Balanced amino-acid specific molecular dynamics force field for the realistic simulation of both folded and disordered proteins

Lei Yu,[1] Da-Wei Li,[2] and Rafael Brüschweiler[1,2,3*]

[1]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA

[2]Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, USA

[3]Department of Biological Chemistry and Pharmacology, The Ohio State University, Columbus, Ohio 43210, USA

[*]To whom correspondence should be addressed:

Rafael Brüschweiler, Ph.D., E-mail: bruschweiler.1@osu.edu

Abstract

Molecular dynamics (MD) simulations provide a unique atomic-level description of the structure and dynamics of proteins, which is essential for the mechanistic understanding of protein interactions and function in living organisms. However, traditional MD force fields that are optimized for folded proteins often generate overly compact structures and incorrect characteristics of intrinsically disordered proteins (IDPs) and protein regions (IDRs), thereby limiting the quantitative insights that can be gained from MD simulations. We introduce the residue-specific protein force field, ff99SBnmr2, which is derived from ff99SBnmr1 by balancing the backbone dihedral angle potentials in a residue-specific manner to quantitatively reproduce dihedral angle distributions from an experimental coil library. The new force field substantially improves the backbone conformational ensembles of disordered proteins, protein regions, and peptides while keeping well-defined protein structures stable and accurate. This balanced new force field should enable a myriad of applications that require quantitative descriptions of IDPs, IDRs, loop dynamics, and folding/unfolding equilibria in the presence and absence of interaction partners.

INTRODUCTION

Molecular dynamics (MD) computer simulations are increasingly popular to produce an atomic-level description of the structures and dynamics of proteins, which is essential for a proper understanding of atomistic aspects of their function.[1] With ever increasing computational power, all-atom simulations are now routinely performed on the microsecond timescale using high-performance computer clusters and sometimes even significantly beyond that timescale with the help of specialized engines.[2] Overall, molecular mechanics force fields have reached a level of accuracy allowing the *in silico* folding of small fast-folding proteins from extended states into structures that are within atomic resolution of their native structures.[3] Protein ensembles generated from MD simulations provide structural and dynamic properties that are increasingly consistent with experimental data measured by a variety of biophysical techniques, such as small-angle X-ray scattering (SAXS), Förster resonance energy transfer (FRET), and nuclear magnetic resonance (NMR) spectroscopy.[4-7] Although many force fields have proven effective for folded proteins, they often yield unsatisfactory results for an important class of proteins known as intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDRs).[8-11] IDPs and IDRs include proteins and protein regions that do not adopt well-defined 3D structures in their native state. As a result, IDPs mainly populate the "flatter" parts of the potential energy surface compared to globular proteins, allowing them to sample a much broader range of conformations under physiological conditions. Widely encoded in human genes and linked to a variety of human diseases, IDPs/IDRs play crucial biological roles, such as molecular recognition and post-translational modification.[12-13] Currently, a major obstacle in the field is the limited understanding of the structural propensities and short- vs. long-range dynamics of IDPs, hindering our understanding of the molecular mechanism(s) underlying their biological functions.

Traditional force fields optimized for folded proteins often yield unsatisfactory results for IDPs from both a local and a global perspective: they tend to generate overly compact structures and display incorrect local structural characteristics. The radii of gyration of unfolded proteins are significantly underestimated by traditional protein force fields, such as CHARMM22* and AMBER ff99SB in the presence of the explicit TIP3P water solvent.[11] Moreover, the structural ensembles of IDPs derived from simulations highly depend on the specific choice of force fields.[14] The polypeptide chain dimensions, hydrogen bonding, and secondary structure content of IDPs vary widely across different protein force fields and explicit water models. This result indicates that most of these force fields, despite their ability to model various aspects of folded proteins correctly, fail to reproduce the structural and dynamic content of IDPs.

Water models play an important role in protein dynamics simulations and researchers have been focusing on developing balanced protein-water interactions to improve the agreement between simulations and experiments.[15-17] For example, the TIP4P-D water model has been notably successful in reproducing the "extendedness" of IDPs in simulations.[18] This water model corrects London dispersion interactions that are typically underestimated in previous water models. By replacing TIP3P with TIP4P-D, several traditional protein force fields were able to improve their performance of IDP simulations. For globular proteins, the TIP4P-D water model is also able to capture the rotational tumbling of globular proteins better than the TIP3P.[19] However, it has been shown that the TIP4P-D water model tends to destabilize folded structures decreasing the folded population of small folded peptides and proteins.[20]

Protein force field developments are continuing in all major MD force field families as this is a central goal in the pursuit of quantitative molecular biophysics. In AMBER ff14SB,[21] side-chain and backbone parameters were modified to improve the secondary structure content of small peptides and the agreement with NMR χ_1 scalar *J*-coupling constants. In the CHARMM force field family, the most recent force field C36m[7] generates more realistic ensembles for both IDPs and folded proteins by mainly reducing the elevated propensity for left-handed α-helical (α_L) conformations. Alternatively, IDP-specific force fields are being developed by correcting protein conformations in a residue-specific manner using information from experimental coil libraries. Several research groups have adjusted force field parameters, such as local Lennard-Jones interactions and dihedral angle potentials, to fit dihedral angle distributions from experimental coil libraries.[22-23] In search for a unified force field, Robustelli and coworkers conducted an extensive benchmark study on a large selection of modern protein force fields. Because none of them satisfactorily performed for both folded and disordered proteins, they developed a new protein force field a99SB-*disp*.[24] The improved balance of a99SB-*disp* has since been confirmed for both folded and partially disordered systems.[25-26]

In this work, we adopt a hybrid approach by adjusting the overall backbone dihedral angle potentials using extensive experimental coil library information, while retaining the local features within the major Ramachandran regions of our prior force field AMBER ff99SBnmr1 (nmr1)[27] and keeping the force field non-polarizable with all charges fixed[28]. The ff99SBnmr1 force field has been shown to perform particularly well for folded proteins as it was specifically optimized using experimental NMR chemical shift data of intact full-length proteins.[6] By making residue-specific modifications, our aim is to make this force field more suitable for simulating disordered proteins and highly dynamic protein regions, including flexible loops and tails. Through extensive validations against experimental data and comparisons with other force fields, we show that the new force field ff99SBnmr2 (nmr2) is able to provide realistic ensemble descriptions for both folded and disordered proteins which are on par or better than those of other state-of-the-art force fields.

METHODS

Coil libraries for the backbone φ,ψ dihedral angle distributions of disordered polypeptides

The coil library used here has been curated from protein structural fragments contained in the Protein Data Bank (PDB), which are neither α-helices nor β-strands.[29] Coil library entries were retrieved from the Protein Coil Library webserver (http://folding.chemistry.msstate.edu/coil/index.html)[29] based on the following criteria: less than 90% sequence identity, from X-ray crystal structures with < 2.0 Å resolution and an R-factor of 0.25 or better. All residues characterized as turns and those that are succeeded by proline residues were also removed.[30] Backbone dihedral angles φ and ψ were then calculated from the atomic coordinates and sorted by residue types. For each residue type, relative populations of four main regions in the Ramachandran plot were determined, namely α_R ($-180° < \varphi < 0°$, $-120° < \psi < 50°$), α_L ($0° < \varphi < 180°$, $-120° < \psi < 240°$), β ($-180° < \varphi < -90°$, $50° < \psi < 240°$), and PP_{II} ($-90° < \varphi < 0°$, $50° < \psi < 240°$) (Figure 1A) and they are reported in Table S1 in the Supporting Information.

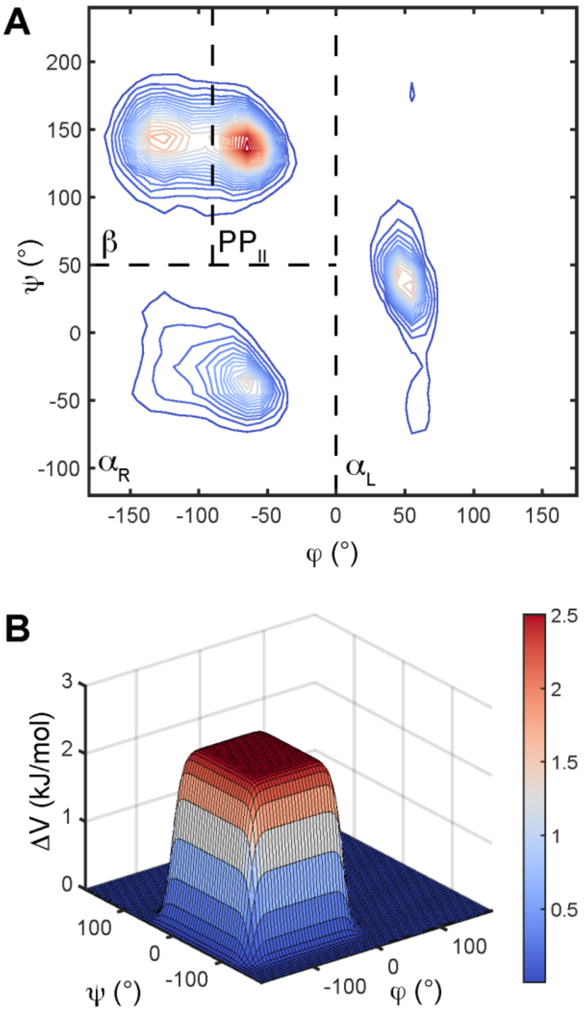


Figure 1. Backbone dihedral angle distributions and corrections. (A) An example of φ,ψ distribution of Asn from the coil library relative to the four main regions of the Ramachandran space (α_R, α_L, β, and PP_{II}). The contours from blue to red depict regions with increasing propensity. (B) Two-dimensional hyperbolic tangent functions were used to construct corrections with a "table-mountain profile" to the dihedral angle potentials. These correction potentials are flat within the region they are applied to, thereby retaining the fine features of the original nmr1 protein potential.

Modification of the protein potential

Dipeptides in the form of ACE-X-NME, where X is one of the 20 amino acid types and ACE and NME are acetyl and *N*-methyl amide capping groups, respectively, were simulated in AMBER ff99SBnmr1/TIP4P-D for 1 μs. For the central residue in each dipeptide, the population *p* from the simulation of any of the Ramachandran regions was subsequently compared with the one obtained from the coil library p_0. The additive correction of potential energy in the dihedral potential ΔV for each region was estimated using the Boltzmann relationship:

$$\Delta V = k_{\mathrm{B}} T \ln(p / p_0) \qquad (1)$$

where k_{B} is the Boltzmann constant and T is the absolute temperature. Residue-specific torsional correction maps (CMAPs) were finally constructed using hyperbolic tangent functions taking the form of "table-mountain" profiles (Figure 1B) and applied to the nmr1 force field, which is the target of the force field modification. The process was performed iteratively until the Ramachandran populations from the coil library were accurately reproduced.

Molecular dynamics simulations

MD simulations were performed using the GROMACS 5.1.2 package.[31] Initial structures of folded proteins were built based on PDB deposits and structures of disordered peptides were prepared using the LEaP program in AmberTools16[32] (see Table S2). The integration time step was set to 2 fs with all bond lengths containing hydrogen atoms constrained by the LINCS algorithm. Na^+ or Cl^- ions were added to neutralize the total charge of the system. A 10 Å cutoff was used for all van der Waals and electrostatic interactions. Particle-Mesh Ewald summation with a grid spacing of 1.2 Å was used to calculate long-range electrostatic interactions. A cubic simulation box that extends 8 Å from the protein surface was used, and periodic boundary conditions were applied along all three dimensions. Energy minimization was performed using the steepest descent algorithm for 50,000 steps. The system was then simulated for 100 ps at constant temperature and constant volume with all protein heavy atoms positionally fixed. Next, the pressure was coupled to 1 atm and the system was simulated for another 100 ps. The final production run was performed in the NPT ensemble at 300 K and 1 atm. MD simulations were performed for 1 μs each using the following combinations of protein force fields and water models: AMBER ff14SB/TIP3P, CHARMM36m/CHARMM-modified TIP3P, AMBER ff99SBnmr1/TIP4P-D, and the coil library-modified force field AMBER ff99SBnmr2/TIP4P-D. In the folding studies, simulations were run for extended simulation times at the experimentally reported melting temperatures. Cα root-mean-square deviations (RMSDs) were calculated comparing conformations in the simulations with the crystal structures. Replica exchange MD simulations were performed for chignolin using nmr2/TIP4P-D (14 replicas, 300 K – 360 K) and nmr2/TIP3P (18 replicas, 300 K – 380 K) with each replica being 1 μs of length. Exchange was attempted every 10 ps and exchange probability was about 0.34. The folding curves were then calculated using the converged trajectories with folded states being defined as structures with RMSD < 2.0 Å. Additional information regarding the simulations is listed in Table S3.

Back-calculation of NMR parameters

NMR ^3J-coupling constants were back-calculated from MD simulations using an appropriate Karplus equation[33] and the ensemble averages were compared with experimentally determined values. Chemical shifts of Cα, Cβ, C', H^N, N, Hα were also predicted from MD trajectories by PPM_One[34] using the static parameter set and RMSDs normalized by predictor errors were calculated to assess the agreement between the ensemble descriptions derived from experiments and simulations. Moreover, residual dipolar coupling (RDC) constants were predicted from simulations by singular value decomposition[35] and the corresponding Q factors (normalized RMSDs) were calculated.[36] ^{15}N-^{1}H NMR S^2 order parameters were predicted from simulations using the isotropic reorientational eigenmode dynamics (iRED)[37] method with a proper length of averaging windows. All experimental NMR parameters used in this study are listed in Table S2. Standard deviations were calculated by dividing each trajectory into 4 blocks of equal length that were analyzed separately.

RESULTS

Backbone dihedral angle distributions of disordered residues

Except for glycine and proline, the dihedral angle distributions of amino acid residues are dominated by α_R, β, and PP_{II} populations (Figure S1). However, relative populations of α_R and β vary from one residue type to another. The previous nmr1 force field generally overestimates the α_R populations, while underestimating β and α_L populations in a number of cases. In addition, significantly underestimated α_L populations for the disorder promoting residues Gly, Arg, Gln, and Lys[38] may be partially responsible for the unsatisfactory performance of traditional force fields in IDP simulations. After modification, the new nmr2 force field achieves semi-quantitative to quantitative agreement between the simulation and the coil library indicating that the dihedral angle potentials of the modified force field now satisfactorily reflect those from the coil library in a residue-specific manner.

^{3}*J*-couplings of disordered peptides

Vicinal $^3J(H^N,H\alpha)$ coupling constants are usually measured with high precision[39] and are directly linked to the backbone φ-dihedral angles via Karplus equations. Thus, the agreement between experimental and back-calculated ^3J-coupling constants is indicative of the quality of the backbone conformational ensemble of a protein. For the dipeptides[40] whose dihedral angle distributions were used for the development of the new force field nmr2, an improved RMSD of 0.31 Hz was obtained as compared to 0.52 Hz by the prior nmr1 force field (Figure 2A). For the middle three residues in each of the 45 capped heptapeptides from the α-synuclein[41] sequence, nmr2 yielded a better RMSD of 0.46 Hz (vs. 0.74 Hz for nmr1) and a better Pearson correlation coefficient R = 0.80 (vs. 0.50 for nmr1) between the experimental and predicted *J*-coupling constants (Figure 2B). Both cases, using the residue-specific average *J*-coupling constants from the coil library as a predictor, gave RMSDs of around 0.4 Hz. The slightly elevated RMSD in α-synuclein peptides using the nmr2 force field may be caused by the fragmentation of the protein in the simulations precluding the occurrence of long-range effects that may be present in the experiment. Some degrees of long-range interactions present in the full-length protein in the NMR measurement,[42-43] although such contributions to protein local backbone conformations are relatively small given the similar RMSDs for the dipeptides and for the IDP peptides. Improved performance by the new force field was also found for a smaller IDP amyloid β(1-40)[44] (Figure S2).

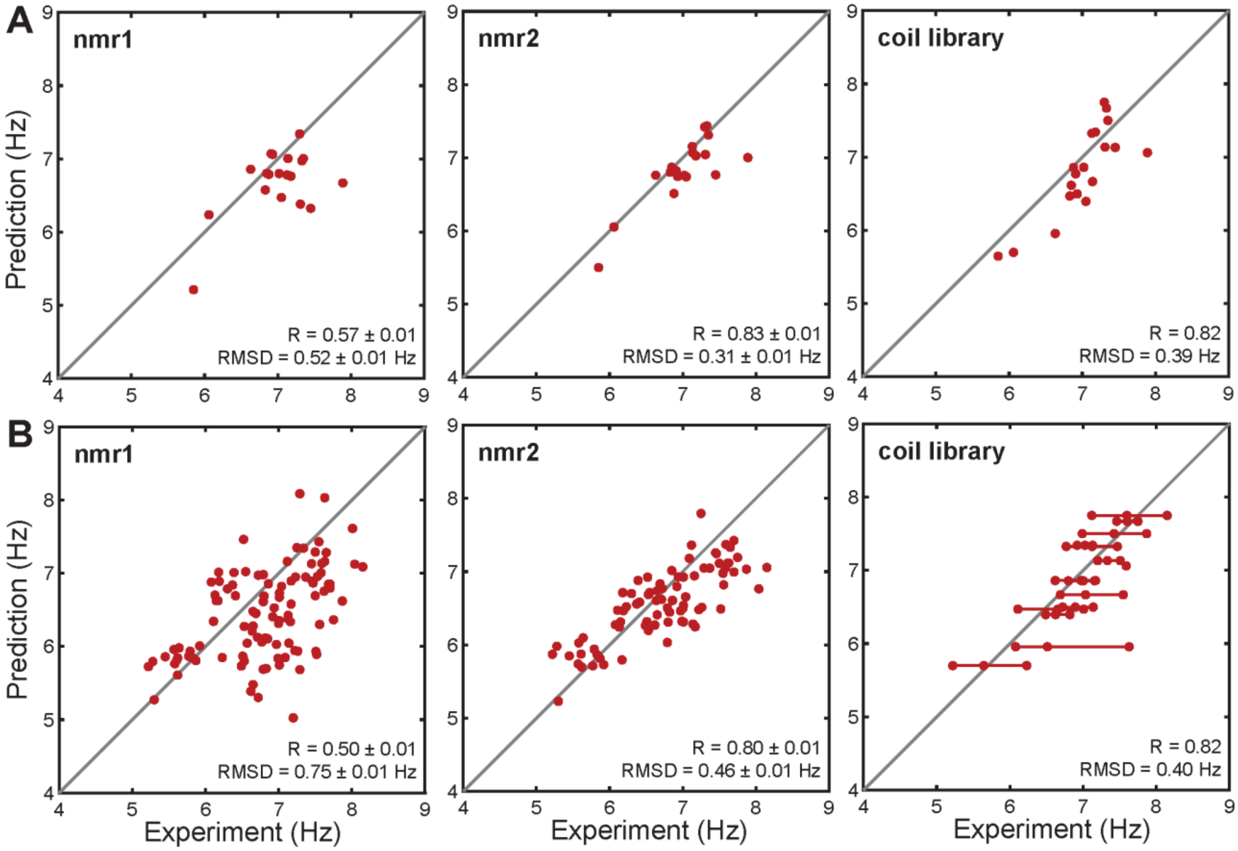


Figure 2. Predicted and experimental $^3J(H^N,H\alpha)$ coupling constants. ^3J-couplings in (A) 20 dipeptides and (B) 45 capped heptapeptides extracted from the sequence of α-synuclein were predicted from MD simulations using two different force fields, nmr1 and nmr2, and from the coil library using the average values for each residue type. These predicted coupling constants were then compared with experimental values. When multiple experimental ^3J-coupling constants are available for the same residue type in the coil library, they are represented by a line connected by the minimum, the average, and the maximum values.

NMR S^2 order parameters

Next, the performance of the new force field was benchmarked for globular proteins with flexible loops and tails and compared with its predecessor nmr1 and two other state-of-the-art protein force fields ff14SB and C36m. NMR N-H S^2 order parameters, which reflect internal protein backbone dynamics on the ps – ns timescale, were calculated from MD simulations using the iRED method[37] and compared with experimental values of the engrailed homeodomain (EnHD)[45], hen egg-white lysozyme (HEWL)[46], and human Interleukin-4[47]. While general agreements were reached between the experimental and predicted S^2 profiles for most force fields, clear discrepancies were observed in some places (Figure 3A–C). Noticeably, ff14SB over-stabilizes the N-terminus of EnHD, giving order parameters between 0.6 and 0.8, which substantially exceed the experimental values (0.2 – 0.4). Both ff14SB and C36m render the C-terminus of EnHD too rigid compared to the experiment. The same trend applies for ff14SB for the loop region comprising residues Ala35 – Thr39 in Interleukin-4. C36m accurately reproduces the S^2 values of Loop AB2 (residues 31 – 40) of Interleukin-4, but over-stabilizes loop residues Val102 – Asn105 of the same protein. In addition, C36m over-estimates the loop dynamics of HEWL at times. Both nmr1 and nmr2 predict the order parameters of EnHD with high accuracy. In Interleukin-4, nmr2 improves the agreement of S^2 values of Loop AB2 over nmr1, but overestimates the mobility of the 3_{10} helix near the N-terminus. Overall, the new force field is capable of probing loop dynamics remarkably accurately, and its performance is comparable to or better than other modern protein force fields.

Slow loop dynamics of Im7

For colicin E7 immunity protein (Im7), S^2 order parameters were calculated from the MD simulation with the nmr2 force field using 12 different iRED time-averaging windows τ_{iRED} ranging from 250 ps to 1 μs. These back-calculated order parameters S^2(MD) were then compared with order parameters extracted from the NMR spin relaxation experiment using the traditional model-free analysis S^2(MF) and those obtained from nanoparticle-assisted R_2 relaxation, denoted as $S^2(\Delta R_2)$, which probe internal dynamics at slower time scales[48] (Figure 3D). S^2(MD) with a 25-ns average time window, required for quantitative comparison with experimental S^2(MF),[49] agree well with S^2(MF) except for residues Val27, Thr30, and Arg61. With larger time-averaging windows of 100 ns and 250 ns, required for quantitative comparisons with experimental $S^2(\Delta R_2)$, S^2(MD) reproduce $S^2(\Delta R_2)$ values well. This demonstrates that protein backbone dynamics of Im7 on times scales from sub-nanoseconds to hundreds of nanoseconds are accurately represented by the nmr2 force field.

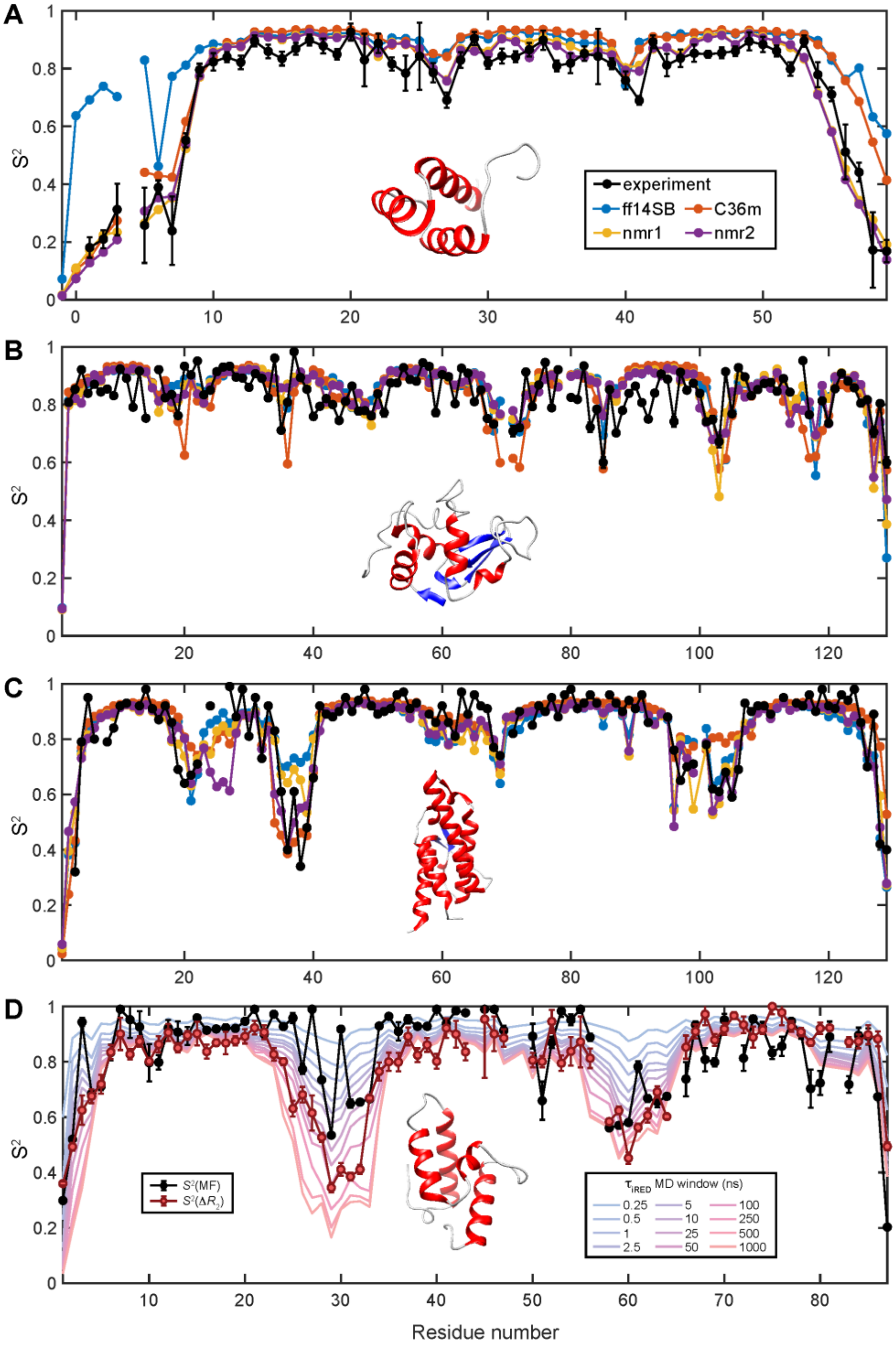


Figure 3. NMR backbone N-H S^2 order parameters of (A) engrailed homeodomain (EnHD), (B) hen egg white lysozyme (HEWL), (C) Interleukin-4, and (D) Im7. (A–C) Each force field is presented by a different color and the experimental data are in black. For each protein, S^2 order parameters were back-calculated from the simulation using the iRED method with a 25-ns time-averaging window and compared with experimental values obtained by model-free analysis. (D) For Im7, the length of the time-averaging window was varied from 25 ps to 1 μs, and the back-

calculated order parameters were compared with both experimental order parameters extracted from model-free analysis S^2(MF) and those obtained from nanoparticle-assisted R_2 relaxation $S^2(\Delta R_2)$, which probe internal motion on a much wider range of timescales from picoseconds to hundreds of nanoseconds. Figure S7 shows the same figure with error bars.

Structural accuracy of globular proteins

The new force field was further validated for globular proteins by comparing experimental and back-calculated RDC constants and chemical shifts (Table 1). RDC Q factors, which measure the agreement of RDCs between experiments and simulations, are similar across different force fields for ubiquitin[50] and B1 immunoglobulin-binding domain of streptococcal protein G (GB1)[51]. The nmr2 force field gives the second-best Q factor for ubiquitin and a slightly higher Q factor for GB1. Two more proteins, namely the N-terminal WW domain of Pin1 (Pin1 WW) and Trp-cage[52], were included in chemical shift back-calculations. Force field nmr2 gives the lowest RMSDs for all proteins except for GB1. For the reference, chemical shifts were also predicted for disordered peptides of α-synuclein and amyloid β with similar RMSDs from the nmr1 and nmr2 force fields. Additional validations of simulations of globular proteins in terms of the average Cα RMSDs relative to crystal structures can be found in Table S4. Together, these results show that the modified force field is able to provide proper stabilization of folded structures and at the same time generate realistic conformational thermal ensembles.

Table 1. Comparisons with experimental NMR parameters.

protein	ff14SB	C36m	nmr1	nmr2
RDC Q factor				
ubiquitin	0.27 ± 0.02	0.22 ± 0.01	0.25 ± 0.01	0.24 ± 0.03
GB1	0.21 ± 0.04	0.22 ± 0.01	0.21 ± 0.11	0.23 ± 0.05
chemical shift back-calculation error (ppm)				
ubiquitin	0.87 ± 0.03	0.79 ± 0.01	0.81 ± 0.02	0.79 ± 0.02
GB1	1.62 ± 0.01	1.70 ± 0.01	1.66 ± 0.01	1.66 ± 0.01
Pin1 WW	0.82 ± 0.01	0.85 ± 0.05	0.82 ± 0.01	0.81 ± 0.01
Trp-cage	1.57 ± 0.01	2.33 ± 0.16	2.10 ± 0.26	1.57 ± 0.01
α-synuclein	N/A	N/A	0.70	0.68
amyloid β	N/A	N/A	0.67	0.66

Reversible peptide/protein folding

The new force field was finally used to study the folding of a mutant of the ten-residue peptide chignolin, CLN025, and the subdomain of villin headpiece. Beginning from its crystal structure, chignolin undergoes reversible folding and unfolding more than three times at the experimental melting temperature (T_m) of 343 K[53] within the duration of the simulation of 3 μs (Figure 4A). The Cα RMSD between the most populated folded conformation in the simulation and the crystal structure is 0.94 Å. For the unfolded states in the simulation, the Cα RMSD gets as high as 9.01 Å, which even exceeds the RMSD between the crystal structure and a fully extended peptide conformation (8.34 Å). For better convergence, replica exchange MD simulations were performed to determine the folding curve of chignolin (Figure 4B). The folded population in the simulation, however, is significantly underestimated, which is a common problem of many modern protein force fields, including ff99SB-*disp*.[24] Similarly, the villin headpiece subdomain samples various intermediates and denatured states at T_m = 343 K[54] in the course of a 5-μs simulation (Figure 4C, D). The Cα RMSD between the most populated conformation in the simulation and the crystal structure (PDB code 1YRF) is 0.76 Å (excluding the N-terminal residue Met41 which is not present in the crystal structure). Such successful simulations of the protein folding/unfolding events demonstrate the potential of the balanced force field for elucidation of atomic-detail folding pathways in real-world applications.

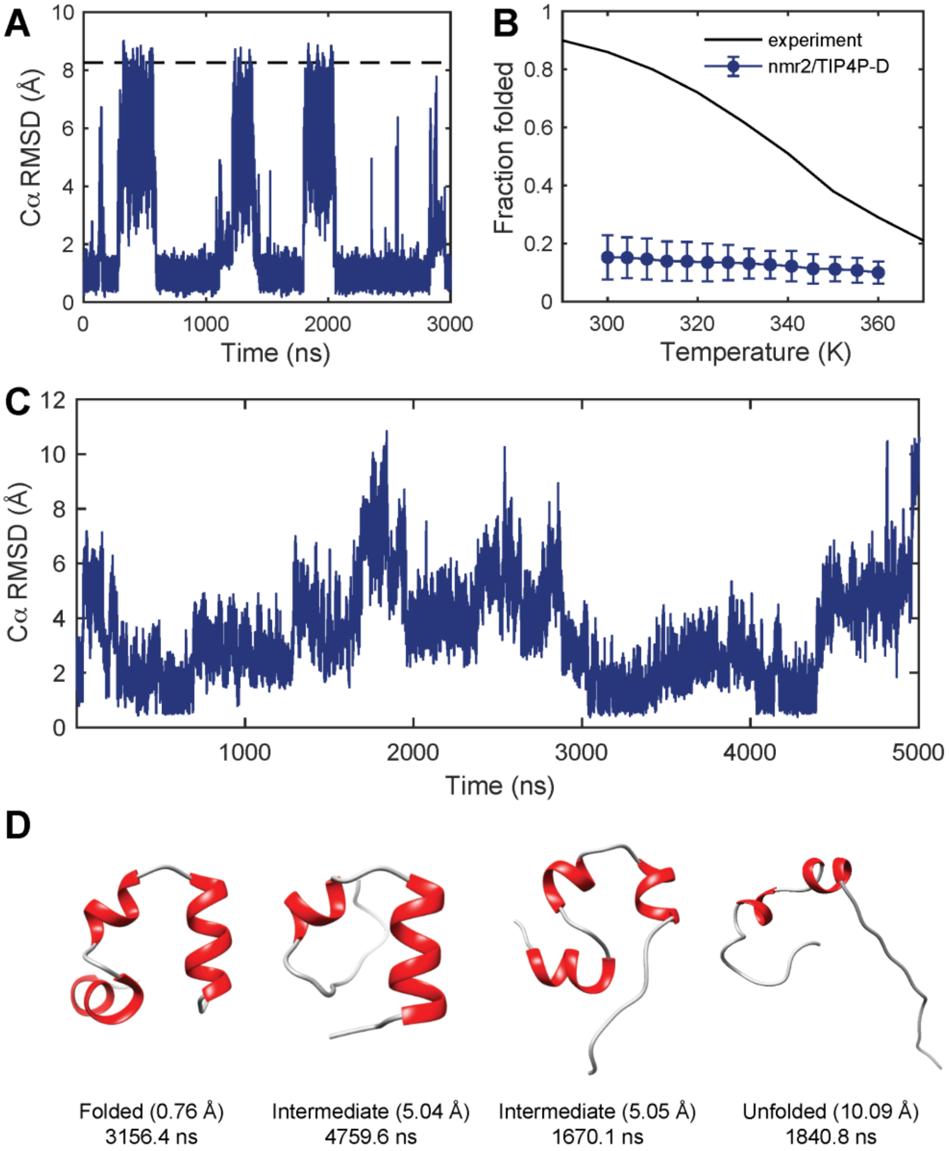


Figure 4. Reversible folding and unfolding of fast-folding peptides and proteins using the new nmr2 force field with TIP4P-D water. (A) Reversible real-time folding and unfolding of the chignolin decapeptide. (B) The folding curve of chignolin from experiment and replica exchange MD simulations. (C) Real-time folding and unfolding of the subdomain of villin headpiece. The dashed line in (A) corresponds to the RMSD between a fully extended conformation and the crystal structure. (D) Representative folded and unfolded structures of the subdomain of villin headpiece are shown with their RMSDs relative to the X-ray crystal structure.

DISCUSSION

The performance of the nmr2 force field can be quantitatively compared with other force fields based on various experimental data sets. The ff14SB force field tends to over-populate helical conformations (α_R region), which is evidenced by underestimated $^3J(H^N,H\alpha)$ coupling constants of dipeptides back-calculated from MD trajectories (Figure S3A). In addition, ff14SB tends to rigidify termini (Figure 3A) and underestimate the loop flexibilities (Figure 3C), which makes it less suitable for modeling the structure and dynamics of disordered proteins and highly flexible protein regions. C36m performs overall better than ff14SB for simulating flexible protein loops. However, the backbone dihedral angle distributions produced by C36m are not optimal for disordered peptides (Figure S3B). The new force field nmr2 is more balanced in this regard. It yields the best ^3J-coupling predictions for dipeptides and significantly improves the backbone ensemble description of disordered peptides (Figure 2). The predicted protein dynamics S^2 profiles produced by the nmr2 force field are among the best when compared with experimental data. When changing the TIP4P-D water model to TIP3P, the flexibility of protein loops and termini were underestimated (Figure S6). The performance of nmr2 for globular protein simulations is also comparable to other protein force fields evaluated here (Table 1). Moreover, nmr2 is able to fold small peptides such as chignolin reversibly with very high, atomistic accuracy of the folded state. Its inability to provide adequate stabilization for the folded structure, however, is likely due to the TIP4P-D water model. When replacing TIP4P-D with TIP3P, chignolin is able to maintain a high population of folded structures determined from replica exchange simulations at elevated temperatures (Figure S5). Therefore, the new protein force field nmr2 performs well for both folded and disordered proteins, thereby fulfilling the premise of this project.

The high accuracy of the new force field is directly related to the way coil library information was used for the modification of nmr1. First, we modified the dihedral angle potentials of nmr1 in a residue-specific manner. Different dihedral potential corrections were used for different types of amino acid residues, which proved useful in fine tuning the conformational ensembles of disordered peptides. Second, when optimizing the dihedral angle potentials, we only used coil library information as the reference for the relative populations of *four* major Ramachandran regions rather than to fit the *entire* Ramachandran space. In this way, we were able to retain the favorable local features of the parent protein force field nmr1, as was demonstrated previously for folded proteins.[6] Using the hybrid dihedral potentials, we are able to ameliorate the destabilizing effect of the TIP4P-D water model on folded protein structures to some extent[18, 20] and to use a single, unified protein force field to model folded and disordered states at the same time.

Comparisons of the radii of gyration of α-synuclein peptides simulated with different combinations of protein force fields and water models permit a better separation of the effects of water models vs. protein force fields. Using the same protein force field nmr1, the disordered peptides became much more extended when the TIP3P water model was replaced by TIP4P-D (Figure 5A). Conversely, the modification of the protein force field had little effect on the chain dimensions when the same water model TIP4P-D was used (Figure 5B). The results show that the new nmr2 force field benefits from the TIP4P-D water model to reproduce the extended structures of disordered peptides and suggest that water models in general have an important role in simulating aqueous protein structures. The effect of the protein force field modification was further analyzed by comparing conformational entropies in the dihedral angle space calculated with up to second-order correlations.[55] Changing water models on average have negligible effects on conformational entropy, whereas the change of the nmr1 protein force field to nmr2 on average increases the entropy by 0.04 k_B per dihedral angle (Figure 5C). Backbone and side-chain contributions to conformational entropies can be found in Figure S4. For these α-synuclein heptapeptides, the increase of conformational entropy from the modification results in an average decrease of 2.64 kJ·mol^{-1} in Gibbs free energy (at 300 K), thereby stabilizing the unfolded states.

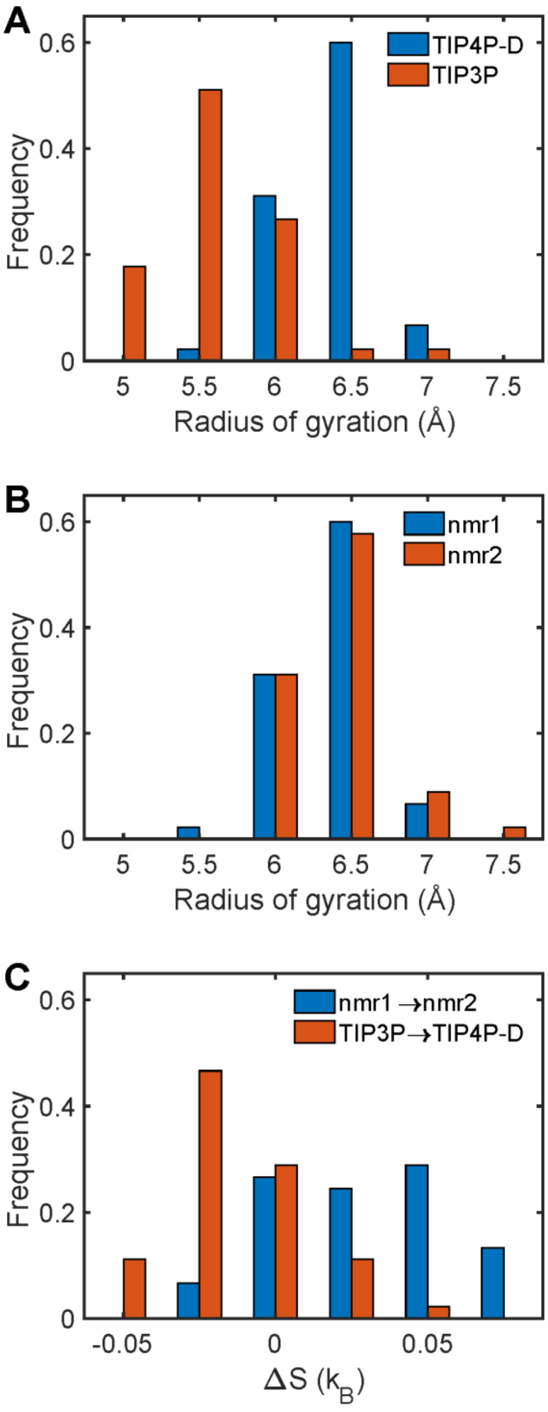


Figure 5. Effects of protein force fields and water models on disordered peptides. (A, B) Relative frequency distributions of the radii of gyration of 45 α-synuclein heptapeptides simulated using (A) two different water models and (B) two different protein force fields. (C) Relative frequency distributions of the change in conformational entropy per dihedral ΔS due to the change of the protein force field and the change of water models. A total of 1142 dihedrals were analyzed.

CONCLUSION

In this work, the novel protein force field ff99SBnmr2 was built by incorporating backbone dihedral angle distributions from coil libraries into the existing ff99SBnmr1 force field. By introducing residue-specific φ,ψ propensities and retaining local features of the prior force field, this new coil library-modified force field ff99SBnmr2 significantly improves the backbone conformational ensembles of proteins while keeping folded protein parts stable. In addition, ff99SBnmr2 is capable of probing loop dynamics accurately with a performance that is comparable to or better than other modern protein force fields. We show that the new force field succeeds in modeling IDPs by providing stabilization of disordered states. As a result, the well-balanced nature of the novel force field is expected to prove useful for the quantitative description of biomolecular processes, including protein-protein and protein-ligand recognition processes involving both rigid and dynamic parts and folding/unfolding equilibria.

ASSOCIATED CONTENT

The Supporting Information is available free of charge via the Internet at http://pubs.acs.org

The data supporting the findings of this study are available within the article and its Supplementary Information files. The new force field AMBER ff99SBnmr2 is downloadable from our website https://research.cbc.osu.edu/bruschweiler.1/protein-force-field/ .

AUTHOR INFORMATION

Corresponding Author

*E-mail: bruschweiler.1@osu.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

This work was supported by the U.S. National Science Foundation (grant MCB-1715505 to R.B.). MD simulations were performed at the Ohio Supercomputer Center.

REFERENCES

1. Karplus, M.; McCammon, J. A., Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002,** *9* (9), 646-652.

2. Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Piana, S.; Shan, Y. B.; Towles, B., Millisecond-Scale Molecular Dynamics Simulations on Anton. *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* **2009**.

3. Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E., How Fast-Folding Proteins Fold. *Science* **2011,** *334* (6055), 517-520.

4. 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* (2), 647-655.

5. 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* (2), e32131.

6. 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* (4), 1409-1414.

7. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr., A. D., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017,** *14* (1), 71-73.

8. 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., Bioinf.* **2011,** *79* (4), 1318-1328.

9. Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T., Structure and Dynamics of the $A\beta_{21-30}$ Peptide from the Interplay of NMR Experiments and Molecular Simulations. *J. Am. Chem. Soc.* **2011,** *133* (30), 11816-11816.

10. 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. USA* **2014,** *111* (45), 15975-15980.

11. 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.

12. Uversky, V. N.; Oldfield, C. J.; Dunker, A. K., Intrinsically Disordered Proteins in Human Diseases: Introducing the D^2 Concept. *Annu. Rev. Biophys.* **2008,** *37*, 215-246.

13. 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.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M., Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014,** *114* (13), 6589-6631.

14. Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H., Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015,** *11* (11), 5513-5524.

15. 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* (11), 5113-5124.

16. Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T., Optimizing Solute–Water van der Waals Interactions To Reproduce Solvation Free Energies. *J. Phys. Chem. B* **2012,** *116* (15), 4524-4534.

17. 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* (7), 3420-3431.

18. Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E., Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015,** *119* (16), 5113-5123.

19. Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T., Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016,** *12* (8), 3926-3947.

20. Wu, H.-N.; Jiang, F.; Wu, Y.-D., Significantly Improved Protein Folding Thermodynamics Using a Dispersion-Corrected Water Model and a New Residue-Specific Force Field. *J. Phys. Chem. Lett.* **2017,** *8* (14), 3199-3205.

21. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-3713.

22. Zhou, C.-Y.; Jiang, F.; Wu, Y.-D., Residue-Specific Force Field Based on Protein Coil Library. RSFF2: Modification of AMBER ff99SB. *J. Phys. Chem. B* **2015,** *119* (3), 1035-1047.

23. Song, D.; Luo, R.; Chen, H.-F., The IDP-Specific Force Field *ff14IDPSFF* Improves the Conformer Sampling of Intrinsically Disordered Proteins. *J. Chem. Inf. Model.* **2017,** *57* (5), 1166-1178.

24. Robustelli, P.; Piana, S.; Shaw, D. E., Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. USA* **2018,** *115* (21), E4758-E4766.

25. Kuzmanic, A.; Pritchard, R. B.; Hansen, D. F.; Gervasio, F. L., Importance of the Force Field Choice in Capturing Functionally Relevant Dynamics in the von Willebrand Factor. *J. Phys. Chem. Lett.* **2019,** *10* (8), 1928-1934.

26. Liu, X.; Chen, J., Residual Structures and Transient Long-Range Interactions of p53 Transactivation Domain: Assessment of Explicit Solvent Protein Force Fields. *J. Chem. Theory Comput.* **2019,** *15* (8), 4708-4720.

27. Li, D.-W.; Brüschweiler, R., NMR-based protein potentials. *Angew. Chem., Int. Ed.* **2010,** *49* (38), 6778-6780.

28. Huang, J.; Lopes, P. E.; Roux, B.; MacKerell, J., A. D., Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model. *J. Phys. Chem. Lett.* **2014,** *5* (18), 3144-3150.

29. Fitzkee, N. C.; Fleming, P. J.; Rose, G. D., The Protein Coil Library: A Structural Database of Nonhelix, Nonstrand Fragments Derived from the PDB. *Proteins: Struct., Funct., Bioinf.* **2005,** *58* (4), 852-854.

30. Jiang, F.; Han, W.; Wu, Y.-D., The intrinsic conformational features of amino acids from a protein coil library and their applications in force field development. *Phys. Chem. Chem. Phys.* **2013,** *15* (10), 3413-3428.

31. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.

32. Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham III, T. E.; T.A. Darden; R.E. Duke; T.J. Giese; H. Gohlke; A.W. Goetz; N. Homeyer; S. Izadi; P. Janowski; J. Kaus; A. Kovalenko; T.S. Lee; S. LeGrand; P. Li; C. Lin; T. Luchko; R. Luo; B. Madej; D. Mermelstein; K.M. Merz; G. Monard; H. Nguyen; H.T. Nguyen; I. Omelyan; A. Onufriev; D.R. Roe; A. Roitberg; C. Sagui; C.L. Simmerling; W.M. Botello-Smith; J. Swails; R.C. Walker; J. Wang; R.M. Wolf; X. Wu; Xiao, L.; Kollman, P. A., AMBER 2016. *University of California, San Francisco* **2016**.

33. Vögeli, B.; Ying, J. F.; Grishaev, A.; Bax, A., Limits on Variations in Protein Backbone Dynamics from Precise Measurements of Scalar Couplings. *J. Am. Chem. Soc.* **2007,** *129* (30), 9377-9385.

34. Li, D.; Brüschweiler, R., PPM_One: a static protein structure based chemical shift predictor. *J. Biomol. NMR* **2015,** *62* (3), 403-409.

35. Showalter, S. A.; Brüschweiler, R., Quantitative Molecular Ensemble Interpretation of NMR Dipolar Couplings without Restraints. *J. Am. Chem. Soc.* **2007,** *129* (14), 4158-4159.

36. Ottiger, M.; Bax, A., Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values. *J. Biomol. NMR* **1999,** *13* (2), 187-191.

37. Prompers, J. J.; Brüschweiler, R., General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation. *J. Am. Chem. Soc.* **2002,** *124* (16), 4522-4534.

38. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z., Intrinsically disordered protein. *J. Mol. Graph. Model.* **2001,** *19* (1), 26-59.

39. Maltsev, A. S.; Ying, J.; Bax, A., Impact of N-terminal Acetylation of α-Synuclein on Its Random Coil and Lipid Binding Properties. *Biochemistry* **2012,** *51* (25), 5004-5013.

40. Avbelj, F.; Grdadolnik, S. G.; Grdadolnik, J.; Baldwin, R. L., Intrinsic backbone preferences are fully present in blocked amino acids. *Proc. Natl. Acad. Sci. USA* **2006,** *103* (5), 1272-1277.

41. Mantsyzov, A. B.; Maltsev, A. S.; Ying, J.; Shen, Y.; Hummer, G.; Bax, A., A maximum entropy approach to the study of residue-specific backbone angle distributions in α-synuclein, an intrinsically disordered protein. *Protein Sci.* **2014,** *23* (9), 1275-1290.

42. Bertoncini, C. W.; Jung, Y.-S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M., Release of long-range tertiary interactions potentiates aggregation of natively unstructured α-synuclein. *Proc. Natl. Acad. Sci. USA* **2005,** *102* (5), 1430-1435.

43. Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M., Mapping Long-Range Interactions in α-Synuclein using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005,** *127* (2), 476-477.

44. Roche, J.; Shen, Y.; Lee, J. H.; Ying, J. F.; Bax, A., Monomeric $A\beta^{1\text{-}40}$ and $A\beta^{1\text{-}42}$ Peptides in Solution Adopt Very Similar Ramachandran Map Distributions That Closely Resemble Random Coil. *Biochemistry* **2016,** *55* (5), 762-775.

45. Religa, T. L., Comparison of multiple crystal structures with NMR data for engrailed homeodomain. *J. Biomol. NMR* **2008,** *40* (3), 189-202.

46. Moorman, V. R.; Valentine, K. G.; Wand, A. J., The dynamical response of hen egg white lysozyme to the binding of a carbohydrate ligand. *Protein Sci.* **2012,** *21* (7), 1066-1073.

47. Redfield, C.; Boyd, J.; Smith, L. J.; Smith, R. A.; Dobson, C. M., Loop Mobility in a Four-Helix-Bundle Protein: ^{15}N NMR Relaxation Measurements on Human Interleukin-4. *Biochemistry* **1992,** *31* (43), 10431-10437.

48. Xie, M.; Yu, L.; Bruschweiler-Li, L.; Xiang, X.; Hansen, A. L.; Brüschweiler, R., Functional protein dynamics on uncharted time scales detected by nanoparticle-assisted NMR spin relaxation. *Sci. Adv.* **2019,** *5* (8), eaax5560.

49. Gu, Y.; Li, D.-W.; Brüschweiler, R., NMR order parameter determination from long molecular dynamics trajectories for objective comparison with experiment. *J. Chem. Theory Comput.* **2014,** *10* (6), 2599-2607.

50. Lakomek, N.-A.; Walter, K. F.; Farès, C.; Lange, O. F.; de Groot, B. L.; Grubmüller, H.; Brüschweiler, R.; Munk, A.; Becker, S.; Meiler, J.; Griesinger, C., Self-consistent residual dipolar coupling based model-free analysis for the robust determination of nanosecond to microsecond protein dynamics. *J. Biomol. NMR* **2008,** *41* (3), 139-155.

51. Bouvignies, G.; Meier, S.; Grzesiek, S.; Blackledge, M., Ultrahigh-Resolution Backbone Structure of Perdeuterated Protein GB1 Using Residual Dipolar Couplings from Two Alignment Media. *Angew. Chem., Int. Ed.* **2006,** *45* (48), 8166-8169.

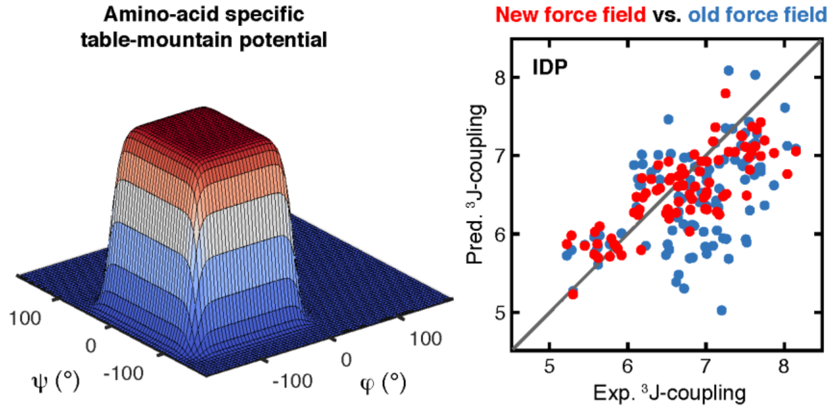
52. Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H., Designing a 20-residue protein. *Nat. Struct. Biol.* **2002,** *9* (6), 425-430.

53. Honda, S.; Akiba, T.; Kato, Y. S.; Sawada, Y.; Sekijima, M.; Ishimura, M.; Ooishi, A.; Watanabe, H.; Odahara, T.; Harata, K., Crystal Structure of a Ten-Amino Acid Protein. *J. Am. Chem. Soc.* **2008,** *130* (46), 15327-15331.

54. McKnight, C. J.; Matsudaira, P. T.; Kim, P. S., NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Biol.* **1997,** *4* (3), 180-184.

55. Wang, J.; Brüschweiler, R., 2D Entropy of Discrete Molecular Ensembles. *J. Chem. Theory Comput.* **2006,** *2* (1), 18-24.

TOC Figure


Amino-acid specific
table-mountain potential
100
0
-100
ψ (°)
-100
0
100
φ (°)
New force field vs. old force field
IDP
8
7
6
5
Pred. ³J-coupling
5
6
7
8
Exp. ³J-coupling