THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Simple and Effective Conformational Sampling Strategy for Intrinsically Disordered Proteins Using the UNRES Web Server

Published as part of The Journal of Physical Chemistry virtual special issue "Protein Folding and Dynamics—An Overview on the Occasion of Harold Scheraga's 100th Birthday".

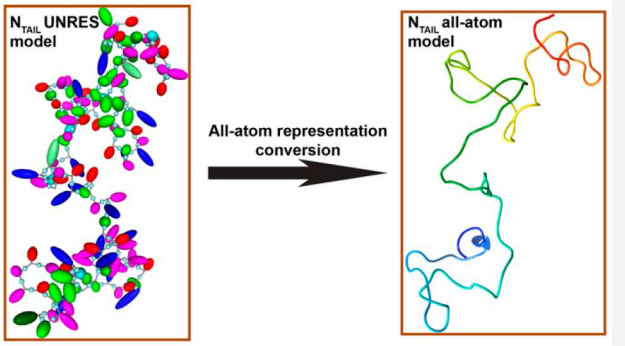
Tongtong Li, Emily Hendrix, and Yi He*

Cite This: *J. Phys. Chem. B* 2023, 127, 2177−2186

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Intrinsically disordered proteins (IDPs) contain more charged amino acids than folded proteins, resulting in a lack of hydrophobic core(s) and a tendency to adopt rapidly interconverting structures rather than well-defined structures. The structural heterogeneity of IDPs, encoded by the amino acid sequence, is closely related to their unique roles in biological pathways, which require them to interact with different binding partners. Recently, Robustelli and co-workers have demonstrated that a balanced all-atom force field can be used to sample heterogeneous structures of disordered proteins (*Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E4758−E4766). However, such a solution requires extensive computational resources, such as Anton supercomputers. Here, we propose a simple and effective solution to sample the conformational space of IDPs using a publicly available web server, namely, the UNited-RESidue (UNRES) web server. Our proposed solution requires no investment in computational resources and no prior knowledge of UNRES. UNRES Replica Exchange Molecular Dynamics (REMD) simulations were carried out on a set of eight disordered proteins at temperatures spanning from 270 to 430 K. Utilizing the latest UNRES force field designed for structured proteins, with proper selections of temperatures, we were able to produce comparable results to all-atom force fields as reported in work done by Robustelli and co-workers. In addition, NMR observables and the radius of gyration calculated from UNRES ensembles were directly compared with the experimental data to further evaluate the accuracy of the UNRES model at all temperatures. Our results suggest that carrying out the UNRES simulations at optimal temperatures using the UNRES web server can be a good alternative to sample heterogeneous structures of IDPs.



INTRODUCTION

Unlike folded proteins, intrinsically disordered proteins (IDPs) exhibit many unique behaviors because of their biophysical properties encoded in their amino acid sequences. IDPs are typically composed of more polar and charged amino acids[1,2] but fewer hydrophobic residues. IDPs prefer to frequently switch among conformers rather than folding into well-defined structures. A previous study proposed the fly cast mechanism,[3] noting that the flexibility of disordered proteins endows them with greater capture radii, which speeds up their recognition of specific binding sites. However, this advantage in structural recognition, due to the structural flexibility and heterogeneity of IDPs, poses a significant challenge in characterizing their structures and exploring their biological functions. The experimental techniques, such as X-ray Diffraction (XRD),[4] cryo-electron microscopy (cryo-EM),[5−7] and integrative methods,[8,9] are designed to determine the stable folded structures of proteins. Nuclear Magnetic Resonance (NMR), small-angle X-ray scattering (SAXS), and fluorescence resonance energy transfer (FRET) are practical tools for revealing structural information on some IDP systems,[10−12] but the structural information is often not sufficient to reconstruct the corresponding structural ensembles of IDPs. Moreover, because of the significant fluctuations/variations in structures of IDPs, it is difficult to generate one or a few structures to represent the whole structural ensembles for IDPs. Besides experimental methods, computational methods can be a good alternative or a complementary tool for exploring the structural ensembles of IDPs.[13−16] Most of the force fields designed for simulating proteins[17−19] are for folded proteins and tend to yield structures deviating from actual

Received: December 22, 2022
Revised: February 6, 2023
Published: February 24, 2023

states when used to simulate disordered proteins.[20] Recently, Robustelli et al.[21] have developed an all-atom force field, the a99SB-*disp*, for both folded and disordered proteins. The new force field was validated with extensive experimental data, including the NMR data [chemical shifts (CS), residual dipolar couplings (RDCs), *J* scalar coupling constants, paramagnetic relaxation enhancements (PREs)] and radius of gyrations (Rg). Compared with other popular force fields, the force field a99SB-*disp* predicted structural properties reached a better agreement with the experimental data for both folded and disordered proteins. Despite all the progress made in building all-atom force fields for IDPs,[22] all-atom simulations still require considerable computational resources to effectively sample the conformational space of IDPs. Therefore, physics-based coarse-grained force fields, such as the UNited RESidue (UNRES) force field,[23−27] are useful tools to simulate IDPs. In the UNRES force field, the geometry of peptide chains is represented by two groups, the simplified side chain and the peptide group. Besides the simplification of the representation of proteins, the protein surrounding environment and all other degrees of freedom have been integrated out and embedded implicitly in UNRES energy equations when deriving the potential of mean force corresponding to the UNRES representation.[28,29] With all these simplifications, UNRES can be used to sample conformational spaces for proteins. The UNRES force field has been extensively used in protein folding simulations[30−36] and protein structural predictions in the Critical Assessment of Methods of Protein Structure Prediction (CASP) experiments.[23,37−41] The release of the UNRES web server[42] has provided an easy access option for the general public to use UNRES without prior knowledge and with no commitment to computational resources.

In this work, we will benchmark the performance of the UNRES force field for the simulation of IDPs using the UNRES web server. Eight disordered proteins from a subset of proteins used in the study of Robustelli et al.,[21] including the ParE2-associated antitoxin PaaA2 (hereinafter called PaaA2),[43] the N_{TAIL} domain of the measles virus nucleoprotein (hereinafter called N_{TAIL}),[44] an intrinsically disordered region from the *Saccharomyces cerevisiae* transcription factor Ash1 (hereinafter called Ash1),[45] drkN SH3,[46] Aβ40,[47] ACTR,[48] the cyclin-dependent kinase inhibitor Sic1 (hereinafter called Sic1),[49] and α-synuclein,[50] were simulated. Replica Exchange Molecular Dynamics (REMD)[51,52] simulations were used to enhance the sampling of the conformational space of these IDPs. UNRES results were then converted to all-atom representations to make a direct comparison to both trajectories generated by force field a99SB-*disp* (provided by the D. E. Shaw group) and experimental data (provided by Dr. Robustelli). The force field score function used by Robustelli et al.[21] was used here to calculate the score of the UNRES force field. Compared with the a99SB-*disp* ensembles, the UNRES trajectories reconstructed based on simulation temperatures exhibited different structural similarities at various simulation temperatures. The same situation happens when comparing UNRES predicted experimental observables with real measurements, yielding variable force field scores, which suggests UNRES trajectories at different temperatures display different degrees of similarity with experimentally detected structures. This indicates that the choice of the simulation temperature provides a simple solution to tune the structural property and the predicted structural ensemble of IDPs. In addition, our results indicate that the force field performance improves as the temperature increases until the unique optimal temperature for the benchmark has been met, which is partially related to the primary sequences and compositions of proteins. These results suggest that the UNRES force field can sample structures with different compactness and secondary structure contents if a proper temperature is used. In general, a high temperature, like 400 K, is more favorable for the IDP simulation with the UNRES model.

METHODS

UNRES Force Field. The UNited-RESidue (UNRES) force field[23,25−27,53] is based on the potential of mean force (PMF) of protein chains including the surrounding solvents. UNRES uses a very simple description to represent the geometry of protein chains to reduce the computational cost but still maintain a reasonable accuracy. Each residue in a protein is represented by two interaction sites, the side chain site and the peptide group.[24] All Cα atoms serve only as geometric points to define the geometry of the protein chain. The contributions of the surrounding solvent environment are embedded into the UNRES energy equations during the process of deriving the PMF of the protein system. With such implicit consideration of the solvent environment, it increases the simulation speed significantly. With the recently developed Maximum Likelihood optimization strategy,[54,55] UNRES can predict the structures of proteins solely based on the physics-based energy function. It should be noted that some energy terms are temperature dependent.[56] Temperature is one of the key factors that can affect the energy predicted from conformers and subsequently the sampled structural distribution. UNRES also provides options to start simulations from either PDB structures or a randomly generated structure based on an amino acid sequence, which is essential for the simulation of disordered proteins lacking structures deposited in the protein database bank.

Apart from using UNRES to speed up simulations, enhanced sampling methods, like Multiplexed Replica Exchange Molecular Dynamics (MREMD)[51,52] simulations can effectively sample large conformational space, which are essential for IDP simulations. MREMD simulations,[57] carried out at a range of temperatures, enable the structural exchange between parallel trajectories at different or same temperatures with a particular probability. This is extremely important in the sampling of the IDP structural ensembles by promoting the escape of structures that are trapped in local minima. MREMD simulations can cover a larger conformational space compared to running regular Molecular Dynamics (MD) simulations separately at multiple temperatures.[52] In our work, the latest version of the UNRES force field, the NEWCT-9P force field,[58] was used to sample the conformational space of eight IDPs, PaaA2, N_{TAIL}, Ash1, drkN SH3, Aβ40, ACTR, Sic1, and α-synuclein. While MREMD simulations have been implemented in UNRES, the simulations performed in this work utilized one replica per temperature, thus appropriately classified as REMD simulations, at 14 temperatures (270, 280, 290, 300, 310, 320, 330, 340, 355, 370, 385, 400, 415, and 430 K) using the publicly available UNRES server[42] (http://unres-server.chem.ug.edu.pl). Pulchra[59] was used to convert UNRES trajectories to all-atom trajectories. Hydrogen atoms were subsequently added to the converted trajectories using Rosetta scoring.[60] Random structure generated by UNRES was used as the starting structure for each protein, yielding a total

https://doi.org/10.1021/acs.jpcb.2c08945
J. Phys. Chem. B 2023, 127, 2177−2186

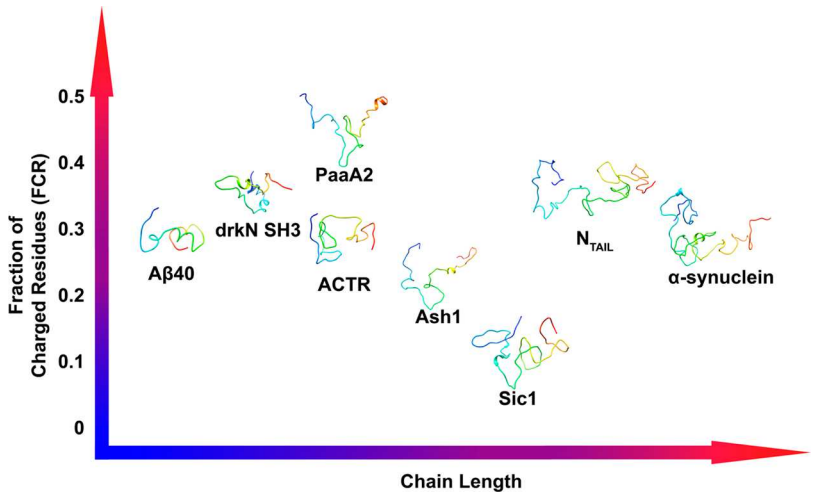


Figure 1. A benchmark set of eight disordered proteins (PaaA2, N_{TAIL}, Ash1, drkN SH3, Aβ40, ACTR, Sic1, and α-synuclein) was illustrated with different chain lengths and fractions of charged residues. Proteins are colored in the rainbow scheme (the N-terminus in blue and the C-terminus in red).

of 16000 snapshots with structures recorded every 10000 steps. And all structures in the UNRES ensembles were used for analysis.

Assessment of Structural Ensembles of UNRES Simulation at Different Temperatures. The a99SB-*disp*,[21] an all-atom force field developed by Robustelli et al., has been validated with extensive NMR and Rg data, which shows good agreement with experiments for disordered proteins and simultaneously maintains accuracy for folded proteins. Therefore, the structural ensembles sampled by the a99SB-*disp* force field at 300 K can serve as the reference to assess the performance of the UNRES/REMD simulations at various temperatures. All structures in the all-atom ensembles generated using the a99SB-*disp* force field at 300 K were used for structural comparison. Structural properties, like root-mean-square-fluctuation (RMSF), helical propensity, protein contacts, and relative energy surfaces, demonstrate the structural similarity between a99SB-*disp* ensembles and UNRES ensembles, especially those UNRES ensembles at high temperatures. In the contact map analysis, contacts were assumed to be present based on the pairwise distances between any heavy atoms of two residues (i, i + 5 with a cutoff of 5.5 Å). The probability of contact formation during the trajectories was used to make the contact map.

To further evaluate the performance of the UNRES model at different temperatures in-depth, we also compared experimental observables predicted from UNRES ensembles with the previously reported experimental measurements. With the reported a99SB-*disp* RMSDs against experimental observables,[21] we have made a direct comparison of the performance of the UNRES force field to the a99SB-disp force field. Specifically, four parameters, NMR chemical shifts (CS), residual dipolar couplings (RDCs), backbone $^3J_{HNH\alpha}$ scalar coupling constants, and radius of gyration (Rg), are used to assess the force field quantitatively. Chemical shifts are calculated with Sparta+.[61] RDCs are calculated with PALES[62,63] using a local alignment window of 15 residues.[64] Backbone $^3J_{HNH\alpha}$ scalar coupling constants predicted based on the Karplus equation[65,66] are calculated with the MDTraj software package.[67] The mass-weighted Rg on all atoms are calculated with the MDTraj software package.[67] The following score equations[21] are used to normalize the UNRES scores at various temperatures with the consideration of the reported RMSDs of the a99SB-*disp* force field.[21]

$$CS_{Score} = \frac{1}{N}\sum_{i=1}^{N}\frac{CS_{rmsd}}{CS_{Norm}}$$

where CS_{rmsd} is the root-mean-square deviation (rmsd) of UNRES/a99SB-*disp* predicted secondary chemical shifts against the experimentally measured chemical shifts with the subtraction of the random coil chemical shifts,[68,69] and CS_{Norm} is the minimum CS_{rmsd} among UNRES simulations ranging from 270 to 430 K and the a99SB-*disp* simulation, corresponding to the most similar conformational space in aspect of secondary chemical shifts. The normalized secondary chemical shifts rmsd of atoms Cα, Hα, C′, N, H_N, and Cβ are averaged to obtain the total CS_{Score}.

$$NMR_{Score} = \frac{1}{2}(RDC_{Score} + J_{Score})$$

where NMR_{Score} is the function of normalized rmsd of RDC and the backbone $^3J_{HNH\alpha}$ scalar coupling constants. Here, the scaled RDC Q-factor[70] is compared between the UNRES/a99SB-*disp* ensemble and experimental values instead of RDC rmsd. The normalized way for the scaled RDC Q-factors and the backbone $^3J_{HNH\alpha}$ scalar coupling constants is the same as secondary chemical shifts. Additionally, when the averaged Rg of the UNRES/a99SB-*disp* ensemble is within the range of experimental error, $Rg_{Penalty}$ is equal to 0. Otherwise, $Rg_{Penalty}$ will be calculated with the following equation.

$$Rg_{Penalty} = \frac{|Rg_{exp} - Rg_{Sim}| - Rg_{Exp\ error}}{Rg_{exp}}$$

$$FF_{Score} = \frac{CS_{Score} + NMR_{Score}}{2} + Rg_{Penalty}$$

The force field score is defined with the rmsd of NMR parameters combined with Rg penalty. The CS_{rmsd}, RDC Q-factor, and Rg for the a99SB-*disp* force field were obtained from the previous work.[21] With the scores at different temperatures, the optimal simulation temperature for each IDP in the benchmark set can be eventually determined.

https://doi.org/10.1021/acs.jpcb.2c08945
J. Phys. Chem. B 2023, 127, 2177−2186

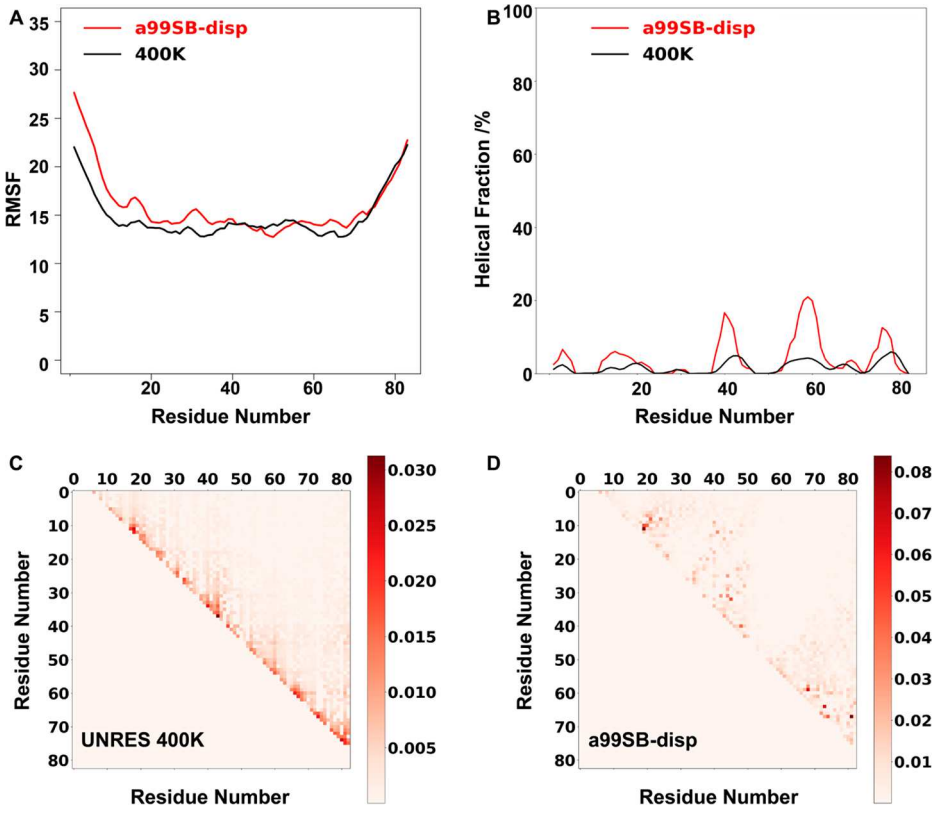


Figure 2. Ash1 structural similarity between UNRES and a99SB-*disp* ensembles. (A) RMSF and (B) helical fraction of Ash1 UNRES ensemble at 400 K (black lines) and the a99SB-*disp* ensembles (red lines). Contact maps of Ash1 are obtained from the UNRES ensemble at 400 K (C) and the a99SB-*disp* ensemble (D). Color bars in panels C and D indicate the frequency to form contacts between two residues in the corresponding trajectories, with darker red indicating a higher tendency of contact formation.

RESULTS

Eight intrinsically disordered proteins (IDPs) with various chain lengths and disorder levels, including PaaA2, N_{TAIL}, Ash1, drkN SH3, Aβ40, ACTR, Sic1, and α-synuclein, as shown in Figure 1, were widely used as benchmarks for force field validation and parametrization.[21,71−73] Here, the UNited-RESidue (UNRES) force field was evaluated using the above-mentioned benchmark set by comparing the NMR parameters, such as chemical shifts (CS), residual dipolar couplings (RDCs), and backbone $^{3}J_{HNH\alpha}$ scalar coupling constants, together with the radius of gyration (Rg) distributions, predicted from UNRES to the corresponding experimental data. The NMR parameters are reported to be highly sensitive to proteins structures. For example, the chemical shifts are widely used to characterize local structures of proteins and can predict the secondary structures, the torsional angles, and the protein flexibility.[74] The backbone $^{3}J_{HNH\alpha}$ scalar coupling constants are related to the geometry of peptide chains. Based on the empirical knowledge on J scalar coupling constants, the Karplus equation can estimate J scalar coupling constants from structures.[65] RDCs indicate orientations of internuclear vector relative to the magnetic field[75,76] and thereby can provide information for both the long-range and short-range contacts. Previous work also proposed that RDCs observed in disordered proteins are related to the local secondary structures.[77,78] Therefore, it is feasible to use the NMR parameters as the criteria to validate the structural properties of the benchmark set. Besides the NMR parameters mentioned above, the radius of gyration (Rg), showing the overall compactness of proteins, is also used to assess the structural similarity.

Evaluation of the Structural Similarity of UNRES Ensembles against a99SB-*disp* Ensembles. Structural analyses, such as RMSF, helical fraction, contact map, and relative energy surface analysis, were conducted to evaluate the structural similarity between UNRES and a99SB-*disp* ensembles. First of all, RMSF was performed to evaluate the flexibility of the disordered proteins. As shown in Figure 2A and Figure S1, we take the result of the Ash1 ensemble from the UNRES simulation at 400 K as an example. Fluctuations with a similar magnitude can be achieved in the RMSF of disordered proteins by adjusting the simulation temperature. We further analyzed the secondary structures of the benchmark set. When the simulation temperature is adjusted, the secondary structural content changes accordingly. As shown in Figure 2B, Ash1 exhibits a similar helical probability between UNRES simulations at 400 K and a99SB-*disp* simulation. In Figure S2C, when the temperature is below 340 K and keeps dropping, the protein's structure in tendency of forming helix in regions of $residue_{52-63}$ and $residue_{74-82}$ will deviate from the a99SB-*disp* simulation, and the low-temperature UNRES simulation tends to overestimate the stability of helices. The weak local structural preference suggests structural similarity and the characteristic low tendency for secondary structures. It is interesting to note that, at 270/280 K, the helical fraction is unexpectedly low. In general, as the temperature increases, they begin to sample structures resembling those sampled with a99SB-*disp*, as shown in Figure S2. We also compared the contacts formed in UNRES and a99SB-*disp* ensembles. As

https://doi.org/10.1021/acs.jpcb.2c08945
J. Phys. Chem. B 2023, 127, 2177−2186

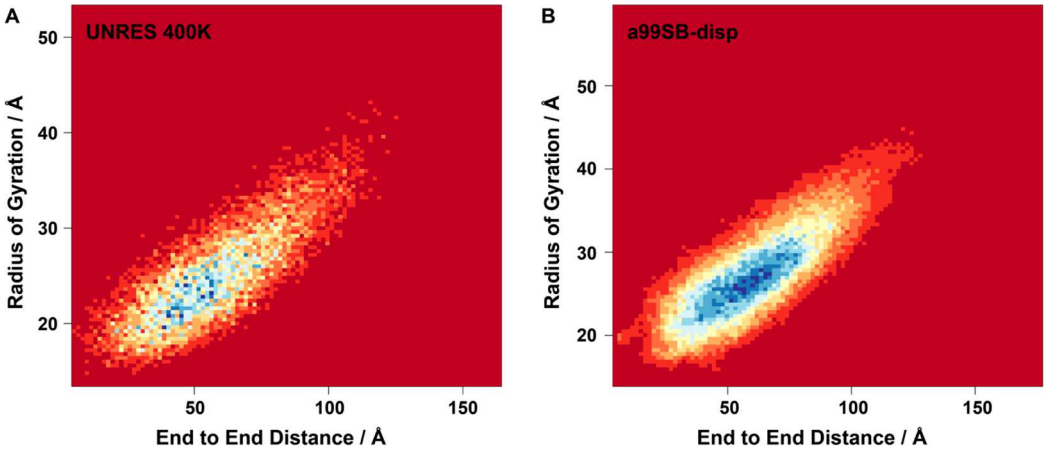


Figure 3. Relative energy surfaces of Ash1 UNRES and a99SB-*disp* ensembles projected onto radius of gyration (Rg) and end-to-end distance.

Table 1. Eight IDPs in the Benchmark Set and Their Amino Acid Compositions

			Amino Acids				
Proteins	Temperature/K	Length	Hydrophobic	Polar	Negative	Positive	Net Charge
PaaA2	320	71	26	14	16	15	−1
N_{TAIL}	385	132	49	35	23	25	2
Ash1	400	83	28	36	1	18	17
drkN SH3	415	59	24	13	13	9	−4
Aβ40	430	40	23	5	6	6	0
ACTR	430	71	33	19	13	6	−7
Sic1	430	92	41	40	0	11	11
α-synuclein	430	140	73	27	24	16	−8

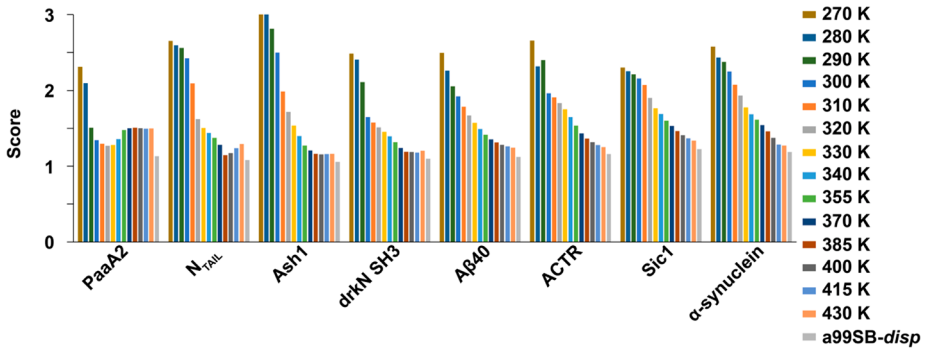


Figure 4. UNRES scores at temperatures from 270 to 430 K and a99SB-disp scores for the benchmark set.

observed in Figure 2C,D, Ash1 exhibits similar contact patterns, indicating that UNRES simulation can predict structures resembling a99SB-*disp* ensembles. Contacts formed at the N-terminus and C-terminus are of higher frequency than contacts formed between $residue_{10-30}$ and $residue_{60-70}$. Contact map analyses of the remaining IDPs in the benchmark set can be found in the Supporting Information (Figure S3).

Additionally, we generated a relative energy surface projected onto the radius of gyration and end-to-end distance. As shown in Figure 3, the Rg and end-to-end distance distributions of the Ash1 UNRES ensemble (400 K) are in good agreement with the a99SB-*disp* ensemble. Since the Ash1 a99SB-*disp* simulation generated 10 times more frames, as shown in Figure 3B, the abundance of data points yielded a clear energy basin. We can also see the structural preference for the same region in the UNRES ensemble, as shown in Figure 3A. However, for some IDP UNRES/a99SB-*disp* ensembles, such as PaaA2 and α-synuclein, a clear deviation of the relative energy surface was observed, as shown in Figure S4.

Validation of the UNRES Force Field with Experimental Measurements. Based on the calculated NMR observable root-mean-square deviation (rmsd) and the Rg penalty, the force field scores were calculated. With the force field scores, we determined the optimal temperatures of UNRES simulations for each IDP in the benchmark set, as shown in Table 1. Subsequent structural analyses demonstrate that structural ensembles yielded by UNRES simulations at the optimal temperature are more similar to experimentally detected structures than at other temperatures. The temperature dependence of force field scores reveals that temperature is a tunable parameter to produce a more reliable structural ensemble using UNRES/REMD simulations. This work also provides insights into the structural prediction of other understudied disordered proteins of interest.

https://doi.org/10.1021/acs.jpcb.2c08945
J. Phys. Chem. B 2023, 127, 2177−2186

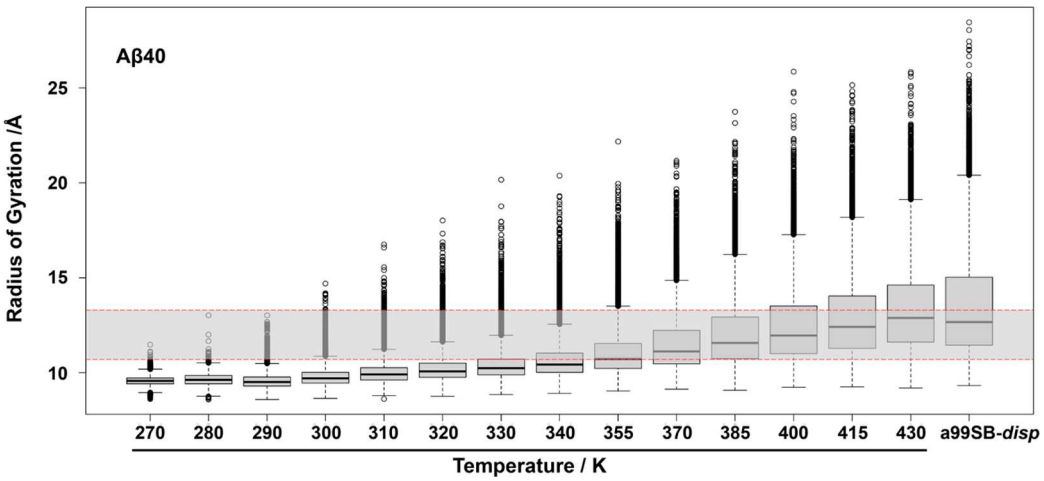


Figure 5. Radius of gyration (Rg) distributions of Aβ40 predicted from UNRES simulations at temperatures spanning from 270 to 430 K, with the experimental values highlighted in gray shadow. The Rg distribution predicted from the Aβ40 a99SB-*disp* ensemble is also shown on the right.

The force field score assesses the similarity between experimental observables obtained from UNRES ensembles and experimental measurements by comparing the overall compactness and local structural features. As shown in Figure 4, IDPs in the benchmark set have higher force field scores (score 1 is comparable to the experimental data) at lower temperatures; as the temperature increases, they begin to sample structures with characteristics resembling experimental states. For example, the UNRES ensembles of Aβ40, ACTR, Sic1, and α-synuclein achieve optimal performance at 430 K. But extremely high temperatures do not always yield the best outcomes, because disordered proteins are not always completely disordered; sometimes, they can form transient secondary structures in some regions. If the simulation temperature is very high, the sparsely populated local structures will be destroyed, leading to structural deviation. For example, PaaA2 reached the best outcomes at 320 K and N_{TAIL} at 385 K, as observed in Figure 4. Consequently, UNRES simulations at the optimal temperature strike a better balance between the disordered character of IDPs and the local structural tendency. Such a temperature dependency feature in structural similarity is similar to the evaluation result against the all-atom ensembles in the previous section. In general, as observed in Figure 4, temperatures above 385 K are favorable for IDP simulations, so NMR observable rmsd, Rg penalty, and the force field score at 400 K with the exception of PaaA2 at 320 K are shown in Table S1. For comparing the accuracy of UNRES ensembles relative to the a99SB-*disp* force field, a99SB-*disp* scores are also shown as the rightmost bar for each IDP in Figure 4. It is obvious that a99SB-*disp* scores are close to 1, suggesting that it achieved the best results among all UNRES and a99SB-*disp* ensembles. And the UNRES ensembles at higher simulation temperatures are more accurate compared to those UNRES ensembles at lower temperatures. In addition, as shown in Table 1, the optimal temperature of the benchmark varies from case to case, because the amino acid composition of IDPs has an essential influence on their secondary structural preference and the overall compactness. The post-transcriptional modification, such as phosphorylation, is such a case in nature, which is a common way to activate IDPs by changing the biophysical properties of specific residues.

Global and Local Structural Analysis. With the force field scores at different temperatures, the optical temperature corresponding to the most reliable structural ensemble can be identified. To evaluate the structural similarity in detail, we further analyzed structural properties such as radius of gyration (Rg) distribution and secondary chemical shifts for the benchmark.

The Rg distribution, indicating the global compactness of proteins, shifts to greater sizes and simultaneously exhibits a broader distribution as the temperature increases. Take Aβ40 as an example, as the temperature increases, the average Rg increases from below 10 Å to falling within the experimentally determined values[79,80] (12.0 ± 1.3 Å). As shown in Figure 5, from the perspective of the Rg distributions, a temperature from 370 to 430 K is a better choice, which achieves comparable results to the a99SB-*disp* ensemble. More factors affecting the IDP structural ensemble (reflected in NMR observable assessment) will help in the final determination of the optimal temperature. Rg distributions of the remaining IDPs in the benchmark set can be found in the Supporting Information (Figures S5 and S6).

Apart from the Rg distribution analysis, NMR secondary chemical shifts, as an indicator closely related to protein structures, are also employed to demonstrate the structural similarity between UNRES structures and experimental detected structures. As observed in Figure 6, secondary

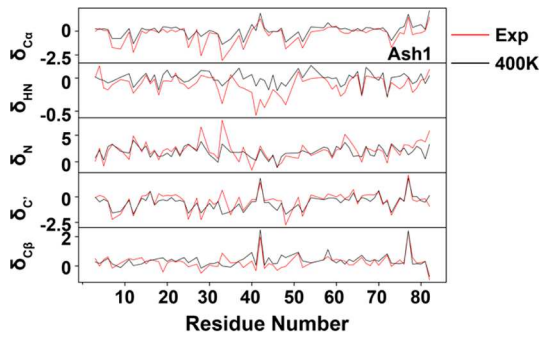


Figure 6. Secondary chemical shifts for atoms C, H, and N of Ash1. The lines for experimental secondary chemical shifts with subtraction of random coil chemical shifts are colored in red, and lines for the UNRES values are in black.

https://doi.org/10.1021/acs.jpcb.2c08945
J. Phys. Chem. B 2023, 127, 2177−2186

chemical shifts of atoms Cα, C, Hα, HN, Cβ, and N, obtained from the Ash1 UNRES ensemble at 400 K, agree with the experimental values, exhibiting some deviations at δHN. A previous work claimed that the backbone $^{3}J_{HNH\alpha}$ scalar coupling constants can be affected by adding amide hydrogen in the idealized position,[81] which may also slightly affect other NMR observables. It is known that secondary chemical shifts are closely related to residues' structural preference to either α-helix or β-sheet; thereby, the consistence illustrates that residues adopt alike secondary structures.

DISCUSSION

The UNRES Force Field can Simulate Disordered Proteins with Reasonable Accuracy. By tuning the simulation temperature, the UNRES simulation can generate structural ensembles similar to the a99SB-*disp* simulations at the optimal temperature. Through the study of RMSF, secondary structural analysis, the contact map, and the relative energy surface in detail, we found that the UNRES force field can probe some local structural properties and predict conformers with contact patterns similar to the a99SB-*disp* simulations. Also, relative energy surfaces projected onto Rg and the end-to-end distance suggest that most of the IDP UNRES ensembles at the optimal simulation temperature exhibit good agreement with the corresponding a99SB-*disp* ensembles, as shown in Figure 3 and Figure S4. Furthermore, the NMR observables and Rg predicted from UNRES ensembles were compared with the experimental values, indicating the temperature dependence in force field scores. Rg analyses suggest that the UNRES model is capable of yielding less compacted conformers at high temperatures. And the local structural similarity can be demonstrated by the secondary chemical shift comparisons in Figure 6 and Figure S7. Therefore, the UNRES force field can be optimized for the structural prediction of disordered proteins with reasonable accuracy by simply adjusting the temperature. In addition, we need to keep in mind that the UNRES force field, a physics-based coarse-grained model, can simulate proteins more efficiently than the currently popular all-atom force fields optimized for disordered proteins with moderate sacrifice on accuracy. And simulations can be performed in the freely accessible UNRES web server.[42]

The temperature indeed improves the simulation quality; however, in some cases, we noticed that the structural ensembles generated from the UNRES simulations, even simulation trajectories at the optimal temperature, deviated from that obtained from the all-atom a99SB-*disp* force field and from the experimental data. As observed in Figure 4, in general, the a99SB-*disp* force field shows a better agreement with the experimental data compared to the UNRES force field. While the UNRES force field may not exhibit the same level of accuracy as a99SB-disp, it is a physics-based coarse-grained force field that can yield an excellent sampling of the conformational space with appropriate temperature selection, without the need for high computational demands. It is worthwhile to note that, in spite of the structural deviation from a99SB-*disp*, structural features calculated from the a99SB-*disp* are not exactly the same as those observed in experiments, such as Rg distributions in Figures S5 and S6, which highlight the challenges in studying IDPs using computer simulations. Moreover, it takes a few weeks to simulate proteins consisting of more than 100 residues. Based on the limitations of our techniques, there are two ways to further improve the performance: (1) the UNRES force field needs to be parametrized specifically for IDPs and (2) experimental data can be introduced to guide protein simulations. In particular, the combination of a variety of experimental techniques and computational methods has shown the power to explore complex biological systems from different aspects.[12,82]

ASSOCIATED CONTENT

Data Availability Statement

All simulation data and scripts can be downloaded at https://tinyurl.com/5n8s697z.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.2c08945.

NMR observable root-mean-square deviation (rmsd), Rg penalty, and scores of UNRES ensembles, root-mean-square-fluctuation of proteins, helical fraction of proteins, contact map of UNRES ensembles, relative energy surfaces of UNRES and a99SB-disp ensembles, and radius of gyration distributions, secondary chemical shifts (PDF)

AUTHOR INFORMATION

Corresponding Author

Yi He − *Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico 87131, United States; Translational Informatics Division, Department of Internal Medicine, University of New Mexico, Albuquerque, New Mexico 87131, United States;* orcid.org/0000-0002-6884-5312; Email: yihe@unm.edu

Authors

Tongtong Li − *Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico 87131, United States*

Emily Hendrix − *Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico 87131, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c08945

Author Contributions

The manuscript was a collaborative effort of all authors, who have made significant contributions. All authors have reviewed and approved the final version of the manuscript, ensuring that it accurately reflects their collective work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was funded by the National Science Foundation (Grant No. 2137558). This work was also supported by the Substance Use Disorders Grand Challenge Pilot Research Award, the Research Allocations Committee (RAC) Award, the startup fund from the University of New Mexico, and the University of New Mexico Office of the Vice President for Research WeR1 Faculty Success Program. We are grateful to Dr. Robustelli for their invaluable comments and support, which greatly enhanced the quality of the manuscript. Their assistance with the data analysis scripts and experimental data was greatly appreciated and helped to make this work a success.

REFERENCES

(1) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. Sequence Complexity of Disordered Protein. *Proteins Struct. Funct. Genet.* **2001**, *42*, 38−48.

(2) Vucetic, S.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. Flavors of Protein Disorder. *Proteins Struct. Funct. Genet.* **2003**, *52*, 573−584.

(3) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. Speeding Molecular Recognition by Using the Folding Funnel: The Fly-Casting Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8868−8873.

(4) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by x-Ray Analysis. *Nature* **1958**, *181*, 662−666.

(5) Henderson, R. From Electron Crystallography to Single Particle CryoEM (Nobel Lecture). *Angew. Chemie - Int. Ed.* **2018**, *57*, 10804−10825.

(6) Frank, J. Single-Particle Reconstruction of Biological Molecules—Story in a Sample (Nobel Lecture). *Angew. Chemie - Int. Ed.* **2018**, *57*, 10826−10841.

(7) Dubochet, J. On the Development of Electron Cryo-Microscopy (Nobel Lecture). *Angew. Chemie - Int. Ed.* **2018**, *57*, 10842−10846.

(8) Rout, M. P.; Sali, A. Principles for Integrative Structural Biology Studies. *Cell* **2019**, *177*, 1384−1403.

(9) Geraets, J. A.; Pothula, K. R.; Schröder, G. F. Integrating Cryo-EM and NMR Data. *Curr. Opin. Struct. Biol.* **2020**, *61*, 173−181.

(10) Bernadó, P.; Svergun, D. I. Analysis of Intrinsically Disordered Proteins by Small-Angle X-Ray Scattering. In *Methods Mol. Biol.*; Uversky, V. N., Dunker, A. K., Eds.; Springer New York: New York, NY, 2012; Vol. *896*, pp 107−122.

(11) Leblanc, S. J.; Kulkarni, P.; Weninger, K. R. Single Molecule FRET: A Powerful Tool to Study Intrinsically Disordered Proteins. *Biomolecules* **2018**, *8*, 140.

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

(13) Das, P.; Matysiak, S.; Mittal, J. Looking at the Disordered Proteins through the Computational Microscope. *ACS Cent. Sci.* **2018**, *4*, 534−542.

(14) Bhattacharya, S.; Lin, X. Recent Advances in Computational Protocols Addressing Intrinsically Disordered Proteins. *Biomolecules* **2019**, *9*, 146.

(15) Gil Pineda, L. I.; Milko, L. N.; He, Y. Performance of CHARMM36m with Modified Water Model in Simulating Intrinsically Disordered Proteins: A Case Study. *Biophys. Reports* **2020**, *6*, 80−87.

(16) Reid, L. M.; Guzzetti, I.; Svensson, T.; Carlsson, A. C.; Su, W.; Leek, T.; von Sydow, L.; Czechtizky, W.; Miljak, M.; Verma, C.; et al. How Well Does Molecular Simulation Reproduce Environment-Specific Conformations of the Intrinsically Disordered Peptides PLP, TP2 and ONEG? *Chem. Sci.* **2022**, *13*, 1957−1971.

(17) 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.

(18) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(19) 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.

(20) 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*, 5513−5524.

(21) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E4758−E4766.

(22) Mu, J.; Liu, H.; Zhang, J.; Luo, R.; Chen, H. F. Recent Force Field Strategies for Intrinsically Disordered Proteins. *J. Chem. Inf. Model.* **2021**, *61*, 1037−1047.

(23) He, Y. Y.; Xiao, Y.; Liwo, A.; Scheraga, H. A. Exploring the Parameter Space of the Coarse-Grained UNRES Force Field by Random Search: Selecting a Transferable Medium-Resolution Force Field. *J. Comput. Chem.* **2009**, *30*, 2127−2135.

(24) Liwo, A.; He, Y.; Scheraga, H. A. Coarse-Grained Force Field: General Folding Theory. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16890−16901.

(25) Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagieła, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A.; et al. A Unified Coarse-Grained Model of Biological Macromolecules Based on Mean-Field Multipole-Multipole Interactions. *J. Mol. Model.* **2014**, *20*, 2306.

(26) Sieradzan, A. K.; Krupa, P.; Scheraga, H. A.; Liwo, A.; Czaplewski, C. Physics-Based Potentials for the Coupling between Backbone- and Side-Chain-Local Conformational States in the United Residue (UNRES) Force Field for Protein Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 817−831.

(27) Krupa, P.; Hałabis, A.; Zmudzińska, W.; Ołdziej, S.; Scheraga, H. A.; Liwo, A. Maximum Likelihood Calibration of the UNRES Force Field for Simulation of Protein Structure and Dynamics. *J. Chem. Inf. Model.* **2017**, *57*, 2364−2377.

(28) Sieradzan, A. K.; Makowski, M.; Augustynowicz, A.; Liwo, A. A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. I. Backbone Potentials of Coarse-Grained Polypeptide Chains. *J. Chem. Phys.* **2017**, *146*, 124106.

(29) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. Cumulant-Based Expressions for the Multibody Terms for the Correlation between Local and Electrostatic Interactions in the United-Residue Force Field. *J. Chem. Phys.* **2001**, *115*, 2323−2347.

(30) Liwo, A.; Khalili, M.; Scheraga, H. A. Ab Initio Simulations of Protein-Folding Pathways by Molecular Dynamics with the United-Residue Model of Polypeptide Chains. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2362−2367.

(31) He, Y.; Zhou, R.; Huang, Y.; Xiao, Y. Foldable Subunits of Helix Protein. *Comput. Biol. Chem.* **2009**, *33*, 325−328.

(32) Maisuradze, G. G.; Senet, P.; Czaplewski, C.; Liwo, A.; Scheraga, H. A. Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field. *J. Phys. Chem. A* **2010**, *114*, 4471−4485.

(33) Rojas, A.; Liwo, A.; Browne, D.; Scheraga, H. A. Mechanism of Fiber Assembly: Treatment of Aβ Peptide Aggregation with a Coarse-Grained United-Residue Force Field. *J. Mol. Biol.* **2010**, *404*, 537−552.

(34) He, Y.; Liwo, A.; Weinstein, H.; Scheraga, H. A. PDZ Binding to the BAR Domain of PICK1 Is Elucidated by Coarse-Grained Molecular Dynamics. *J. Mol. Biol.* **2011**, *405*, 298−314.

(35) Gołaś, E.; Maisuradze, G. G.; Senet, P.; Ołdziej, S.; Czaplewski, C.; Scheraga, H. A.; Liwo, A. Simulation of the Opening and Closing of Hsp70 Chaperones by Coarse-Grained Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 1750−1764.

(36) Stevens, A. O.; He, Y. Residue-Level Contact Reveals Modular Domain Interactions of PICK1 Are Driven by Both Electrostatic and Hydrophobic Forces. *Front. Mol. Biosci.* **2021**, *7*, 616135.

(37) He, Y.; Mozolewska, M. A.; Krupa, P.; Sieradzan, A. K.; Wirecki, T. K.; Liwo, A.; Kachlishvili, K.; Rackovsky, S.; Jagiela, D.; Ślusarz, R.; et al. Lessons from Application of the UNRES Force Field to Predictions of Structures of CASP10 Targets. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14936−14941.

(38) Krupa, P.; Mozolewska, M. A.; Wiśniewska, M.; Yin, Y.; He, Y.; Sieradzan, A. K.; Ganzynkowicz, R.; Lipska, A. G.; Karczyńska, A.; Slusarz, M.; et al. Performance of Protein-Structure Predictions with the Physics-Based UNRES Force Field in CASP11. *Bioinformatics* **2016**, *32*, 3270−3278.

(39) Karczyńska, A.; Mozolewska, M. A.; Krupa, P.; Giełdoń, A.; Bojarski, K. K.; Zaborowski, B.; Liwo, A.; Ślusarz, R.; Ślusarz, M.; Lee,

https://doi.org/10.1021/acs.jpcb.2c08945
J. Phys. Chem. B 2023, 127, 2177−2186

J.; et al. Use of the UNRES Force Field in Template-Assisted Prediction of Protein Structures and the Refinement of Server Models: Test with CASP12 Targets. *J. Mol. Graph. Model.* **2018**, *83*, 92−99.
(40) Karczyńska, A. S.; Zięba, K.; Uciechowska, U.; Mozolewska, M. A.; Krupa, P.; Lubecka, E. A.; Lipska, A. G.; Sikorska, C.; Samsonov, S. A.; Sieradzan, A. K.; et al. Improved Consensus-Fragment Selection in Template-Assisted Prediction of Protein Structures with the UNRES Force Field in CASP13. *J. Chem. Inf. Model.* **2020**, *60*, 1844−1864.
(41) Antoniak, A.; Biskupek, I.; Bojarski, K. K.; Czaplewski, C.; Giełdoń, A.; Kogut, M.; Kogut, M. M.; Krupa, P.; Lipska, A. G.; Liwo, A.; et al. Modeling Protein Structures with the Coarse-Grained UNRES Force Field in the CASP14 Experiment. *J. Mol. Graph. Model.* **2021**, *108*, 108008.
(42) Czaplewski, C.; Karczyńska, A.; Sieradzan, A. K.; Liwo, A. UNRES Server for Physics-Based Coarse-Grained Simulations and Prediction of Protein Structure, Dynamics and Thermodynamics. *Nucleic Acids Res.* **2018**, *46*, W304−W309.
(43) Sterckx, Y. G. J.; Volkov, A. N.; Vranken, W. F.; Kragelj, J.; Jensen, M. R.; Buts, L.; Garcia-Pino, A.; Jové, T.; Van Melderen, L.; Blackledge, M.; et al. Small-Angle X-Ray Scattering- and Nuclear Magnetic Resonance-Derived Conformational Ensemble of the Highly Flexible Antitoxin PaaA2. *Structure* **2014**, *22*, 854−865.
(44) Jensen, M. R.; Communie, G.; Ribeiro, E. A.; Martinez, N.; Desfosses, A.; Salmon, L.; Mollica, L.; Gabel, F.; Jamin, M.; Longhi, S.; et al. Intrinsic Disorder in Measles Virus Nucleocapsids. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 9839−9844.
(45) Martin, E. W.; Holehouse, A. S.; Grace, C. R.; Hughes, A.; Pappu, R. V.; Mittag, T. Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. *J. Am. Chem. Soc.* **2016**, *138*, 15323−15335.
(46) Zhang, O.; Forman-Kay, J. D. Structural Characterization of Folded and Unfolded States of an SH3 Domain in Equilibrium in Aqueous Buffer. *Biochemistry* **1995**, *34*, 6784−6794.
(47) Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E. The Alzheimer's Peptides Aβ40 and 42 Adopt Distinct Conformations in Water: A Combined MD/NMR Study. *J. Mol. Biol.* **2007**, *368*, 1448−1457.
(48) Iešmantavičius, V.; Jensen, M. R.; Ozenne, V.; Blackledge, M.; Poulsen, F. M.; Kjaergaard, M. Modulation of the Intrinsic Helix Propensity of an Intrinsically Disordered Protein Reveals Long-Range Helix-Helix Interactions. *J. Am. Chem. Soc.* **2013**, *135*, 10155−10163.
(49) Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D. Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure* **2010**, *18*, 494−506.
(50) 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. U. S. A.* **2005**, *102*, 1430−1435.
(51) Rhee, Y. M.; Pande, V. S. Multiplexed-Replica Exchange Molecular Dynamics Method for Protein Folding Simulation. *Biophys. J.* **2003**, *84*, 775−786.
(52) Czaplewski, C.; Kalinowski, S.; Liwo, A.; Scheraga, H. A. Application of Multiplexed Replica Exchange Molecular Dynamics to the UNRES Force Field: Tests with α and A+β Proteins. *J. Chem. Theory Comput.* **2009**, *5*, 627−640.
(53) Liwo, A.; Oldziej, S.; Czaplewski, C.; Kleinerman, D. S.; Blood, P.; Scheraga, H. A. Implementation of Molecular Dynamics and Its Extensions with the Coarse-Grained UNRES Force Field on Massively Parallel Systems: Toward Millisecond-Scale Simulations of Protein Structure, Dynamics, and Thermodynamics. *J. Chem. Theory Comput.* **2010**, *6*, 890−909.
(54) Zaborowski, B.; Jagieła, D.; Czaplewski, C.; Hałabis, A.; Lewandowska, A.; Zmudzińska, W.; Ołdziej, S.; Karczyńska, A.; Omieczynski, C.; Wirecki, T.; et al. A Maximum-Likelihood Approach to Force-Field Calibration. *J. Chem. Inf. Model.* **2015**, *55*, 2050−2070.
(55) He, Y.; Liwo, A.; Scheraga, H. A. Optimization of a Nucleic Acids United-RESidue 2-Point Model (NARES-2P) with a Maximum-Likelihood Approach. *J. Chem. Phys.* **2015**, *143*, 243111.
(56) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Ołdziej, S.; Wachucik, K.; Scheraga, H. A. Modification and Optimization of the United-Residue (UNRES) Potential Energy Function for Canonical Simulations. I. Temperature Dependence of the Effective Energy Function and Tests of the Optimization Method with Single Training Proteins. *J. Phys. Chem. B* **2007**, *111*, 260−285.
(57) Rhee, Y. M.; Pande, V. S. Multiplexed-Replica Exchange Molecular Dynamics Method for Protein Folding Simulation. *Biophys. J.* **2003**, *84*, 775−786.
(58) Liwo, A.; Sieradzan, A. K.; Lipska, A. G.; Czaplewski, C.; Joung, I.; Zmudzińska, W.; Hałabis, A.; Ołdziej, S. A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. III. Determination of Scale-Consistent Backbone-Local and Correlation Potentials in the UNRES Force Field and Force-F. *J. Chem. Phys.* **2019**, *150*, 155104.
(59) Rotkiewicz, P.; Skolnick, J. Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations. *J. Comput. Chem.* **2008**, *29*, 1460−1465.
(60) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031−3048.
(61) 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.
(62) Zweckstetter, M.; Bax, A. Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR [11]. *J. Am. Chem. Soc.* **2000**, *122*, 3791−3792.
(63) Zweckstetter, M. NMR: Prediction of Molecular Alignment from Structure Using the PALES Software. *Nat. Protoc.* **2008**, *3*, 679−690.
(64) Marsh, J. A.; Baker, J. M. R.; Tollinger, M.; Forman-Kay, J. D. Calculation of Residual Dipolar Couplings from Disordered State Ensembles Using Local Alignment. *J. Am. Chem. Soc.* **2008**, *130*, 7804−7805.
(65) Karplus, M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.* **1959**, *30*, 11−15.
(66) Vuister, G. W.; Bax, A. Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J(HNHα) Coupling Constants in 15N-Enriched Proteins. *J. Am. Chem. Soc.* **1993**, *115*, 7772−7777.
(67) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.
(68) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. 1H, 13C and 15N Random Coil NMR Chemical Shifts of the Common Amino Acids. I. Investigations of Nearest-Neighbor Effects. *J. Biomol. NMR* **1995**, *5*, 332.
(69) Spera, S.; Bax, A. Empirical Correlation between Protein Backbone Conformation and C.Alpha. and C.Beta. 13C Nuclear Magnetic Resonance Chemical Shifts. *J. Am. Chem. Soc.* **1991**, *113*, 5490−5492.
(70) 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.
(71) Zhang, Y.; Liu, H.; Yang, S.; Luo, R.; Chen, H. F. Well-Balanced Force Field Ff03 CMAP for Folded and Disordered Proteins. *J. Chem. Theory Comput.* **2019**, *15*, 6769−6780.

https://doi.org/10.1021/acs.jpcb.2c08945
J. Phys. Chem. B 2023, 127, 2177−2186

(72) Latham, A. P.; Zhang, B. Maximum Entropy Optimized Force Field for Intrinsically Disordered Proteins. *J. Chem. Theory Comput.* **2020**, *16*, 773−781.
(73) Qiu, Y.; Shan, W.; Zhang, H. Force Field Benchmark of Amino Acids. 3. Hydration with Scaled Lennard-Jones Interactions. *J. Chem. Inf. Model.* **2021**, *61*, 3571−3582.
(74) Berjanskii, M. V.; Wishart, D. S. Unraveling the Meaning of Chemical Shifts in Protein NMR. *Biochim. Biophys. Acta - Proteins Proteomics* **2017**, *1865*, 1564−1576.
(75) Saupe, A.; Englert, G. Saupe_NMROrientatedMole_PhyRevLet_1963. *Phys. Rev. Lett.* **1963**, *11*, 462−464.
(76) Bothner-By, A. A. In *Magnetic Field Induced Alignment of Molecules, in "Encyclopedia of Nuclear Magnetic Resonance*; Grant, D. M., Harris, R. K., Eds.; Wiley: Chichester, 1996; pp 2932−2938.
(77) Sanchez-Martinez, M.; Crehuet, R. Application of the Maximum Entropy Principle to Determine Ensembles of Intrinsically Disordered Proteins from Residual Dipolar Couplings. *Phys. Chem. Chem. Phys.* **2014**, *16*, 26030−26039.
(78) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J. A.; Dyson, H. J.; Wright, P. E. Structural Characterization of Unfolded States of Apomyoglobin Using Residual Dipolar Couplings. *J. Mol. Biol.* **2004**, *340*, 1131−1142.
(79) Granata, D.; Baftizadeh, F.; Habchi, J.; Galvagnion, C.; De Simone, A.; Camilloni, C.; Laio, A.; Vendruscolo, M. The Inverted Free Energy Landscape of an Intrinsically Disordered Peptide by Simulations and Experiments. *Sci. Rep.* **2015**, *5*, 15449.
(80) Choy, W. Y.; Mulder, F. A. A.; Crowhurst, K. A.; Muhandiram, D. R.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E. Distribution of Molecular Size within an Unfolded State Ensemble Using Small-Angle X-Ray Scattering and Pulse Field Gradient NMR Techniques. *J. Mol. Biol.* **2002**, *316*, 101−112.
(81) Vögeli, B.; Ying, J.; Grishaev, A.; Bax, A. Limits on Variations in Protein Backbone Dynamics from Precise Measurements of Scalar Couplings. *J. Am. Chem. Soc.* **2007**, *129*, 9377−9385.
(82) Chan-Yao-Chong, M.; Durand, D.; Ha-Duong, T. Molecular Dynamics Simulations Combined with Nuclear Magnetic Resonance and/or Small-Angle X-Ray Scattering Data for Characterizing Intrinsically Disordered Protein Conformational Ensembles. *J. Chem. Inf. Model.* **2019**, *59*, 1743−1758.

https://doi.org/10.1021/acs.jpcb.2c08945
J. Phys. Chem. B 2023, 127, 2177−2186