Refining all-atom protein force fields for polar-rich, prion-like, low complexity intrinsically disordered proteins

Wai Shing Tang[†], **Nicolas L. Fawzi**[¶], **Jeetain Mittal**[‡,*]

[†] Department of Physics, Brown University, Providence, RI 02912, USA

[¶] Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, Providence, RI 02912, USA

[‡] Department of Chemical and Biomolecular Engineering, Lehigh University, Bethlehem, PA 18015, USA

Abstract

Significant efforts in the last decade have given us highly accurate all-atom protein force fields for molecular dynamics (MD) simulations of folded and disordered proteins. These simulations, complemented with experimental data, provide new insights into molecular interactions that underlie the physical properties of proteins, especially for intrinsically disordered proteins (IDPs) for which defining the heterogeneous structural ensemble is hugely challenging by experiments alone. Consequently, the accuracy of these protein force fields is of utmost importance in order to ensure reliable simulated conformational data. Here, we first assess the accuracy of current state-of-the-art force fields for IDPs (ff99SBws and ff03ws) applied to disordered proteins of low amino acid sequence complexity that can undergo liquid-liquid phase separation. Based on a detailed comparison of NMR chemical shifts between simulation and experiment on several IDPs, we find that regions surrounding specific polar residues result in simulated ensembles with exaggerated helicity when compared to experiment. To resolve this discrepancy, we introduce residue-specific modifications to the backbone torsion potential of three residues (Ser, Thr, Gln) in the ff99SBws force field. The modified force field, ff99SBws-STQ, provides a more accurate representation of helical structure propensity in these LC domains without compromising faithful representation of helicity in a region with distinct sequence composition. Our refinement strategy also suggests a path forward for integrating experimental data in the assessment of residue-specific deficiencies in the current physics-based force fields and improve these force fields further for their broader applicability.

Graphical Abstract

[*] jem309@lehigh.edu.
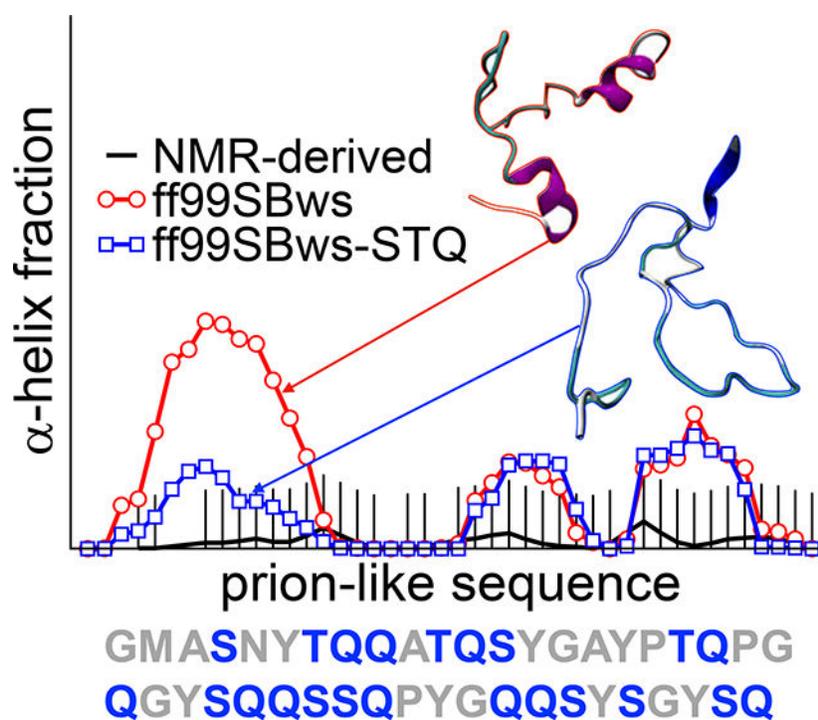
Introduction

Intrinsically disordered proteins (IDPs) and regions (IDRs) are biologically functional without adopting a single well-defined folded structure.[1] They are present in a significant fraction in the proteome of all organisms[1–3] and participate in essential physiological and pathological functions including stress response[4] and signaling.[5] They are also involved in cell regulation and various neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) through the formation of protein aggregates.[6] IDPs are also involved in cellular liquid-liquid phase separation (LLPS), which underlies the formation of some membraneless organelles[7,8] and may serve as precursors for pathogenic aggregates.[9,10] Many IDPs can contain partial secondary structures, including helices[11,12] and β-sheet structures,[13,14] that can contribute to their intermolecular interactions and self-assembly processes. Therefore, accurately probing these transient secondary structures in experiments and simulations is essential for understanding the underlying molecular mechanisms of protein assembly in LLPS.

Experimental techniques such as X-ray crystallography and cryogenic electron microscopy (cryo-EM) can resolve folded protein structures down to angstrom-scale but are unable to provide detailed spatiotemporal information on the equilibrium structural ensembles of disordered proteins because these techniques and sample preparations quench and/or cannot observe the heterogeneous nature and rapid interconversion between a large number of microstates.[15] Nuclear magnetic resonance (NMR) in solution has provided a wealth of information on the local structural and dynamical properties of IDPs to generate important insights into their biological function. It is often helpful to combine NMR data with molecular modeling techniques to generate conformational ensembles consistent with the

experimental observables such as chemical shifts and J-coupling to obtain atomic details obscured by averaging.[16–20] It is also highly desirable to obtain the necessary information on IDP properties directly from physics-based all-atom transferable force fields without requiring any experimental input on the system of interest. Such models can provide unprecedented details on the structural and dynamical properties as well as atomic interactions responsible for the observed behavior.[21,22] This can be critical in establishing IDP sequence-structure/motions-function relationships and provide hypotheses for experimental work.[23]

With the rapid development of computer hardware and advanced sampling techniques to conduct molecular dynamics (MD) simulations based on atomistic protein force fields over the past decade, it became apparent that improvements in their accuracy and transferability were necessary to realize the potential for molecular simulations to serve as a "computational microscope."[24,25] Progressive efforts in the last 10–15 years have resulted in significant improvements in the accuracy of all-atom protein force fields beginning with a substantial step forward on achieving secondary structure balance and thereby demonstrating true transferability in terms of their application to proteins of different structural classes. [26–29] The empirical approach by Best and Hummer to use helical propensity data on a short peptide,[26] Ac-(AAQAA)$_3$-NH$_2$, to achieve such secondary structure balance also demonstrated that the potential energy functions only needed fine-tuning as opposed to a significant reparameterization as in previous force field development efforts.

This strategy to modify only backbone-torsion-potentials yielded many improved force field variants, marked by a symbol '*' next to their name, such as ff03*, ff99SB*, charmm22*. [26,30] In these force fields, modifications on backbone torsion potentials were applied uniformly to all amino acids (except proline and glycine) and was not tuned to capture any residue-specific biases in the force field. Since then, several studies have taken additional steps to make residue-specific corrections to side-chain[31] and backbone[32] torsion potentials, as well as more elaborate bottoms-up parameterization of energy function parameters,[33–36] resulting in highly accurate protein force fields.

The use of these optimized force fields has been especially fruitful for studying IDPs by overcoming challenges of using the experimental techniques alone.[37–40] At the same time, the flat conformational energy landscapes of IDPs can make conformational ensembles more sensitive to systematic errors in the model. Therefore, simulations of IDPs have been used successfully for evaluating the remaining deficiencies of these force fields and providing roadmaps for their continued improvement.[41] For example, the unfolded states of proteins remained too collapsed in the force fields tuned for secondary structure balance alone (such as ff03*), which was also corroborated by available experimental data from single-molecule FRET and SAXS.[42] Some improvements in preventing unfolded state collapse were visible by changing the water model (TIP4P/2005)[43] in protein simulations while trying to improve the temperature cooperativity of the helix-coil transition and the associated solvation effects. The new protein force fields, marked by symbol 'w' for an improved water model TIP4P/2005 (ff03w and ff99SBw), provided significant improvement in protein-protein interactions but only modest success in capturing IDP dimensions.[44] Based on the accumulated evidence from other studies, protein-water interactions were scaled (strengthened)[45] to model the size

of solvated disordered proteins accurately with force fields suffixed by 'ws' such as ff03ws and ff99SBws.[46] Our previous work based on these optimization strategies has leveraged the force field improvements to probe the secondary structure, single-chain configurations, and contact formation of IDPs and demonstrated their applicability to reproduce the structure and dynamics of unfolded proteins accurately. [47–50,]

At this point, it is instructive to ask if there are areas of concern about the applicability of these refined force fields for IDPs. The recent surge of interest in biomolecular LLPS and the underlying atomic interactions responsible for stabilizing the protein-rich condensed phase has motivated us to look at low complexity (LC) IDPs[51] in which a few amino acid types dominate their composition, and at prion-like domains, named for their polar-rich residue sequence composition that resembles the composition of yeast prion proteins.[52] These LC IDPs present new challenges in conducting accurate simulations as any minor deficiencies in particular amino acid types that dominate the sequence composition can propagate additively. Therefore, we believe that LC protein sequences are useful benchmarks to assess the accuracy of current state-of-the-art all-atom force fields.

Methods

All-atom MD simulations were conducted on 44-residue fragments from the disordered regions of FUS LC (FUS LC_{0-43}, FUS LC_{37-97}, FUS LC_{41-84}, FUS LC_{77-120}, FUS $LC_{120-163}$), TDP-43 CTD (TDP-$43_{310-350}$), RNA Pol II CTD heptads (RNA Pol $II_{1927-1970}$), and hnRNPA2 LC (hnRNPA2$_{190-233}$, hnRNPA2$_{265-308}$). The length of the peptides was chosen, such that sampling the conformational space of the protein is computationally tractable. The GROMACS 4.6.7 MD engine[53] was used for simulations, with PLUMED 2.2.4 plugin.[54] Simulations were conducted using different variants of Amber force fields as specified in the text (see SI Tables S2–S4), using the original and modified versions of ff03ws and ff99SBws[46] which includes TIP4P/2005 explicit water molecules.[43] Production simulations were conducted using replica exchange molecular dynamics (REMD)[55] in a well-tempered ensemble (WTE),[56,57] a type of metadynamics method.[58] This combination is known as parallel tempering in the well-tempered ensemble (PT-WTE) and is useful for reducing the number of replicas required to explore the conformational space efficiently. Temperature is held constant for each replica using Langevin dynamics as a thermostat with a friction constant of 1 ps^{-1} and a time step of 2 fs. Electrostatic interactions are calculated using the particle-mesh Ewald method[59] up to fourth order with a real-space cutoff distance of 1.0 nm. A 1.0 nm cutoff distance is used for the van der Waals interactions. For each protein, a random protein structure is generated and solvated in a truncated octahedron box with a face-to-face distance of 6.5 nm. The system is initially equilibrated for 100 ps at each temperature used for parallel tempering. In our simulations, 16 replicas at temperatures ranged from 300 to 500 K are constructed. Differences in temperatures between the replicas are determined such that the exchange acceptance probability is about 30%. The positions and velocities of the thermally equilibrated system are then used as the initial configuration of the PT-WTE simulations. WTE was applied using Gaussian functions of width $\sigma_i = 750$ kJ/mol added at a $\tau_G = 20$ ps interval, with a bias factor of $\gamma = (T + \Delta T)/\Delta T = 36$ and initial hill height of 2.5 kJ/mol. Chemical shifts and helical propensities in these simulations are calculated using the SPARTA+ algorithm[60] and secondary structure is computed using

DSSP.[61] Experimental secondary structure propensities were also be derived from NMR secondary chemical shifts by using the δ2d software.[62] RMSD (root-mean-squared deviation) from experimental chemical shifts (see SI text) was calculated as $(x_{sim} - x_{expt})^2/N$ where N is the number of residues sampled, x_{sim} is the simulated chemical shifts from SPARTA+ and x_{expt} is the NMR chemical shifts. For each peptide, two residues from each terminus and residues missing from experimental data are not included in the analysis. RMSD is calculated for each residue type based on the chemical shift differences between simulations and experiments. The relevant input files can be downloaded at https://bitbucket.org/jeetain/all-atom_ff_refinements.

Results and Discussion

We conduct parallel tempering (PT) MD simulations in a well-tempered ensemble (WTE)[56] to enhance conformational sampling (see Methods) with two suitable force fields for IDPs (ff03ws and ff99SBws) on a set of prion-like domains including domains from TDP-43, Fused in Sarcoma (FUS), and hnRNPA2, as well as the C-terminal heptad repeat domain of RNA Polymerase II (RNA Pol II),.[47,48,50,63] Similar to our previous work, we use 44-residue fragments of these proteins to reduce the computational cost associated with simulations of full-length proteins using a multi-replica approach necessary to obtain converged equilibrium properties. We use the lowest temperature (300 K) replica for all the analyses presented in this paper.

First, we focus on the results of two protein fragments, TDP-43$_{310-350}$ and FUS LC$_{0-43}$, as an example of IDPs where NMR data suggest partial helical structure and complete disorder, respectively. We calculate the chemical shifts from the simulated ensembles using the SPARTA+ algorithm[60] and subtract the sequence-based random-coil values from the Poulsen webserver[64] to compute secondary chemical shifts, which can be used to infer secondary structure populations. Figure 1a shows an excellent agreement between the simulation and experimental data in the helix-forming region (δC_α- δC_β values, a well-established metric of helical (positive values) or sheet (negative values) computed by calculating the difference between the C_α and C_β observed chemical shifts and those expected for a completely random coil disordered protein) at residues 321–330 in TDP-43$_{310-350}$, an alanine rich region embedded in the LC domain.[12] However, in the case of FUS LC$_{0-43}$, simulation data suggests the formation of helical structures in the N-terminal region, residues 6–13, which is not supported by the experimental data.[22,65] We also compute the secondary structure propensities from the simulation data directly using the DSSP algorithm[61] and compare these to the predicted values based on the experimental NMR chemical shift data using the δ2d webserver[62] (Fig. 1b). Again, we see excellent agreement between simulation and experiment regarding the presence of partial helicity in the alanine-rich segment of TDP-43, but apparent overpopulation of helix in several regions of the polar rich FUS LC. Overall, both the force fields display similar behavior with some minor differences.

Robustelli *et al.* had shown a similar behavior for ff03ws on other IDPs with various partially populated secondary structures and proposed a new force field to provide a better representation of folded and disordered proteins.[33] We also conducted additional simulations using their ff99SB-disp force field.[33] These data are shown in supporting information (SI)

Fig. S1. Consistent with previous findings that partial helical structures may be underpopulated in ff99SB-disp[33], simuations with ff99SB-disp show underpopulation of helix in TDP-43$_{310-350}$ while lower overall helicity for the disordered FUS LC$_{0-43}$. However, it is interesting that the simulated ensemble using ff99SB-disp also populates α-helical structures, like ff03ws and ff99SBws, in the N-terminal part of FUS LC$_{0-43}$ which are not present in the experimental data.

The analysis above clearly suggests that further fine-tuning of the current state-of-the-art force fields is needed to make them suitable for low-complexity sequences such as FUS LC$_{0-43}$. The observed overpopulation of helical structures in the case of FUS LC$_{0-43}$ is most likely related to the presence of specific amino acids for which the balance in helical and extended structures is not optimal. In our previous work, we had used the Lifson-Roig helix growth parameter (w) to assess residue-level helical propensities with a model Ala-based host-guest peptide that helped us identify residues with significant deviations from the experiment and propose refinements.[66] Moving forward, it will be useful to develop a strategy that can take advantage of the available NMR data on the sequences of interest and does not require us to interpret the secondary structure propensities in terms of a helix-coil transition model in the context of a host-guest peptide.

Here, we calculate the differences between the secondary chemical shifts between the simulation and experimental data, $(\delta C_\alpha - \delta C_\beta) = (\delta C_\alpha - \delta C_\beta)_{sim} - (\delta C_\alpha - \delta C_\beta)_{expt}$, for each residue in a variety of IDPs for which NMR data are available. We selected many peptide fragments from a set of prion-like IDPs such as FUS LC, TDP-43 CTD, hnRNPA2 LC, RNA Pol II C-terminal domain for the proposed comparison (SI Table S1). These LC protein sequences are mainly composed of polar and aromatic amino acids with a low occurrence of nonpolar aliphatic amino acids. The $(\delta C_\alpha - \delta C_\beta)$ data are shown in Fig. 2 as a for each residue type and each simulated peptide (Fig 2a,b) and the average over all the peptides with the two force fields (Fig 2c). We find that the residues Ser and Thr in ff99SBws and ff03ws, and Gln in ff99SBws, which constitute a significant fraction of the amino acid composition in these sequences, deviate significantly from the experimental values. The positive $(\delta C_\alpha - \delta C_\beta)$ values reflect the overpopulation of helical structures in these residues. Numerically, the average deviations for threonine are the most significant (~ +1.0 ppm) followed by serine (~ +0.5 ppm) and glutamine (~ +0.3 ppm) in ff99SBws. Also, the observed deviations in different sequences (Fig. 2a,b) may reveal sequence-dependence to some degree, but the overall trend is qualitatively consistent.

Based on the analysis above, we decided to test the effect of changes in the backbone torsion potential parameters of three amino acids (Ser, Thr, and Gln, referred hereafter as STQ group) in ff99SBws. Best and Hummer proposed the following form for the backbone dihedral angle ψ correction: $V_1(\psi; k_\psi, \delta_\psi) = k_\psi[1+\cos(\psi - \delta_\psi)]$, where k_ψ and δ_ψ are the magnitude and phase offset of V_1, respectively. For ff99SBws, $k_\psi = 2.0$ kJ/mol and δ_ψ =105.4° are used to correct the intrinsic bias toward β-sheet structures for the ff99SB force field.[26] After testing for various values, we find that a reduction in k_ψ from 2.0 kJ/mol to 1.0 kJ/mol is sufficient and appropriate to shift the bias away from helical configurations. To make a broader assessment on how changes in the potential parameters for the STQ residues affect the observed secondary structure propensities, we tested several combinations of the

modified residues (S, T, ST, TQ, STQ; Figure 3) for three peptides from our dataset, FUS LC$_{0-43}$, TDP-43$_{310-350}$, and RNA Pol II$_{1927-1970}$. We conduct PT-WTE simulations for these three peptides with variants of the ff99SBws force field, where the suffix letter(s) represent the amino acids for which the k_ψ is modified. For example, in the case of ff99SBws-ST, torsion potential is changed for Ser and Thr residues.

The (δC_α- δC_β) comparison between the ff99SBws and its new variants are shown in the SI Fig. S2 for ST, TQ, and STQ. We note that a change in only one of the residue types (S or T) k_ψ is not sufficient to observe any meaningful differences from the ff99SBws force field (Table 1). For force field variants with two or more residues k_ψ modified, we find that the deviations consistent with overly helical simulations are reduced for FUS LC$_{0-43}$ but are minimally affected for TDP-43$_{310-350}$ and RNA Pol II$_{1927-1970}$. For reference, an additional comparison between ff99SBws, ff99SBws-STQ, and ff99SB-disp force fields is presented in SI Fig. S3, showing that ff99SB-disp does not fully capture the balance of partial helical structure in these sequences. To quantify the overall differences on the whole sequence, we computed the root mean squared deviation (RMSD) between the simulation and experimental data and report it in Table 1. The RMSD values highlight the benefits of applying the proposed residue-specific changes for all the peptides, but this is especially true for the FUS LC. We also confirm that these changes do not negatively impact the observed (δC_α- δC_β) behavior of other neighboring residues.

To visualize the local structural changes, we also calculate the secondary structure propensities from the simulated ensembles using the DSSP algorithm for these different force field variants and compare it with experimental values derived from chemical shifts using δ2D method (Fig. 3). For FUS LC$_{0-43}$, the modified force fields (ST, TQ, and STQ) significantly reduce the helicity in the N-terminal region as compared to the ff99SBws. The helical fraction at all residue positions is below 20%, which we believe may be at or close to the detection limit for interpretation of the NMR data using methods based on chemical shifts such as δ2D given the possibility of many different low population states with different structures making up the conformational ensemble. For TDP-43$_{310-350}$, we do not observe a significant change in the helical population in the previously identified region (residues 321–330) among different force fields, a considerable decrease in helicity (from 40% to 20%) is observed between residues 340–345. The latter (STQ) helps improve the agreement with the experiment, which is different than what is seen in the case of ff99SB-disp that shows reduced helicity everywhere for this peptide (SI Fig. S3). RNA Pol II$_{1927-1970}$ shows relatively little changes in the observed behavior presumably since it is almost entirely disordered, which was already captured well in the ff99SBws force field.

To check if inter-residue contacts are affected by the changes in the force field parameter, we compute contact maps for FUS$_{0-43}$ and TDP-43$_{310-350}$, as shown in SI Fig. S4 and S5. As expected, a significant reduction in helical contacts (i,i+4) is observed, but otherwise, the contacts are relatively similar between the structural ensemble generated based on ff99SBws and ff99SBws-STQ. We note that the changes in the peptide dimensions due to the STQ dihedral corrections are relatively small (SI Fig. S6), and therefore, not expected to lead to significant changes in the formation of non-local intramolecular contacts.

To assess the convergence of our simulation data in terms of secondary structure propensity, we also plot residue-level secondary structure as a function of simulation time for each replica moving through the temperature space (SI Fig. S7 and S8). These data highlight many transitions between different states and the transient nature of populated structures, which ultimately contributes to the convergence of the estimated propensities reported in the paper.

We also check the performance of these new parameters with an additional test case, which is commonly employed to test the suitability of dihedral corrections, i.e., $(AAQAA)_3$ peptide. As shown in SI Fig. S9, there is a small reduction in the helical propensity of this peptide, but the overall results are still consistent with the NMR data.

We believe that the changes proposed here should improve the accuracy of the force field in general, but more work is needed to see if certain sequences containing STQ residues do not show improvement in agreement with experimental observables compared to the ff99SBws model. Nonetheless, we note that many recent studies have proposed force fields with residue-specific torsion and nonbonded parameters to improve their accuracy with a particular focus on IDPs. It is possible to incorporate some of these changes in our proposed force field further to improve the resulting IDP properties of arbitrary sequences, but significant work is needed to test the combination of parameters that will yield an optimal behavior. We plan to focus on this in our future work.

Conclusion

In conclusion, our proposed residue-type specific modifications on backbone torsion potentials of ff99SBws improved the accuracy of all-atom simulation of low-complexity IDPs (FUS_{0-43}, $TDP-43_{310-350,}$ and RNA Pol $II_{1927-1970}$) in terms of their agreement with the ^{13}C NMR chemical shifts, and hence secondary structure propensities. In this study, we adjusted the torsion potential correction parameter for three polar amino acids (S, T, and Q) to illustrate a promising path for all-atom force field optimization for IDPs using NMR spectroscopy data. Most importantly, the new force field (ff99SBws-STQ) is still transferable and does not require any experimental input while simulating a specific protein system of interest. We expect that ff99SBws-STQ will provide a more accurate description of the structural and dynamical properties of IDPs involved in the formation of biomolecular condensates via LLPS and become an indispensable model for future simulations of the many disordered proteins rich in polar residues.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

References

(1). Van Der Lee R; Buljan M; Lang B; Weatheritt RJ; Daughdrill GW; Dunker AK; Fuxreiter M; Gough J; Gsponer J; Jones DT et al. Classification of Intrinsically Disordered Regions and Proteins. Chem. Rev. 2014, 114 (13), 6589–6631. 10.1021/cr400525m. [PubMed: 24773235]

(2). Dunker AK; Obradovic Z; Romero P; Garner EC; Brown CJ Intrinsic Protein Disorder in Complete Genomes. Genome Inform. Ser. Workshop Genome Inform. 2000, 11, 161–171. 10.11234/gi1990.11.161.

(3). Tompa P Intrinsically Unstructured Proteins. Trends in Biochemical Sciences. 2002 10.1016/S0968-0004(02)02169-2.

(4). Chavali S; Gunnarsson A; Babu MM Intrinsically Disordered Proteins Adaptively Reorganize Cellular Matter During Stress. Trends in Biochemical Sciences. 2017 10.1016/j.tibs.2017.04.007.

(5). Wright PE; Dyson HJ Intrinsically Disordered Proteins in Cellular Signalling and Regulation. Nat. Rev. Mol. Cell Biol. 2015, 16 (1), 18–29. 10.1038/nrm3920. [PubMed: 25531225]

(6). Hofmann JW; Seeley WW; Huang EJ RNA Binding Proteins and the Pathogenesis of Frontotemporal Lobar Degeneration. Annual Review of Pathology: Mechanisms of Disease. 2019 10.1146/annurev-pathmechdis-012418-012955.

(7). Brangwynne CP; Eckmann CR; Courson DS; Rybarska A; Hoege C; Gharakhani J; Jülicher F; Hyman AA Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. Science (80-.). 2009, 324 (5935), 1729–1732. 10.1126/science.1172046.

(8). Shin Y; Brangwynne CP Liquid Phase Condensation in Cell Physiology and Disease. Science (80-.). 2017, 357 (6357). 10.1126/science.aaf4382.

(9). Li YR; King OD; Shorter J; Gitler AD Stress Granules as Crucibles of ALS Pathogenesis. J. Cell Biol. 2013, 201 (3), 361–372. 10.1083/jcb.201302044. [PubMed: 23629963]

(10). Molliex A; Temirov J; Lee J; Coughlin M; Kanagaraj AP; Kim HJ; Mittag T; Taylor JP Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell 2015, 163 (1), 123–133. 10.1016/j.cell.2015.09.015. [PubMed: 26406374]

(11). Mittal J; Yoo TH; Georgiou G; Truskett TM Structural Ensemble of an Intrinsically Disordered Polypeptide. J. Phys. Chem. B 2013, 117 (1), 118–124. 10.1021/jp308984e. [PubMed: 23205890]

(12). Conicella AE; Zerze GH; Mittal J; Fawzi NL ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Structure 2016, 24 (9), 1537–1549. 10.1016/j.str.2016.07.007. [PubMed: 27545621]

(13). Murray DT; Kato M; Lin Y; Thurber KR; Hung I; McKnight SL; Tycko R Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. Cell 2017, 171 (3), 615–627.e16. 10.1016/j.cell.2017.08.048. [PubMed: 28942918]

(14). Hughes MP; Sawaya MR; Boyer DR; Goldschmidt L; Rodriguez JA; Cascio D; Chong L; Gonen T; Eisenberg DS Atomic Structures of Low-Complexity Protein Segments Reveal Kinked b Sheets That Assemble Networks. Science (80-.). 2018, 359 (6376), 698–701. 10.1126/science.aan6398.

(15). Chouard T Structural Biology: Breaking the Protein Rules. Nature 2011 10.1038/471151a.

(16). Brookes DH; Head-Gordon T Experimental Inferential Structure Determination of Ensembles for Intrinsically Disordered Proteins. J. Am. Chem. Soc. 2016, 138 (13), 4530–4538. 10.1021/jacs.6b00351. [PubMed: 26967199]

(17). Löhr T; Jussupow A; Camilloni C Metadynamic Metainference: Convergence towards Force Field Independent Structural Ensembles of a Disordered Peptide. J. Chem. Phys. 2017, 146 (16). 10.1063/1.4981211.

(18). Bottaro S; Bengtsen T; Lindorff-Larsen K Integrating Molecular Simulation and Experimental Data: A Bayesian/Maximum Entropy Reweighting Approach. bioRxiv 2018, 457952 10.1101/457952.

(19). Köfinger J; Stelzl LS; Reuter K; Allande C; Reichel K; Hummer G Efficient Ensemble Refinement by Reweighting. J. Chem. Theory Comput. 2019, 15 (5), 3390–3401. 10.1021/acs.jctc.8b01231. [PubMed: 30939006]

(20). Gomes G-NW; Krzeminski M; Martin E; Mittag T; Head-Gordon T; Forman-Kay J; Gradinaru C Integrating SmFRET, SAXS and NMR Data to Infer Structural Ensembles of an Intrinsically-Disordered Protein. bioRxiv 2020, 2020.02.05.935890 10.1101/2020.02.05.935890.

(21). Lindorff-Larsen K; Trbovic N; Maragakis P; Piana S; Shaw DE Structure and Dynamics of an Unfolded Protein Examined by Molecular Dynamics Simulation. J. Am. Chem. Soc. 2012, 134 (8), 3787–3791. 10.1021/ja209931w. [PubMed: 22339051]

(22). Murthy AC; Dignon GL; Kan Y; Zerze GH; Parekh SH; Mittal J; Fawzi NL Molecular Interactions Underlying Liquid–liquid Phase Separation of the FUS Low-Complexity Domain. Nat. Struct. Mol. Biol. 2019, 26 (7), 637–648. 10.1038/s41594-019-0250-x. [PubMed: 31270472]

(23). Conicella AE; Dignon GL; Zerze GH; Schmidt HB; D'Ordine AM; Kim YC; Rohatgi R; Ayala YM; Mittal J; Fawzi NL TDP-43 α-Helical Structure Tunes Liquid-Liquid Phase Separation and Function. Proc. Natl. Acad. Sci. U. S. A. 2020, 117 (11), 5883–5894. 10.1073/pnas.1912055117. [PubMed: 32132204]

(24). Best R B; Nicolae-Viorel B; Gerhard H Are Current Molecular Dynamics Force Fields Too Helical? Biophys. J. 2008, 95 (12), 1–8. [PubMed: 18441029]

(25). Freddolino PL; Harrison CB; Liu Y; Schulten K Challenges in Protein-Folding Simulations. Nat. Phys. 2010, 6 (10), 751–758. 10.1038/nphys1713. [PubMed: 21297873]

(26). Best RB; Hummer G Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. J. Phys. Chem. B 2009, 113 (26), 9004–9015. 10.1021/jp901540t. [PubMed: 19514729]

(27). Mittal J; Best RB Tackling Force-Field Bias in Protein Folding Simulations: Folding of Villin HP35 and Pin WW Domains in Explicit Water. Biophys. J. 2010, 99 (3), L26–L28. 10.1016/j.bpj.2010.05.005. [PubMed: 20682244]

(28). Best RB; Mittal J Balance between α and β Structures in Ab Initio Protein Folding. J. Phys. Chem. B 2010, 114 (26), 8790–8798. 10.1021/jp102575b. [PubMed: 20536262]

(29). Lindorff-Larsen K; Piana S; Dror RO; Shaw DE How Fast-Folding Proteins Fold. Science (80-.). 2011, 334 (6055), 517–520. 10.1126/science.1208351.

(30). Piana S; Lindorff-Larsen K; Shaw DE How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? Biophys. J. 2011, 100 (9), L47–L49. 10.1016/j.bpj.2011.03.051. [PubMed: 21539772]

(31). Kresten L-L; Stefano P; Kim P; Paul M; John LK; Ron OD; David ES; Lindorff-Larsen K; Piana S; Palmo K et al. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. Proteins Struct. Funct. Bioinforma. 2010, 78 (8), 1950–1958. 10.1002/prot.22711.

(32). Zhou CY; Jiang F; Wu YD Residue-Specific Force Field Based on Protein Coil Library. RSFF2: Modification of AMBER ff99SB. J. Phys. Chem. B 2015, 119 (3), 1035–1047. 10.1021/jp5064676. [PubMed: 25358113]

(33). Robustelli P; Piana S; Shaw DE Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. Proc. Natl. Acad. Sci. U. S. A. 2018, 115 (21), E4758–E4766. 10.1073/pnas.1800690115. [PubMed: 29735687]

(34). Piana S; Robustelli P; Tan D; Chen S; Shaw DE Development of a Force Field for the Simulation of Single-Chain Proteins and Protein-Protein Complexes. J. Chem. Theory Comput. 2020, 16 (4), 2494–2507. 10.1021/acs.jctc.9b00251. [PubMed: 31914313]

(35). Best RB; Zhu X; Shim J; Lopes PEM; Mittal J; Feig M; MacKerell AD Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain X1and X2Dihedral Angles. J. Chem. Theory Comput. 2012, 8 (9), 3257–3273. 10.1021/ct300400x. [PubMed: 23341755]

(36). Maier JA; Martinez C; Kasavajhala K; Wickstrom L; Hauser KE; Simmerling C ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. 2015 10.1021/acs.jctc.5b00255.

(37). Best RB Computational and Theoretical Advances in Studies of Intrinsically Disordered Proteins. Current Opinion in Structural Biology. 2017 10.1016/j.sbi.2017.01.006.

(38). Levine ZA; Shea JE Simulations of Disordered Proteins and Systems with Conformational Heterogeneity. Curr. Opin. Struct. Biol. 2017, 43, 95–103. 10.1016/j.sbi.2016.11.006. [PubMed: 27988422]

(39). Huang J; MacKerell AD Force Field Development and Simulations of Intrinsically Disordered Proteins. Curr. Opin. Struct. Biol. 2018, 48, 40–48. 10.1016/j.sbi.2017.10.008. [PubMed: 29080468]

(40). Das P; Matysiak S; Mittal J Looking at the Disordered Proteins through the Computational Microscope. ACS Cent. Sci. 2018, 4 (5), 534–542. 10.1021/acscentsci.7b00626. [PubMed: 29805999]

(41). Rauscher S; Gapsys V; Gajda MJ; Zweckstetter M; De Groot BL; 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. 10.1021/acs.jctc.5b00736. [PubMed: 26574339]

(42). Best RB; Mittal J Microscopic Events in β-hairpin Folding from Alternative Unfolded Ensembles. Proc. Natl. Acad. Sci. 2011, 108 (27), 11087–11092. 10.1073/pnas.1016685108. [PubMed: 21690352]

(43). Abascal JLF; Vega C A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. J. Chem. Phys. 2005, 123 (23). 10.1063/1.2121687.

(44). Best RB; Mittal J Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. J. Phys. Chem. B 2010, 114 (46), 14916–14923. 10.1021/jp108618d. [PubMed: 21038907]

(45). Nerenberg P; Head-Gordon T Optimizing Protein-Solvent Force Fields To Reproduce Conformational Preferences of Model Peptides. J. Chem. Theo. Comp. 2011, 7, 1220–1230. 10.1021/ct2000183.

(46). Best RB; Zheng WW; 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. 10.1021/ct500569b. [PubMed: 25400522]

(47). Conicella AE; Dignon GL; Zerze GH; Schmidt HB; D'Ordine AM; Kim YC; Rohatgi R; Ayala YM; Mittal J; Fawzi NL TDP-43 α-Helical Structure Tunes Liquid–Liquid Phase Separation and Function. Proc. Natl. Acad. Sci. U. S. A. 2020, 117 (11), 5883–5894. 10.1073/pnas.1912055117. [PubMed: 32132204]

(48). Ryan VH; Dignon GL; Zerze GH; Chabata CV; Silva R; Conicella AE; Amaya J; Burke KA; Mittal J; Fawzi NL Mechanistic View of HnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. Mol. Cell 2018, 69 (3), 465–479.e7. 10.1016/j.molcel.2017.12.022. [PubMed: 29358076]

(49). Zerze GH; Best RB; Mittal J Sequence- and Temperature-Dependent Properties of Unfolded and Disordered Proteins from Atomistic Simulations. J. Phys. Chem. B 2015, 119 (46), 14622–14630. 10.1021/acs.jpcb.5b08619. [PubMed: 26498157]

(50). Monahan Z; Ryan VH; Janke AM; Burke KA; Rhoads SN; Zerze GH; O'Meally R; Dignon GL; Conicella AE; Zheng W et al. Phosphorylation of the FUS Low-complexity Domain Disrupts Phase Separation, Aggregation, and Toxicity. EMBO J. 2017, 36 (20), 2951–2967. 10.15252/embj.201696394. [PubMed: 28790177]

(51). Mier P; Paladin L; Tamana S; Petrosian S; Hajdu-Soltész B; Urbanek A; Gruca A; Plewczynski D; Grynberg M; Bernadó P et al. Disentangling the Complexity of Low Complexity Proteins. Briefings in Bioinformatics. 2020 10.1093/bib/bbz007.

(52). Alberti S; Halfmann R; King O; Kapila A; Lindquist S A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. Cell 2009 10.1016/j.cell.2009.02.044.

(53). Lindahl E; Hess B; Kutzner C; van der Spoel D Gromacs 4.0: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory Comput. 2008, 4, 435–447. [PubMed: 26620784]

(54). Tribello GA; Bonomi M; Branduardi D; Camilloni C; Bussi G PLUMED 2: New Feathers for an Old Bird. Comput. Phys. Commun. 2014, 185 (2), 604–613. 10.1016/j.cpc.2013.09.018.

(55). Sugita Y; Y. O; Sugita Y; Okamoto Y Replica-Exchange Molecular Dynamics Method for Protein Folding. Chem. Phys. Lett. 1999, 314 (November), 141.

(56). Bonomi M; Parrinello M Enhanced Sampling in the Well-Tempered Ensemble. Phys. Rev. Lett. 2010, 104 (19), 3–6. 10.1103/PhysRevLett.104.190601.

(57). Deighan M; Bonomi M; Pfaendtner J Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble. J. Chem. Theory Comput. 2012, 8 (7), 2189–2192. 10.1021/ct300297t. [PubMed: 26588950]

(58). Laio A; Parrinello M Escaping Free-Energy Minima. Proc. Natl. Acad. Sci. U. S. A. 2002, 99 (20), 12562. [PubMed: 12271136]

(59). Essmann U; Perera L; Berkowitz ML; Darden T; Lee H; Pedersen LG A Smooth Particle Mesh Ewald Method. J. Chem. Phys. 1995, 103 (19), 8577–8593. 10.1063/1.470117.

(60). 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 (1), 13–22. 10.1007/s10858-010-9433-9. [PubMed: 20628786]

(61). Kabsch W; Sander C Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-bonded and Geometrical Features. Biopolymers 1983, 22 (12), 2577–2637. 10.1002/bip.360221211. [PubMed: 6667333]

(62). Camilloni C; Simone, De A; Vranken WF; Vendruscolo M Determination of Secondary Structure Populations in Disordered States of Proteins Using NMR Chemical Shifts Determination of Secondary Structure Populations in Disordered States of Proteins Using NMR Chemical Shifts. Biochemistry 2012, 51 (11), 2224–2231. 10.1021/bi3001825. [PubMed: 22360139]

(63). Janke AM; Seo DH; Rahmanian V; Conicella AE; Mathews KL; Burke KA; Mittal J; Fawzi NL Lysines in the RNA Polymerase II C-Terminal Domain Contribute to TAF15 Fibril Recruitment. Biochemistry 2018, 57 (17), 2549–2563. 10.1021/acs.biochem.7b00310. [PubMed: 28945358]

(64). Kjaergaard M; Brander S; Poulsen FM Random Coil Chemical Shifts for Intrinsically Disordered Proteins: Effects of Temperature and pH. J. Biomol. NMR 2011, 49, 139–149. [PubMed: 21234644]

(65). Burke KA; Janke AM; Rhine CL; Fawzi NL Residue-by-Residue View of In Vitro FUS Granules That Bind the C-Terminal Domain of RNA Polymerase II. Mol. Cell 2015, 60 (2), 231–241. 10.1016/j.molcel.2015.09.006. [PubMed: 26455390]

(66). Best RB; De Sancho D; Mittal J Residue-Specific α-Helix Propensities from Molecular Simulation. Biophys. J. 2012, 102 (6), 1462–1467. 10.1016/j.bpj.2012.02.024. [PubMed: 22455930]
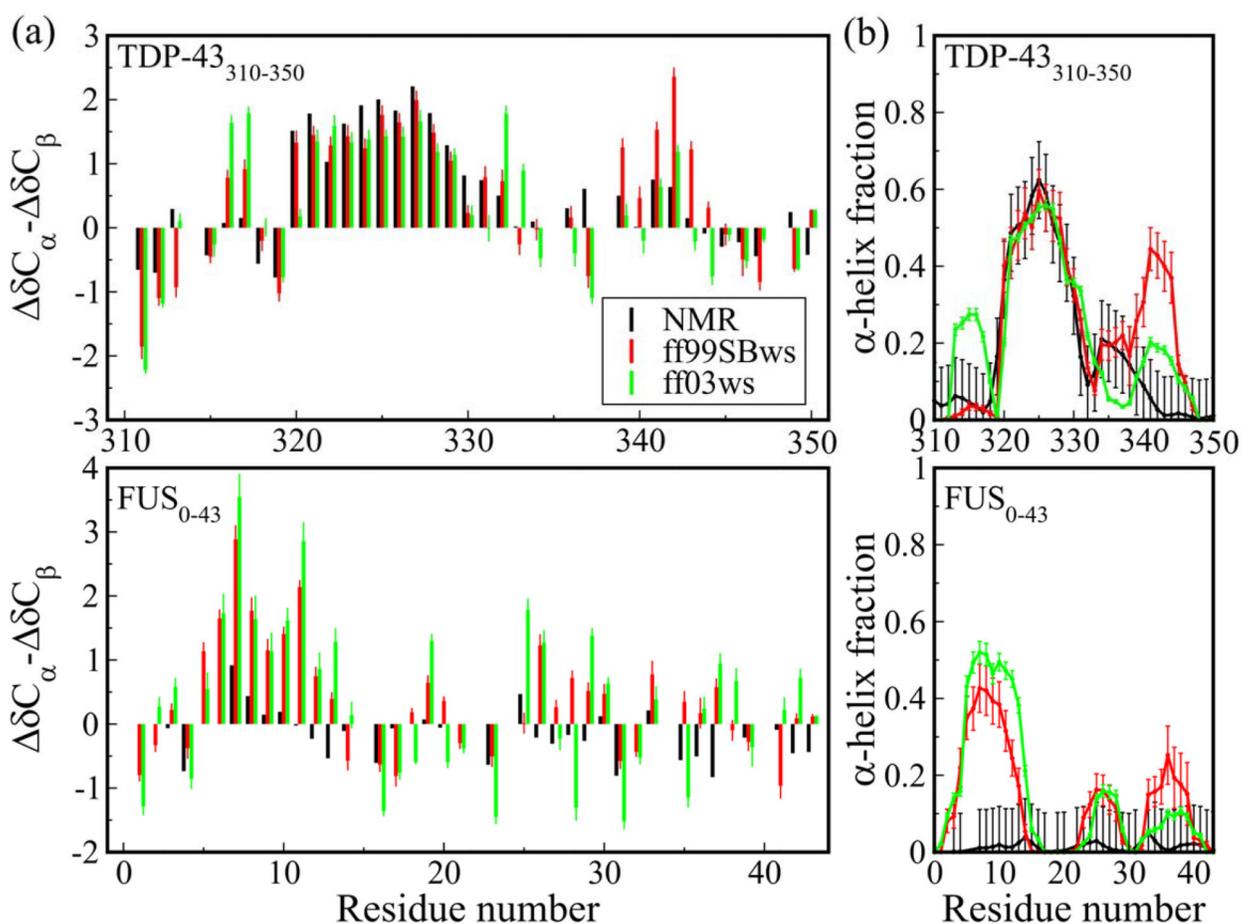
Figure 1.
(*a*) Chemical shifts $\Delta\delta C_\alpha - \Delta\delta C_\beta$ and (*b*) α-helix propensities of TDP-43$_{310-350}$ and FUS$_{0-43}$ using all-atom force fields ff99SBws and ff03ws, compared to experimental results. Chemical shifts are calculated from protein structure using SPARTA+ algorithm. Secondary structures are assigned by DSSP. Experimental values of secondary structure propensities are calculated from NMR chemical shifts using the δ2D program.
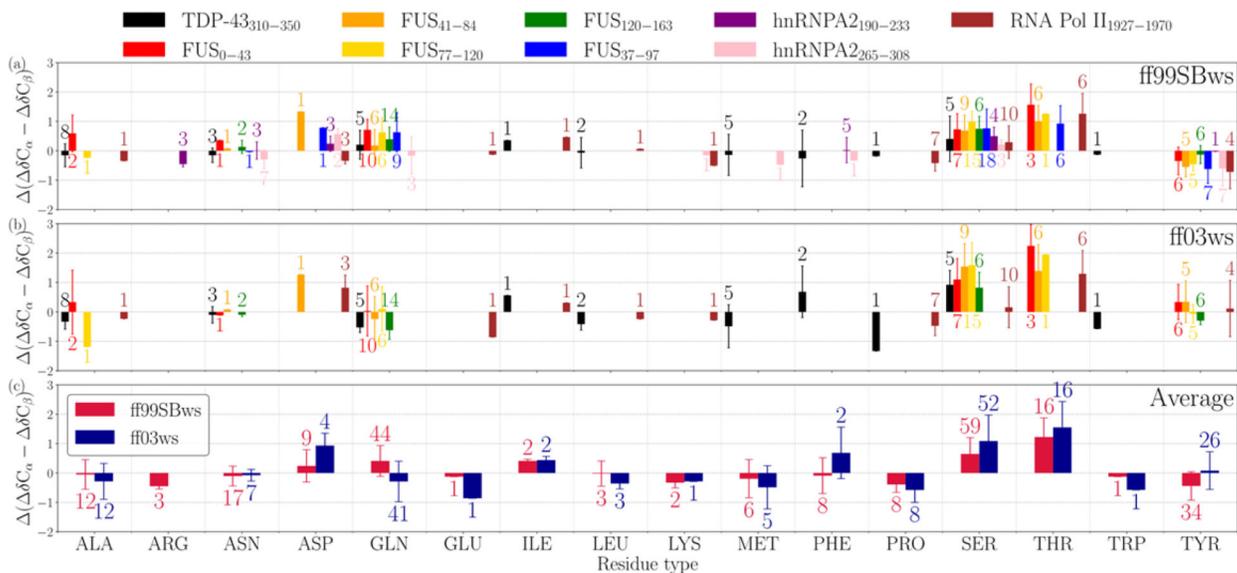
Figure 2.

Deviation of chemical shifts $(\delta C_\alpha - \delta C_\beta)$ (in ppm) in simulations from NMR ^{13}C chemical shifts. **(a)** and **(b)** shows $(\delta C_\alpha - \delta C_\beta)$ of each peptide from ff99SBws and ff03ws simulations, respectively. **(c)** shows the average of all peptides in the same force field. Error bars are the standard deviations among the sampled residues. Annotated numbers next to the bar are the number of the residues of that type present in the chosen peptide(s).
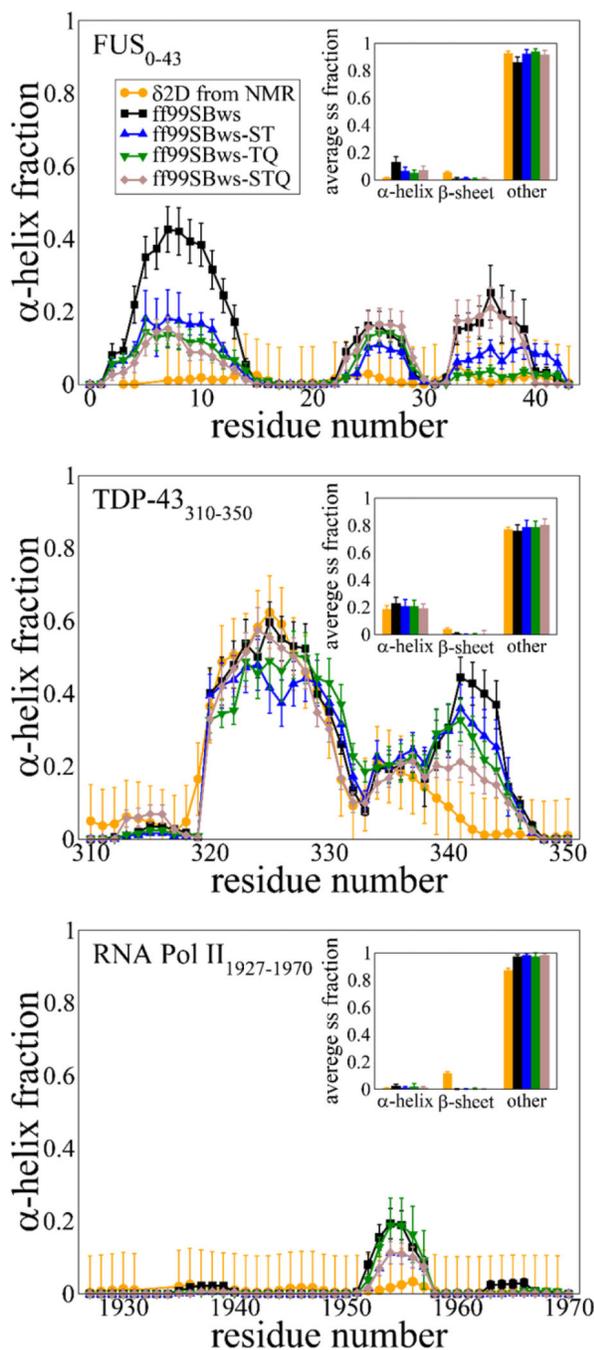
Figure 3.
Secondary structure propensities of (a) FUS$_{0-43}$, (b) TDP-43$_{310-350}$, and (c) RNA Pol II$_{1927-1970}$. Error bars for DSSP values are deviations between 10-ns segments in the simulation trajectories. Error bars for δ2d population per residue are 0.1 (10%).

Table 1:

Root-mean-squared deviation (RMSD) of simulated $\delta C_{\alpha} - \delta C_{\beta}$ (in ppm) from NMR ^{13}C chemical shifts of three IDPs for ff99SBws, modified ff99SBws variants (S,T,ST,TQ,STQ) (this study), and ff99SB-disp.

	FUS$_{0-43}$	TDP-43$_{310-350}$	RNA Pol II$_{1927-1970}$
ff99SBws	0.90 ± 0.11	0.64 ± 0.09	0.68 ± 0.11
ff99SB-disp	0.90 ± 0.11	0.78 ± 0.07	0.74 ± 0.11
ff99SBws-S	0.86 ± 0.10		
ff99SBws-T	0.98 ± 0.13		
ff99SBws-ST	0.64 ± 0.07	0.71 ± 0.08	0.74 ± 0.08
ff99SBws-TQ	0.62 ± 0.08	0.66 ± 0.06	0.71 ± 0.10
ff99SBws-STQ	0.54 ± 0.07	0.51 ± 0.04	0.74 ± 0.08