conformational sampling was obtained using replica exchange. Taken together, our results demonstrate an unexpectedly high sensitivity of IDP conformational ensembles to differences between the force fields. For example, for CHARMM 36, an unexpectedly high propensity for left-handed α-helix was found, which was strongly reduced by relatively small changes in the Lennard-Jones parameters of aliphatic carbons. A comparison of the secondary structure content of the RS and FG peptides showed that, for these two very different IDPs, force field is a stronger determinant of secondary structure content than peptide sequence. Major differences in chain dimensions were also found: ensembles span the entire range from collapsed globule-like to highly expanded. There was consensus, though, among all of the force fields that the studied peptides are disordered (that is, populating many conformations rather than a single, well-defined native structure). Even this similarity is tenuous because the vast majority of the CHARMM 36 ensemble contains stretches of left-handed α-helix.

A key finding of this study is that the conformational ensemble obtained using CHARMM 22* with charmm-modified TIP3P agrees best with all available experimental data. Maintaining the correct balance between solvent−solvent, solvent−solute, and solute−solute interactions has been an optimization criterion in the development of the CHARMM force fields, even in much earlier versions.[31,101] With a correct balance of interactions, the quality of water as a solvent for proteins should be well described, and this indeed appears to be the case for the RS peptide.

This study provides a systematic benchmark of eight force fields for a set of IDPs. Further computational-experimental studies of other IDPs are a necessary next step to delineate the accuracy of force fields for simulations of IDPs of different length and sequence composition.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00736.

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524

Force field ensembles, contact maps, example inverse claw conformation, secondary structure content for different sequences and force fields, force field differences in backbone dihedral angles, comparison to SAXS, scalar couplings, chemical shifts and RDCs, Ramachandran plots, list of simulations, Monte Carlo move set, and Karplus relations used to calculate the scalar couplings (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: sarah.rauscher@mpibpc.mpg.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Robert Best for providing the GROMACS ports for Amber ff03w and Amber ff03ws and Alexander MacKerell and Michael Feig for discussions on CHARMM 36 development. S.R. is supported by a postdoctoral fellowship from the Alexander von Humboldt Foundation. M.Z. was supported by the DFG Collaborative Research Center 860, Project B2. Compute time was provided through an allocation by the Gauss Supercomputing Center on the SuperMUC supercomputer at the Leibniz Rechenzentrum in Garching.

REFERENCES

(1) van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; et al. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114*, 6589−6631.

(2) Tompa, P.; Davey, N. E.; Gibson, T. J.; Babu, M. M. A Million Peptide Motifs for the Molecular Biologist. *Mol. Cell* **2014**, *55*, 161−169.

(3) Lim, R. Y. H.; Fahrenkrog, B.; Köser, J.; Schwarz-Herion, K.; Deng, J.; Aebi, U. Nanomechanical Basis of Selective Gating by the Nuclear Pore Complex. *Science* **2007**, *318*, 640−643.

(4) Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D.; et al. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **2015**, *57*, 936−947.

(5) Weber, S. C.; Brangwynne, C. P. Getting RNA and Protein in Phase. *Cell* **2012**, *149*, 1188−1191.

(6) Jensen, M. R.; Zweckstetter, M.; Huang, J. R.; Blackledge, M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chem. Rev.* **2014**, *114*, 6632−6660.

(7) Rauscher, S.; Pomès, R. Molecular Simulations of Protein Disorder. *Biochem. Cell Biol.* **2010**, *88*, 269−290.

(8) Varadi, M.; Kosol, S.; Lebrun, P.; Valentini, E.; Blackledge, M.; Dunker, A. K.; Felli, I. C.; Forman-Kay, J. D.; Kriwacki, R. W.; Pierattelli, R.; et al. pE-DB: a Database of Structural Ensembles of Intrinsically Disordered and of Unfolded Proteins. *Nucleic Acids Res.* **2014**, *42*, D326−D335.

(9) Brucale, M.; Schuler, B.; Samorì, B. Single-Molecule Studies of Intrinsically Disordered Proteins. *Chem. Rev.* **2014**, *114*, 3281−3317.

(10) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry* **1999**, *38*, 16424−16431.

(11) Camilloni, C.; Vendruscolo, M. Statistical Mechanics of the Denatured State of a Protein Using Replica-Averaged Metadynamics. *J. Am. Chem. Soc.* **2014**, *136*, 8982−8991.

(12) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Mapping Long-Range Interactions in A-Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005**, *127*, 476−477.

(13) Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M. Determination of the Free Energy Landscape of α-Synuclein Using Spin Label Nuclear Magnetic Resonance Measurements. *J. Am. Chem. Soc.* **2009**, *131*, 18314−18326.

(14) Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W. Y.; Forman-Kay, J. D. Characterization of Disordered Proteins with ENSEMBLE. *Bioinformatics* **2013**, *29*, 398−399.

(15) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *132*, 8407−8418.

(16) Xiang, S.; Gapsys, V.; Kim, H. Y.; Bessonov, S.; Hsiao, H. H.; Möhlmann, S.; Klaukien, V.; Ficner, R.; Becker, S.; Urlaub, H.; et al. Phosphorylation Drives a Dynamic Switch in Serine/Arginine-Rich Proteins. *Structure* **2013**, *21*, 2162−2174.

(17) Ball, K. A.; Wemmer, D. E.; Head-Gordon, T. Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-β Peptides. *J. Phys. Chem. B* **2014**, *118*, 6405−6416.

(18) Gurry, T.; Ullman, O.; Fisher, C. K.; Perovic, I.; Pochapsky, T.; Stultz, C. M. The Dynamic Structure of α -Synuclein Multimers. *J. Am. Chem. Soc.* **2013**, *135*, 3865−3872.

(19) Schwalbe, M.; Ozenne, V.; Bibow, S.; Jaremko, M.; Jaremko, L.; Gajda, M.; Jensen, M. R.; Biernat, J.; Becker, S.; Mandelkow, E.; et al. Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and α-Synuclein in Solution From NMR and Small Angle Scattering. *Structure* **2014**, *22*, 238−249.

(20) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüllller, H.; Griesinger, C.; de Groot, B. L. Recognition Dynamics Up to Microseconds Revealed From an RDC-Derived Ubiquitin Ensemble in Solution. *Science* **2008**, *320*, 1471−1475.

(21) Vitalis, A.; Pappu, R. V. ABSINTH: a New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions. *J. Comput. Chem.* **2009**, *30*, 673−699.

(22) Smith, W. W.; Schreck, C. F.; Hashem, N.; Soltani, S.; Nath, A.; Rhoades, E.; O'Hern, C. S. Molecular Simulations of the Fluctuating Conformational Dynamics of Intrinsically Disordered Proteins. *Phys. Rev. E* **2012**, *86*, 041910.

(23) Yedvabny, E.; Nerenberg, P. S.; So, C.; Head-Gordon, T. Disordered Structural Ensembles of Vasopressin and Oxytocin and Their Mutants. *J. Phys. Chem. B* **2015**, *119*, 896−905.

(24) Stanley, N.; Esteban-Martin, S.; De Fabritiis, G. Kinetic Modulation of a Disordered Protein Domain by Phosphorylation. *Nat. Commun.* **2014**, *5*, 5272.

(25) Gerben, S. R.; Lemkul, J. A.; Brown, A. M.; Bevan, D. R. Comparing Atomistic Molecular Mechanics Force Fields for a Difficult Target: a Case Study on the Alzheimer's Amyloid β-Peptide. *J. Biomol. Struct. Dyn.* **2014**, *32*, 1817−1832.

(26) 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.

(27) 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.

(28) Sgourakis, N. G.; Merced-Serrano, M.; Boutsidis, C.; Drineas, P.; Du, Z.; Wang, C.; García, A. E. Atomic-Level Characterization of the Ensemble of the Aβ(1−42) Monomer in Water Using Unbiased Molecular Dynamics Simulations and Spectral Algorithms. *J. Mol. Biol.* **2011**, *405*, 570−583.

(29) Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. Structure and Dynamics of the Aβ 21−30 Peptide From the Interplay of NMR Experiments and Molecular Simulations. *J. Am. Chem. Soc.* **2008**, *130*, 6145−6158.

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524

(30) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585−590.
(31) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(32) 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.
(33) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47−L49.
(34) 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.
(35) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields Against Experimental Data. *PLoS One* **2012**, *7*, e32131.
(36) Beauchamp, K. A.; Lin, Y. S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409−1414.
(37) Lange, O. F.; van der Spoel, D.; de Groot, B. L. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J.* **2010**, *99*, 647−655.
(38) Cino, E. A.; Choy, W. Y.; Karttunen, M. Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 2725−2740.
(39) Petrov, D.; Zagrovic, B. Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? *PLoS Comput. Biol.* **2014**, *10*, e1003638.
(40) Sethi, A.; Tian, J.; Vu, D. M.; Gnanakaran, S. Identification of Minimally Interacting Modules in an Intrinsically Disordered Protein. *Biophys. J.* **2012**, *103*, 748−757.
(41) Levine, Z. A.; Larini, L.; LaPointe, N. E.; Feinstein, S. C.; Shea, J. E. Regulation and Aggregation of Intrinsically Disordered Peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2758−2763.
(42) Pantelopulos, G. A.; Mukherjee, S.; Voelz, V. A. Microsecond Simulations of Mdm2 and Its Complex with p53 Yield Insight Into Force Field Accuracy and Conformational Dynamics. *Proteins: Struct., Funct., Genet.* **2015**, *83*, 1665−1676.
(43) Hoffmann, K. Q.; McGovern, M.; Chiu, C.-C.; de Pablo, J. J. Secondary Structure of Rat and Human Amylin Across Force Fields. *PLoS One* **2015**, *10*, e0134091.
(44) Rosenman, D. J.; Connors, C. R.; Chen, W.; Wang, C.; García, A. E. Aβ Monomers Transiently Sample Oligomer and Fibril-Like Configurations: Ensemble Characterization Using a Combined MD/NMR Approach. *J. Mol. Biol.* **2013**, *425*, 3338−3359.
(45) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. Net Charge Per Residue Modulates Conformational Ensembles of Intrinsically Disordered Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8183−8188.
(46) Fluitt, A. M.; de Pablo, J. J. An Analysis of Biomolecular Force Fields for Simulations of Polyglutamine in Solution. *Biophys. J.* **2015**, *109*, 1009−1018.
(47) Best, R. B.; Mittal, J. Free-Energy Landscape of the GB1 Hairpin in All-Atom Explicit Solvent Simulations with Different Force Fields: Similarities and Differences. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 1318−1328.
(48) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the Accuracy of Physical Models Used in Protein-Folding Simulations: Quantitative Evidence From Long Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98−105.
(49) Skinner, J. J.; Yu, W.; Gichana, E. K.; Baxa, M. C.; Hinshaw, J. R.; Freed, K. F.; Sosnick, T. R. Benchmarking All-Atom Simulations Using Hydrogen Exchange. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15975−15980.
(50) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. Structure and Dynamics of an Unfolded Protein Examined by Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **2012**, *134*, 3787−3791.
(51) 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.
(52) Mercadante, D.; Milles, S.; Fuertes, G.; Svergun, D. I.; Lemke, E. A.; Gräter, F. Kirkwood−Buff Approach Rescues Overcollapse of a Disordered Protein in Canonical Protein Force Fields. *J. Phys. Chem. B* **2015**, *119*, 7975−7984.
(53) Henriques, J.; Cragnell, C.; Skepö, M. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 3420−3431.
(54) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(55) 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.* **1983**, *79*, 926−935.
(56) Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.
(57) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(58) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ1and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(59) Nehrbass, U.; Kern, H.; Mutvei, A.; Horstmann, H.; Marshallsay, B.; Hurt, E. C. NSP1: a Yeast Nuclear Envelope Protein Localized at the Nuclear Pores Exerts Its Essential Function by Its Carboxy-Terminal Domain. *Cell* **1990**, *61*, 979−989.
(60) Patel, S. S.; Belmont, B. J.; Sante, J. M.; Rexach, M. F. Natively Unfolded Nucleoporins Gate Protein Diffusion Across the Nuclear Pore Complex. *Cell* **2007**, *129*, 83−96.
(61) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(62) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(63) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(64) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(65) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(66) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: a New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(67) Das, R. K.; Pappu, R. V. Conformations of Intrinsically Disordered Proteins Are Influenced by Linear Sequence Distributions of Oppositely Charged Residues. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13392−13397.
(68) Mao, A. H.; Pappu, R. V. Crystal Lattice Properties Fully Determine Short-Range Interaction Parameters for Alkali and Halide Ions. *J. Chem. Phys.* **2012**, *137*, 064104.
(69) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dyanmics. *J. Mol. Graphics* **1996**, *14*, 33−38.

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524

(70) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.
(71) Flyvbjerg, H.; Petersen, H. G. Error Estimates on Averages of Correlated Data. *J. Chem. Phys.* **1989**, *91*, 461−466.
(72) Svergun, D.; Barberato, C.; Koch, M. H. J. CRYSOL - a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules From Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28*, 768−773.
(73) 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.
(74) Chen, P.-C.; Hub, J. S. Validating Solution Ensembles From Molecular Dynamics Simulation by Wide-Angle X-Ray Scattering Data. *Biophys. J.* **2014**, *107*, 435−447.
(75) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. *Primus*: A Windows PC-Based System for Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2003**, *36*, 1277−1282.
(76) Ortega, A.; Amorós, D.; García de la Torre, J. Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins From Atomic- and Residue-Level Models. *Biophys. J.* **2011**, *101*, 892−898.
(77) Vuister, G. W.; Bax, A. Quantitative J Correlation: a New Approach for Measuring Homonuclear Three-Bond J(HNHα) Coupling Constants in ^{15}N-Enriched Proteins. *J. Am. Chem. Soc.* **1993**, *115*, 7772−7777.
(78) Cornilescu, G.; Bax, A.; Case, D. A. Large Variations in One-Bond 13Cα− 13Cβ J Couplings in Polypeptides Correlate with Backbone Conformation. *J. Am. Chem. Soc.* **2000**, *122*, 2168−2171.
(79) Schmidt, J. M.; Howard, M. J.; Maestre-Martínez, M.; Pérez, C. S.; Löhr, F. Variation in Protein Cα-Related One-Bond J couplings. *Magn. Reson. Chem.* **2009**, *47*, 16−30.
(80) Schmidt, J. M. Asymmetric Karplus Curves for the Protein Side-Chain ^{3}J Couplings. *J. Biomol. NMR* **2007**, *37*, 287−301.
(81) Vuister, G. W.; Delaglio, F.; Bax, A. The Use of $^1J_{C\alpha H\alpha}$ Coupling Constants as a Probe for Protein Backbone Conformation. *J. Biomol. NMR* **1993**, *3*, 67−80.
(82) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. SHIFTX2: Significantly Improved Protein Chemical Shift Prediction. *J. Biomol. NMR* **2011**, *50*, 43−57.
(83) Shen, Y.; Bax, A. SPARTA+: a Modest Improvement in Empirical NMR Chemical Shift Prediction by Means of an Artificial Neural Network. *J. Biomol. NMR* **2010**, *48*, 13−22.
(84) Roessle, M. W.; Klaering, R.; Ristau, U.; Robrahn, B.; Jahn, D.; Gehrmann, T.; Konarev, P.; Round, A.; Fiedler, S.; Hermes, C.; et al. Upgrade of the Small-Angle X-Ray Scattering Beamline X33 at the European Molecular Biology Laboratory, Hamburg. *J. Appl. Crystallogr.* **2007**, *40*, s190−s194.
(85) Round, A. R.; Franke, D.; Moritz, S.; Huchler, R.; Fritsche, M.; Malthan, D.; Klaering, R.; Svergun, D. I.; Roessle, M. Automated Sample-Changing Robot for Solution Scattering Experiments at the EMBL Hamburg SAXS Station X33. *J. Appl. Crystallogr.* **2008**, *41*, 913−917.
(86) Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I. New Developments in the ATSAS Program Package for Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2012**, *45*, 342−350.
(87) Hamelberg, D.; Shen, T.; McCammon, J. A. A Proposed Signaling Motif for Nuclear Import in mRNA Processing via the Formation of Arginine Claw. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14947−14951.
(88) Sellis, D.; Drosou, V.; Vlachakis, D.; Voukkalis, N.; Giannakouros, T.; Vlassi, M. Phosphorylation of the Arginine/Serine Repeats of Lamin B Receptor by SRPK1—Insights From Molecular Dynamics Simulations. *Biochim. Biophys. Acta, Gen. Subj.* **2012**, *1820*, 44−55.
(89) Novotny, M.; Kleywegt, G. J. A Survey of Left-Handed Helices in Protein Structures. *J. Mol. Biol.* **2005**, *347*, 231−241.
(90) Lovell, S. C.; Davis, I. W.; Arendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by Cα Geometry: ϕ, ψ and Cβ Deviation. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 437−450.
(91) Otten, R.; Wood, K.; Mulder, F. A. A. Comprehensive Determination of 3JHNHα for Unfolded Proteins Using 13C′-Resolved Spin-Echo Difference Spectroscopy. *J. Biomol. NMR* **2009**, *45*, 343−349.
(92) Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T. Differences in β-Strand Populations of Monomeric Aβ40 and Aβ42. *Biophys. J.* **2013**, *104*, 2714−2724.
(93) Wang, Y.; Jardetzky, O. Probability-Based Protein Secondary Structure Identification Using Combined NMR Chemical-Shift Data. *Protein Sci.* **2002**, *11*, 852−861.
(94) Wang, Y.; Chu, X.; Longhi, S.; Roche, P.; Han, W.; Wang, E.; Wang, J. Multiscaled Exploration of Coupled Folding and Binding of an Intrinsically Disordered Molecular Recognition Element in Measles Virus Nucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E3743−E3752.
(95) Wang, Y.; Longhi, S.; Roche, P.; Wang, J. Reply to Jensen and Blackledge: Dual Quantifications of Intrinsically Disordered Proteins by NMR Ensembles and Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E1559−E1559.
(96) Montalvao, R. W.; De Simone, A.; Vendruscolo, M. Determination of Structural Fluctuations of Proteins From Structure-Based Calculations of Residual Dipolar Couplings. *J. Biomol. NMR* **2012**, *53*, 281−292.
(97) Wuttke, R.; Hofmann, H.; Nettels, D.; Borgia, M. B.; Mittal, J.; Best, R. B.; Schuler, B. Temperature-Dependent Solvation Modulates the Dimensions of Disordered Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 5213−5218.
(98) Soranno, A.; Koenig, I.; Borgia, M. B.; Hofmann, H.; Zosel, F.; Nettels, D.; Schuler, B. Single-Molecule Spectroscopy Reveals Polymer Effects of Disordered Proteins in Crowded Environments. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 4874−4879.
(99) Finkelstein, A. V.; Ptitsyn, O. *Protein Physics: A Course of Lectures*, 1st ed.; Academic Press, 2002.
(100) Vorobyov, I. V.; Anisimov, V. M.; MacKerell, A. D. Polarizable Empirical Force Field for Alkanes Based on the Classical Drude Oscillator Model. *J. Phys. Chem. B* **2005**, *109*, 18988−18999.
(101) Mackerell, A. D., Jr.; Wiórkiewicz-Kuczera, J.; Karplus, M. An All-Atom Empirical Energy Function for the Simulation of Nucleic Acids. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.

DOI: 10.1021/acs.jctc.5b00736
J. Chem. Theory Comput. 2015, 11, 5513−5524