*Chem Biol Drug Des 2014*; **84:** 253–269

Research Article

# New Force Field on Modeling Intrinsically Disordered Proteins

**Wei Wang[1,†], Wei Ye[1,†], Cheng Jiang[1], Ray Luo[2,*] and Hai-Feng Chen[1,3,*]**

[1]*State Key Laboratory of Microbial metabolism, Department of Bioinformatics and Biostatistics, College of Life Sciences and Biotechnology, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai 200240, China*
[2]*Departments of Molecular Biology and Biochemistry and Biomedical Engineering, University of California, Irvine, CA 92697-3900, USA*
[3]*Shanghai Center for Bioinformation Technology, 1275 Keyuan Road, Shanghai 200235, China*
**Corresponding author: Ray Luo, ray.luo@uci.edu* and Hai-Feng Chen, *haifengchen@sjtu.edu.cn*
[†]*These authors contributed equally to this work.*

**Intrinsically disordered proteins or intrinsically disordered protein regions comprise a large portion of eukaryotic proteomes (between 35% and 51%). These intrinsically disordered proteins were found to link with cancer and various other diseases. However, widely used additive force field parameter sets are insufficient in quantifying the structural properties of intrinsically disordered proteins. Therefore, we explored to a systematic correction of a base additive force field parameter set (chosen as Amber *ff99SBildn*) to correct the biases that was first demonstrated in simulations with the base parameter set. Specifically, the $/\psi$ distributions of disorder-promoting residues were systematically corrected with the CMAP method. Our simulations show that the CMAP corrected Amber parameter set, termed *ff99IDPs*, improves the $/\psi$ distributions of the disorder-promoting residues with respect to the benchmark data of intrinsically disordered protein structures, with root mean-squared percentage deviation less than 0.15% between the simulation and the benchmark. Our further validation shows that the chemical shifts from the *ff99IDPs* simulations are in quantitative agreement with those from reported NMR measurements for two tested IDPs, MeV $N_{TAIL}$, and *p53*. The predicted residue dipolar couplings also show high correlation with experimental data. Interestingly, our simulations show that *ff99IDPs* can still be used to model the ordered state when the intrinsically disordered proteins are in complex, in contrast to *ff99SBildn* that can be applied well only to the ordered complex structures. These findings confirm that the newly proposed Amber *ff99IDPs* parameter set provides a reasonable tool in further studies of intrinsically disordered protein structures. In addition, our study also shows the importance of considering intrinsically disordered protein structures in general-purposed force field developments for both additive and non-additive models.**

**Key words:** CMAP, *ff99IDPs*, IDPs force field, MD simulation, $\varphi/\psi$ dihedrals

Received 28 January 2014, revised 15 February 2014 and accepted for publication 24 February 2014

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) in structured proteins have been recently researched (1) to exist widely in eukaryotic proteomes: Up to 35–51% of proteins were predicted with more than 40 consecutive disordered residues (2,3). It is thus very natural to extend the studies of protein functions carried out by folded and mostly rigid structures to those by intrinsic disordered structures (4–6).

Indeed, IDPs have been proven to be very important in many cellular processes, such as molecular recognition, molecular assembly, and protein post-translational modification (3,6–9). It has been reported that 79% of cancer-associated and 66% of cell-signaling proteins contain IDRs with 30 or more residues (10). For example, tumor suppressor *p53* acts as a connected hub in multiple signaling networks whose N-terminal and C-terminal domains are IDRs (10–13). Furthermore, many diseases are related to these IDPs or IDRs (7,14), such as Parkinson's disease (15,16), Alzheimer's disease (17–19), cancer (10), cardiovascular disease (20), amyloidosis (21), diabetes (7), and others. All these observations show the importance of exploring IDPs and their functional mechanisms.

To obtain better insight into IDPs functions, one important and direct way is to study the relationship between the disordered structures and functions. As a robust computational tool, molecular dynamics (MD) simulation has been widely used for these studies. Nevertheless, the MD simulation community still faces two major limitations: (i) current computation power limits the timescale of MD simulations up to microsecond timescales, which is still far shorter than many important biological timescales (*sampling problem*); (ii) current accuracy in the potential energy functions

 doi: 10.1111/cbdd.12314

may lead to biased conformations and energetics (*force field problem*) (22,23).

For large-scale biological macromolecules, molecular mechanics (MM) empirical functions (force field) are broadly utilized, such as assisted model building with energy refinement (AMBER), Chemistry at HARvard Macromolecular Mechanics (CHARMM), GROningen MOlecular Simulation (GROMOS), and optimized potential for liquid simulations (OPLS) (24–28). All these existing potential energy functions are empirically parameterized, and their test sets are all based on crystal structure databases of folded proteins, such as Protein Data Bank (PDB; 29) and Cambridge Crystal Data Bank (CCDB; 30). Unfortunately, the samples of IDPs in PDB or CCBD are often much lower. This may cause biases that the conformations of IDPs in simulations are over-stabilized. In contrast to fixed charge additive force field, general-purposed polarizable force fields could partly overcome higher energy barrels and sample more conformations between order and disorder structures (31–33). However, the development of polarizable force field has yet reached the stage of routine applications for typical proteins, in part due to the extra computational cost (32,34).

In this study, we explored the direction of a special purpose force field for IDPs based on specific CMAP corrections of key disorder-promoting residues from a standard fixed charge additive force field. The strategy leads to little overhead in the force field complexity and computational overhead in MD simulations of typical IDPs. Of all the function terms that constitute a MM potential energy function, backbone torsion (dihedral) energy is apparently the most significant and directly determines the secondary structure. For example, the force fields of *ff94* and *ff99* over-stabilize α-helical conformations (35). To eliminate this bias, a new set of backbone torsion parameters was successfully introduced with *ff99SB* (36,37).

Specifically, we analyzed the crystal structures in PDB and sampled the backbone dihedrals ($\varphi/\psi$) distributions in both ordered and disordered regions. Our data show that the $\varphi/\psi$ distributions of residues in disordered fragments are significantly different from those in ordered fragments. To introduce corrections in a minimal perturbation strategy, we only corrected the main chain dihedral terms for the eight disorder-promoting residues, A, G, P, R, Q, S, E, and K as reported in the literature (1,6,38,39). Residue-specific grid-based energy correction maps (CMAP; 40,41) were employed to achieve the best possible agreement with database backbone distributions.

Two IDPs with NMR experimental chemical shifts or residual dipolar couplings (RDCs) values were used as representative structures for validating our new IDPs force field (termed *ff99IDPs*): C-terminal domain of measles virus nucleoprotein (MeV $N_{TAIL}$) in both free and bound states (42–47); N-terminal domain of *p53* in the free state (48,49).

## Methods

### Data collection of disordered protein structures

Of 16 548 structures with resolution value $R < 0.25$ were downloaded from PDB. DSSP (50,51) was used to assign the secondary structures and calculate dihedral angles. Five or more consecutive residues classified as 'no secondary structure' by DSSP were used as models of IDPs. Actually, we tested different numbers of residues for IDPs, from 5 to 13 coil residues. We found the $\varphi/\psi$ distributions of four principal regions without significant differences for different numbers of residues (from 5 to 13). The results are shown in Figure S1. To collect enough samples, five or more disordered regions were chosen as the standard of IDPs. Furthermore, its soundness is validated below by independent comparison of simulation and experiment of real IDPs. In summary, 42 774 disordered fragments containing 267 751 pairs of backbone dihedrals were extracted and analyzed. These subset data were used as a Ramachandran plot for IDPs or IDRs. All the disordered dihedrals were taken as parameterization benchmark.

### CMAP method

MM potential energy is calculated in Amber force fields as

$$E_{MM} = E_{bond} + E_{angle} + E_{non-bond} + E_{dihedral} \quad (1)$$

Even with ever improving performance in generations of force fields, distinct difference between MD generated and crystal structure observed $\varphi/\psi$ distributions still exists. In this study, we utilized the method of grid-based energy correction maps (termed CMAP) (40,41,52) to minimize the difference in the dihedral angle distributions. The method is currently applied in the CHARMM simulation package and can also be compatibly deployed in the latest Amber package (53,54). Briefly, an additional energy term of $E_{CMAP}$ was added to the original Amber potential energy function to correct the dihedral energy as eqn (2).

$$E_{MM} = E_{bond} + E_{angle} + E_{non-bond} + E_{dihedral} + E_{CMAP} \quad (2)$$

All other energy terms except the dihedral energy term remain the same as a chosen base additive force field. Furthermore, only the backbone dihedral parameters for the 8 disorder-promoting residues were optimized while the parameters for the rest of the 12 residues remain the same to minimize the perturbation to folded structure distributions.

In this initial study, the parameterization was conducted with di-residue models only (Nme-X-Ace, X means a certain amino acid), which contain $\varphi$ and $\psi$ angles for the central residues X. Similar model compounds were also

used in the Amber *ff94* and *ff99SBildn* parameter developments (37,55). Of course, further optimization based on short peptides and full proteins would further improve the quality of the force field, but the limitations would be the availability of benchmark data.

CMAP is a matrix of corrections on dihedral grids with the corrections between grid points calculated with a two-dimensional bicubic interpolation method (41). The correction matrix for each residue was set up with a dihedral angle grid in the resolution of 15 degrees. Specifically, we used relative conformational free energies ($\Delta G_{i,j}$) converted from $\varphi/\psi$ distributions from the disordered protein structures to compute the correction matrix

$$\Delta G_{i,j} = -\mathrm{RT}\ \ln\ (N_{i,j}/N_0) \tag{3}$$

where $N_{i,j}$ is the population of $\varphi/\psi$ dihedral bin ($i$, $j$) and $N_0$ is the population of the most-populated bin.

Using eqn 3, sparsely populated bins could have huge relative free energies, leading to over-correction. To overcome this limitation, we used an iterative optimization process to determine the CMAP correction matrix self-consistently. Here, CMAP energy terms were calculated at each iteration step with eqn 4

$$E_{i,j}^{\mathrm{CMAP}} = \Delta G_{i,j}^{\mathrm{DB}} - \Delta G_{i,j}^{\mathrm{MD}} \tag{4}$$

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

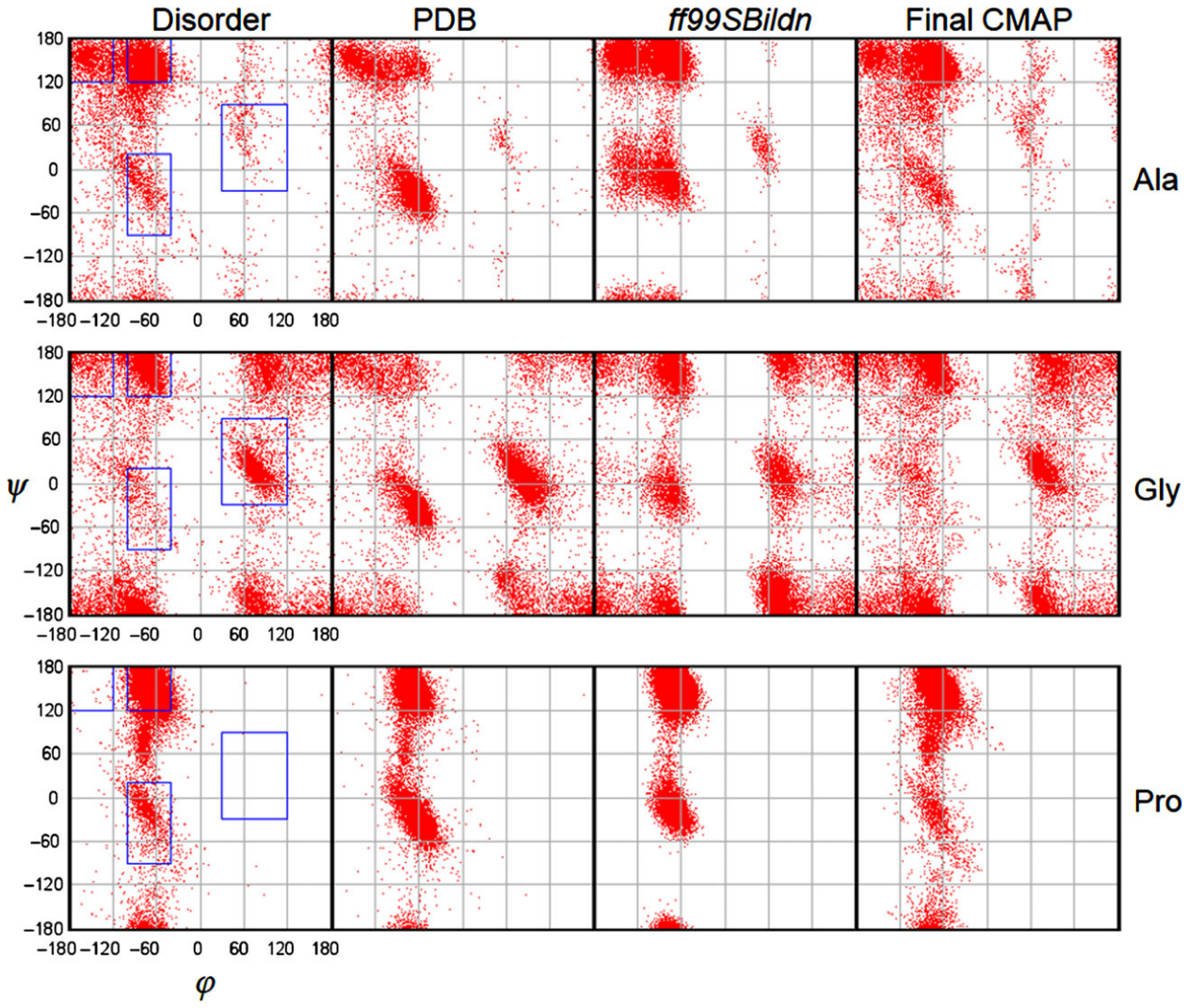


**Figure 1:** Dihedral distribution of disordered benchmark data, PDB structures, *ff99SBildn*, and CMAP optimization for Ala, Gly, and Pro.

and database populations. An optimization was conducted for every disorder-promoting residue to ensure the convergence of RMS*p*, that is, less than 0.15%. In general, up to five iteration steps were needed for all eight disorder-promoting residues.

Once the convergence was achieved, a further two-sample Kolmogorov–Smirnov (KS) test was employed to assess whether the optimized MD population is significantly different from the database population. Note that the KS test, as a nonparametric test, is a good choice for this study (56–58) because the distribution of $\varphi$ and $\psi$ dihedrals in the dihedral angle space (shown in Figure 1) apparently does not satisfy any type of probability distribution (normal or uniform distribution) for a parametric test.

### Interfacing with existing Amber force field tools

The CMAP parameters of the 8 disordered residues were collected in a predefined CMAP parameter file, and the CMAP parameters of the remaining 12 residues were set to zero, so that the base *ff99SBildn* force field was used for these residues in all simulations. Starting topology and co-ordinate files were first generated with the LEaP module with the base force field. The addition of the CMAP term to the topology file was realized with a PERL script. Then, standard procedure was performed with MD simulation.

### Molecular dynamics simulations

All the MD simulations in CMAP optimizations and subsequent validations were conducted with AMBER12. Di-residue models of the eight disorder-promoting amino acids were generated in the LEaP module. Counter-ions were used to maintain system neutrality. All systems were solvated in a truncated octahedron box of TIP3P waters with a buffer of 10 Å. Particle mesh Ewald (PME; 59) was employed to treat long-range electrostatic interactions with the default setting in AMBER12. All the MD simulations were accelerated with the CUDA version of PMEMD (60,61) in GPU cores of NVIDIA® Tesla K10 (Santa Clara, CA, USA). The SHAKE algorithm (62) was used to constrain bonds involving hydrogen atoms. A total of 20 000-step steepest descent minimization was performed to relieve any structural clash in the solvated systems. Then, 20 ps was used to heat up and 10 ps was briefly equilibrated in the NVT ensemble at 298K with PMEMD. Langevin dynamics with a time step of 2 fs were used in the heating and equilibration runs with a friction constant of 1 $ps^{-1}$.

At each iteration step, at least 100-ns MD simulation was performed to collect samples. To reduce correlation among sampling points, data were collected every 5 ps. Because dihedral samples of proline, arginine, and lysine are more than 10 000 (proline has the most samples ~20 000), longer simulations on these residues were performed, which indicate that 100-ns simulations were sufficient for the CMAP optimizations.

In the subsequent validation simulations with the free and bound MeV $N_{TAIL}$ and free *p53*, five trajectories with 100 ns each were collected. As there is no structure available for free IDPs, the initial protein structures were obtained from PDB in bound state: the bound MeV $N_{TAIL}$ structure [residues 484–504, bound with XD of phosphoprotein, 1T6O (63)] and bound *p53* [residues 17–29, bound with MDM2, 1YCR (64)]. Furthermore, these IDPs have experimental chemical shifts or RDCs, and it will be helpful to compare the simulation and experimental observation.

**Table 1:** Dihedral distribution of the four principal regions for IDPs benchmark data and PDB database

| | Population (%) | | |
|---|---|---|---|
| Regions | Disorder | PDB | Difference |
| A | | | |
| $\alpha_R$ | 5.72 (8.03) | 71.77 (79.6) | −66.05 (−71.58) |
| $\alpha_L$ | 1.90 (2.67) | 0.96 (1.06) | 0.94 (1.61) |
| $PP_{II}$ | 52.45 (73.58) | 6.65 (7.37) | 45.80 (66.21) |
| B | 11.21 (15.73) | 10.79 (11.96) | 0.42 (3.76) |
| E | | | |
| $\alpha_R$ | 4.35 (6.91) | 70.7 (81.13) | −66.35 (−74.21) |
| $\alpha_L$ | 2.48 (3.94) | 1.63 (1.87) | 0.85 (2.08) |
| $PP_{II}$ | 45.0 (71.54) | 6.89 (7.91) | 38.10 (63.62) |
| B | 11.07 (17.61) | 7.93 (9.09) | 3.15 (8.51) |
| G | | | |
| $\alpha_R$ | 2.85 (6.6) | 29.33 (42.07) | −26.48 (−35.46) |
| $\alpha_L$ | 16.41 (37.98) | 30.12 (43.2) | −13.71 (−5.22) |
| $PP_{II}$ | 19.56 (45.28) | 4.17 (5.98) | 15.39 (39.29) |
| B | 4.38 (10.14) | 6.10 (8.74) | −1.72 (1.39) |
| K | | | |
| $\alpha_R$ | 5.63 (9.56) | 63.95 (76.06) | −58.33 (−66.5) |
| $\alpha_L$ | 2.29 (3.9) | 2.69 (3.2) | −0.40 (0.7) |
| $PP_{II}$ | 38.34 (65.16) | 8.30 (9.87) | 30.05 (55.29) |
| B | 12.59 (21.39) | 9.14 (10.88) | 3.44 (10.51) |
| P | | | |
| $\alpha_R$ | 4.47 (5.1) | 55.06 (59.24) | −50.59 (−54.14) |
| $\alpha_L$ | 0.01 (0.01) | 0.01 (0.01) | 0 (0) |
| $PP_{II}$ | 83.02 (94.79) | 37.85 (40.73) | 45.17 (54.06) |
| B | 0.09 (0.11) | 0.03 (0.03) | 0.07 (0.08) |
| Q | | | |
| $\alpha_R$ | 3.80 (6.44) | 67.16 (78.46) | −63.35 (−72.02) |
| $\alpha_L$ | 3.22 (5.46) | 2.17 (2.53) | 1.06 (2.93) |
| $PP_{II}$ | 34.26 (58) | 6.62 (7.73) | 27.64 (50.27) |
| B | 17.78 (30.1) | 9.65 (11.28) | 8.13 (18.83) |
| R | | | |
| $\alpha_R$ | 5.65 (9.28) | 62.68 (74.57) | −57.04 (−65.29) |
| $\alpha_L$ | 2.61 (4.29) | 1.94 (2.31) | 0.67 (1.98) |
| $PP_{II}$ | 35.81 (58.84) | 7.97 (9.48) | 27.84 (49.35) |
| B | 16.79 (27.6) | 11.47 (13.64) | 5.33 (13.96) |
| S | | | |
| $\alpha_R$ | 5.30 (7.95) | 56.18 (67.17) | −50.88 (−59.22) |
| $\alpha_L$ | 1.36 (2.04) | 1.58 (1.89) | −0.22 (0.15) |
| $PP_{II}$ | 44.31 (66.47) | 9.34 (11.17) | 34.97 (55.3) |
| B | 15.69 (23.54) | 16.54 (19.77) | −0.84 (3.77) |

Numbers in parentheses are relative fractions within the four principal regions.

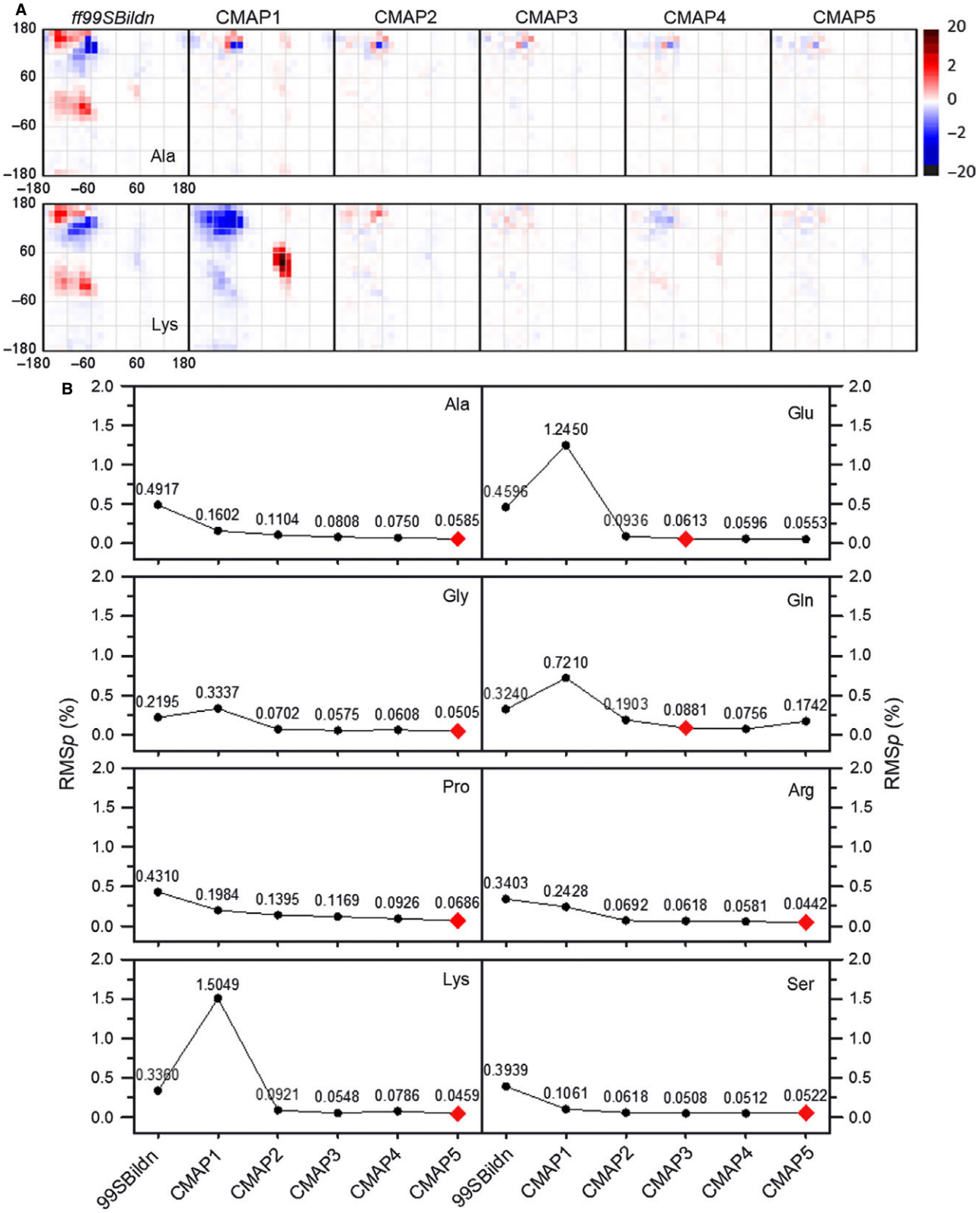


**Figure 2:** CMAP optimization. (A) Difference between disordered survey data and CMAP optimization. (B) RMSp over five steps of optimization.

### Data analyses

Root mean-squared deviations (RMS*d*), fluctuations (RMS*f*), and dihedral angles in MD trajectories were calculated with the PTRAJ module in AMBER12 and AmberTools13 (54). Structural clustering was conducted with the kclust program in the MMTSB tool set (65). Secondary structures of each snapshot were identified with DSSP (50,51). Experimental chemical shift data of free and bound MeV $N_{TAIL}$ and free *p53* were retrieved from the Biological Magnetic Resonance Data Bank (66) with Accession Numbers 6566, 6567, and 17760, respectively. Computational secondary chemical shift data for the simulated structures were calculated with SPARTA version 1.01 (67). Helicity was calculated as the helical (predicted by DSSP as α-helix, 3/10 helix, and π-helix) population during the production runs. Prediction of $NH^N$ RDC was performed with flexible-meccano (FM) software (68).

**Table 2:** Simulation conditions for all systems

| System | Force Field | Counter-ions | Waters | Trajectories | Simulation time (ns) |
|---|---|---|---|---|---|
| Free MeV $N_{TAIL}$ | *ff99IDPs* *ff99SBildn* | 1 $Cl^-$ | 2890 | 10 | 1000 |
| Free *p53* | *ff99IDPs* *ff99SBildn* | 2 $Na^+$ | 2390 | 10 | 1000 |
| Bound MeV $N_{TAIL}$ | *ff99IDPs* *ff99SBildn* | 5 $Cl^-$ | 5097 | 2 | 200 |
| Free MeV $N_{TAIL}$ (PB) | *ff99IDPs* *ff99SBildn* | 1 $Cl^-$ | 0 | 2 | 200 |
| Free *p53* (PB) | *ff99IDPs* *ff99SBildn* | 2 $Na^+$ | 0 | 2 | 200 |

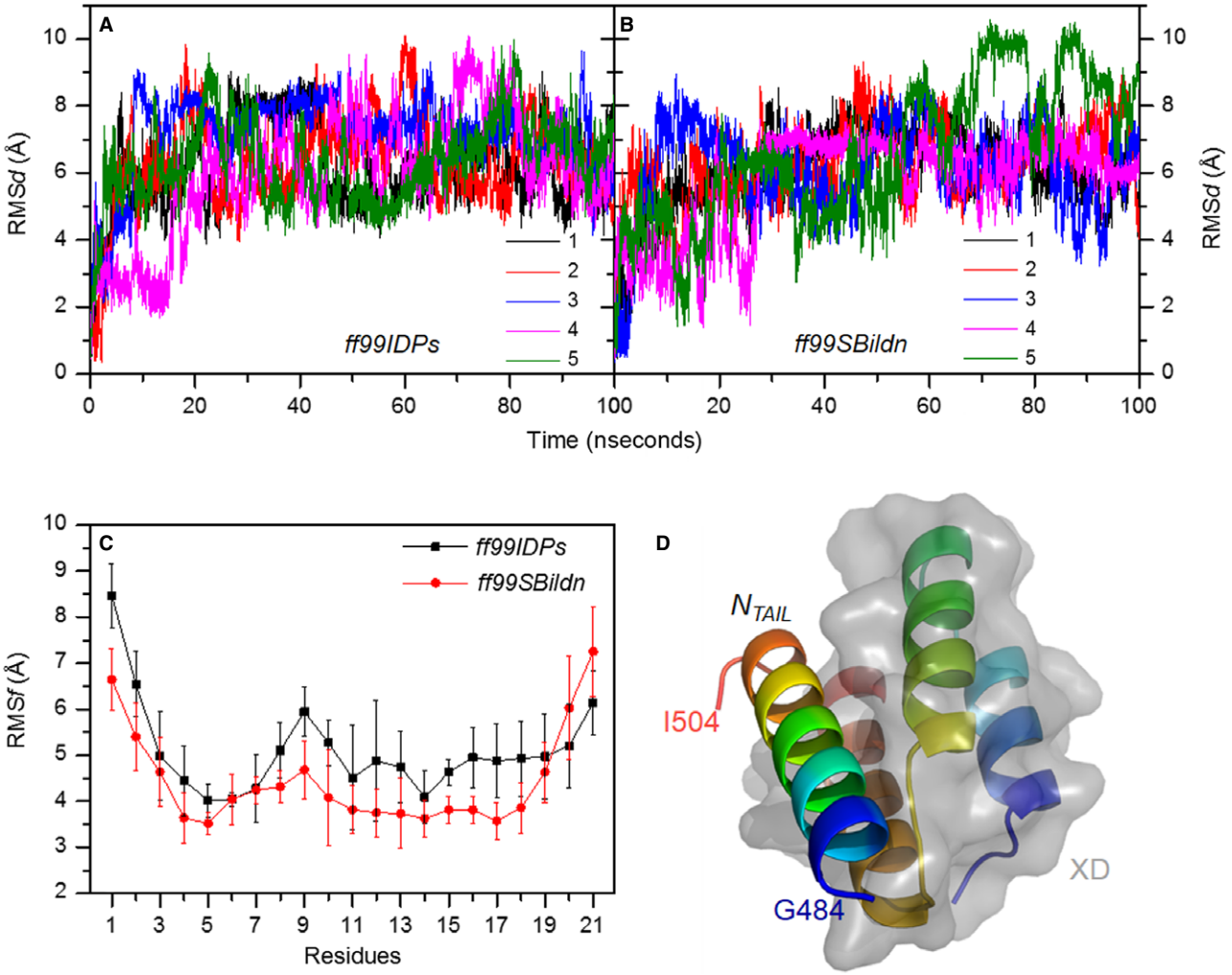


**Figure 3:** RMS*d* and RMS*f* for bound and free MeV $N_{TAIL}$ under different force fields. (A) Cα RMSd of free state with *ff99IDPs*. (B) Cα RMSd of free state with *ff99SBildn*. (C) RMSf for free state with *ff99IDPs* and *ff99SBildn*. (D) Cartoon representation of crystal structure for bound state MeV $N_{TAIL}$ with XD of P protein.

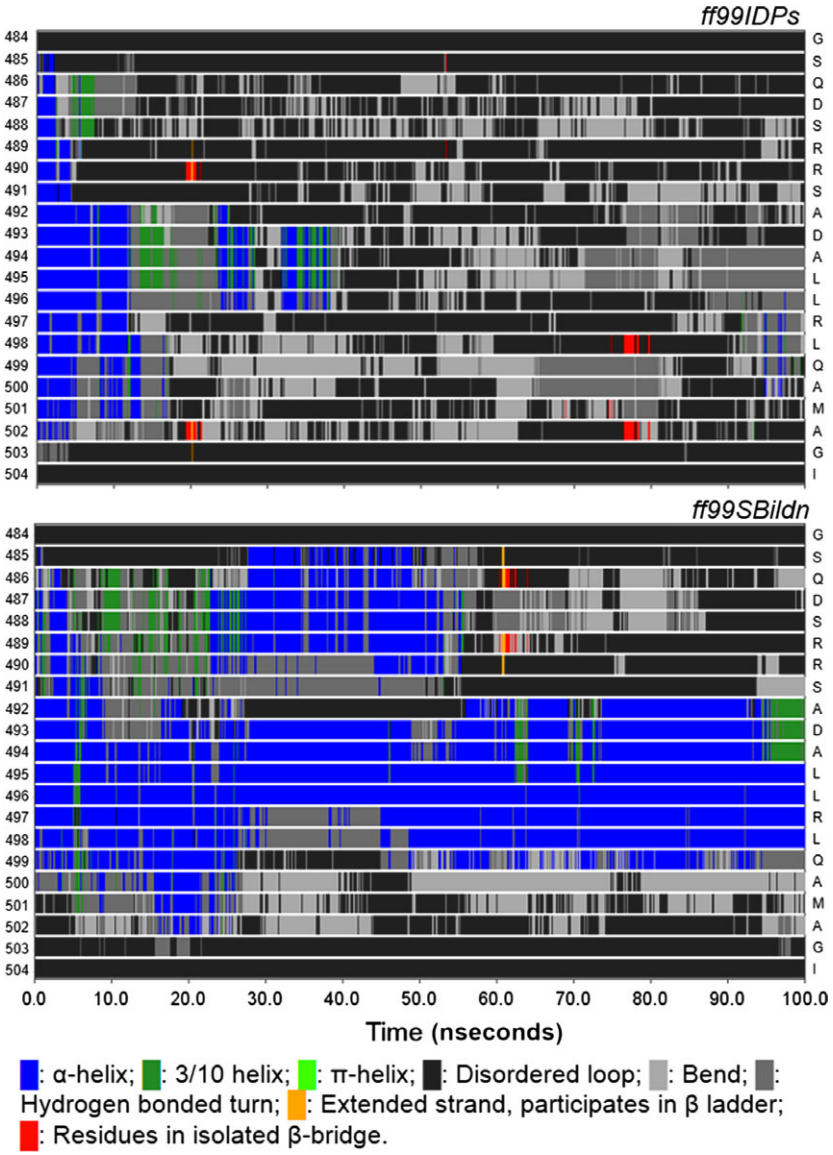


**Figure 4:** Secondary structures of free MeV $N_{TAIL}$ versus simulation time under different force fields.

## Results and Discussion

### *Database statistics*

To justify our development for the IDPs force field, the dihedral angle distributions of eight disorder-promoting residues (A, E, G, K, P, Q, R, and S) were first analyzed. Table 1 shows the statistics of dihedral angle distributions for the loops with five or more residues, all PDB crystal structures and their differences in four principal regions (shown in Figure 1; 37). The data show that the dihedral angle distributions of disordered structures are significantly different from those of PDB structures. At the same time, the published validation studies of *ff99SBildn* show that the force field reproduces well experimental $S^2$ parameters on alpha-helices and beta-sheets, while the main errors are on the loop regions (37). This is consistent with the observations that currently widely used force fields work well for folded proteins but is limited in representing IDPs because their biases toward folded protein structures. To further confirm the property of widely used force field, the dihedral distributions for residues A, G, and P were plotted as scatters and are shown in Figure 1 (other 5 disorder-promoting residues were plotted in Figure S2). The RMS*p* between IDPs and *ff99SBildn* is about 0.4917%, 0.2195%, and 0.4310%, for Ala, Gly, and Pro, respectively. Two-sample KS test was used to check the statistical significance for the cumulative populations for all the dihedral grids between disorder benchmark and *ff99SBildn* simulated data, as shown in Figure S3. The analysis shows that the differences for the distributions are statistically significant,

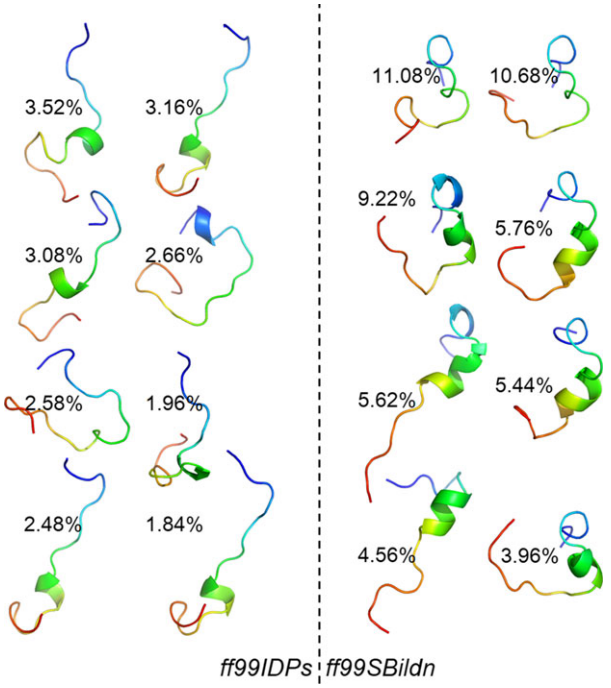


**Figure 5:** Structural representations of the top eight clusters of MeV $N_{TAIL}$ protein with *ff99IDPs* and *ff99SBildn* force fields.

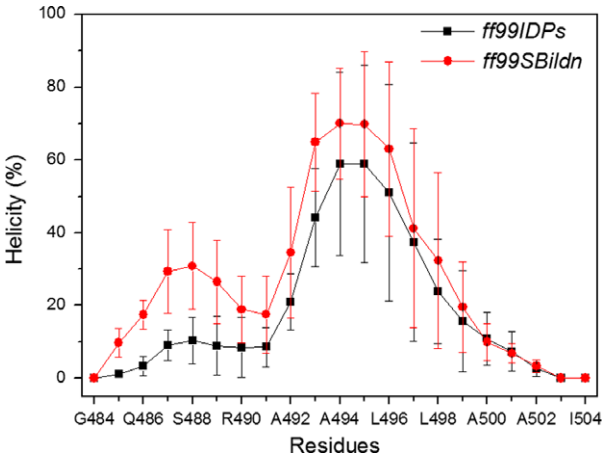


**Figure 6:** The helical propensity of free MeV $N_{TAIL}$ from *ff99SBildn* and *ff99IDPs* simulations.

with the *P* values typically less than 0.001. Therefore, development of new additive force filed is necessary for IDPs.

### CMAP energy term optimization

In the CMAP optimizations, we first tested the necessary sampling time during CMAP optimizations. We compared the RMS*p*s for the different simulation times in the selected step of CMAP optimizations to disordered database dihedral distributions (shown in Figure S2). For the 8 disorder-promoting residues, the fluctuation of RMS*p*s became stable after 60 ns, which indicates that 100-ns simulation is sufficient for the dihedral sampling. More and longer test on proline, lysine, and arginine also suggests that 100-ns trajectories are enough for the equilibration sampling needed for the CMAP optimization.

Over five iteration steps in CMAP optimization, RMS*p* of dihedral distribution between MD and benchmark data for 8 disorder-promoting residues gradually decreases and then remains stable, as shown in Figure 2. The detailed scatters for disorder benchmark, PDB, *ff99SBildn*, and CMAP optimization of five iteration steps are shown in Figure S3. We also performed KS tests on the cumulative populations between disorder benchmark and the final step of CMAP (magenta line in Figure S4), indicating that the cumulative populations for 8 disorder-promoting residues are consistent with the benchmark data. This suggests the optimized CMAP reproduces the benchmark extremely well.

Alanine and lysine are used as two examples to illustrate the optimization process. Heatmaps of RMS*p*s of alanine and lysine for each step are plotted in Figure 2A. The blue blocks represent that the CMAP model has lower population and red ones present that the model has higher population. Population differences between *ff99SBildn* and benchmark data for alanine are concentrated in $\alpha_R$, $PP_{II}$, and $\beta$ regions. For example, $PP_{II}$ region has a population difference of 18.45%. These differences gradually decrease during the optimization iteration (total RMS*p* was lower than 0.06% in the end). However, for lysine, as dihedral populations in the $\alpha_L$ region are low, this region might be over-corrected at the first step of the CMAP optimization (populations in region $\alpha_L$ for the new force field, *ff99SBildn* and the first step of optimization were 2.29%, 0.00%, and 85.26%, respectively). In addition, regions $\alpha_R$ and $\beta$ were also slightly over-corrected (the new force field is lower than *ff99SBildn* but higher than the first step of optimization). These over-corrections would lead RMS*p*s to decrease in the following steps. At the fifth step, population differences for the four regions are all lower than 1%, and the total RMS*p* was 0.05%.

### Force field validation

Two typical IDPs (MeV $N_{TAIL}$ and *p53*) were used to validate the newly developed force field *ff99IDPs*. MeV $N_{TAIL}$ has been shown to have a long intrinsically disordered region, which could fold into highly ordered α-helices upon binding to XD domain of measles virus phosphoprotein (69). N-terminal of *p53* is also a partly IDPs with four-residue helical turn. We simulated free MeV $N_{TAIL}$ (residues 484–504) and free *p53* (residues 17–29) in *ff99IDPs* extensively in explicit solvent. For comparison, the two proteins were also simulated in *ff99SBildn* under the same condition. Overall, each protein was simulated in five independent trajectories of 100 ns in both force fields. RMS*d*, RMS*f*, secondary structure and helicity, $\varphi/\psi$ clustering, secondary chemical shift, and RDC were analyzed and compared with experimental

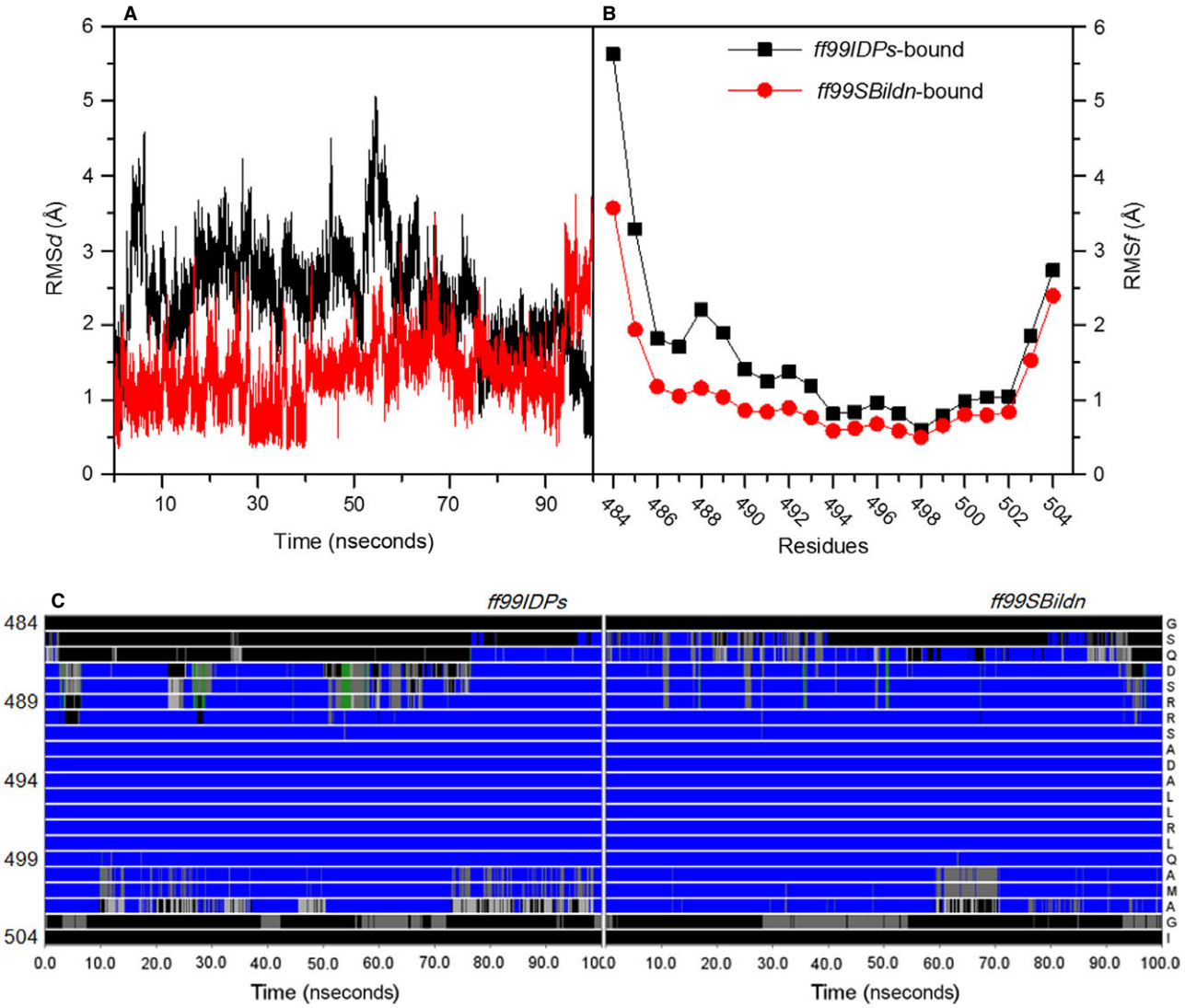


**Figure 7:** RMS*d*, RMS*f*, and secondary structures of bound MeV $N_{TAIL}$ protein with *ff99IDPs* and *ff99SBildn* force fields.

observations. The performances of *ff99IDPs* and *ff99SBildn* were also assessed when MeV $N_{TAIL}$ is in the bound state, for which experimental data are available. Finally, MD simulations for free MeV $N_{TAIL}$ and free *p53* were also conducted in the implicit solvents to check the compatibility of *ff99IDPs* with implicit solvents. Detailed simulation conditions are listed in Table 2.

## Explicit water simulations of MeV $N_{TAIL}$ protein

RMS*d* and RMS*f* for the simulated trajectories are shown in Figure 3A–C. In the free state, RMS*d* of both *ff99IDPs* and *ff99SBildn* trajectories increased quickly and became dynamics equilibration after 50 ns. Average Cα RMS*f*s show a relatively higher flexibility in *ff99IDPs* than that in *ff99SBildn*, which suggests a larger conformational adjustment in the *ff99IDPs* simulations. The cumulative average disorder populations of free MeV $N_{TAIL}$ within 10-ns window are shown in Figure S5 for *ff99IDPs* and *ff99SBildn*, respectively. An increase of approximately 20% disorder population is seen over the course of ~50 ns, after which no further improvement is noted. This suggests that five independent simulations converged to essentially identical families of structures after 50 ns.

The secondary structure of every residue versus simulation time in one representative simulation trajectory for *ff99IDPs* and *ff99SBildn* is shown in Figure 4. In *ff99IDPs* trajectories, all the helices disrupted before ~40 ns and transited into random coils or bends; but in *ff99SBildn* trajectories, despite the disruption of the flexible terminal residues, core helical residues 8–12 remain stable. Detailed information on secondary structure is shown in Figure S6. To obtain further insight on the conformation

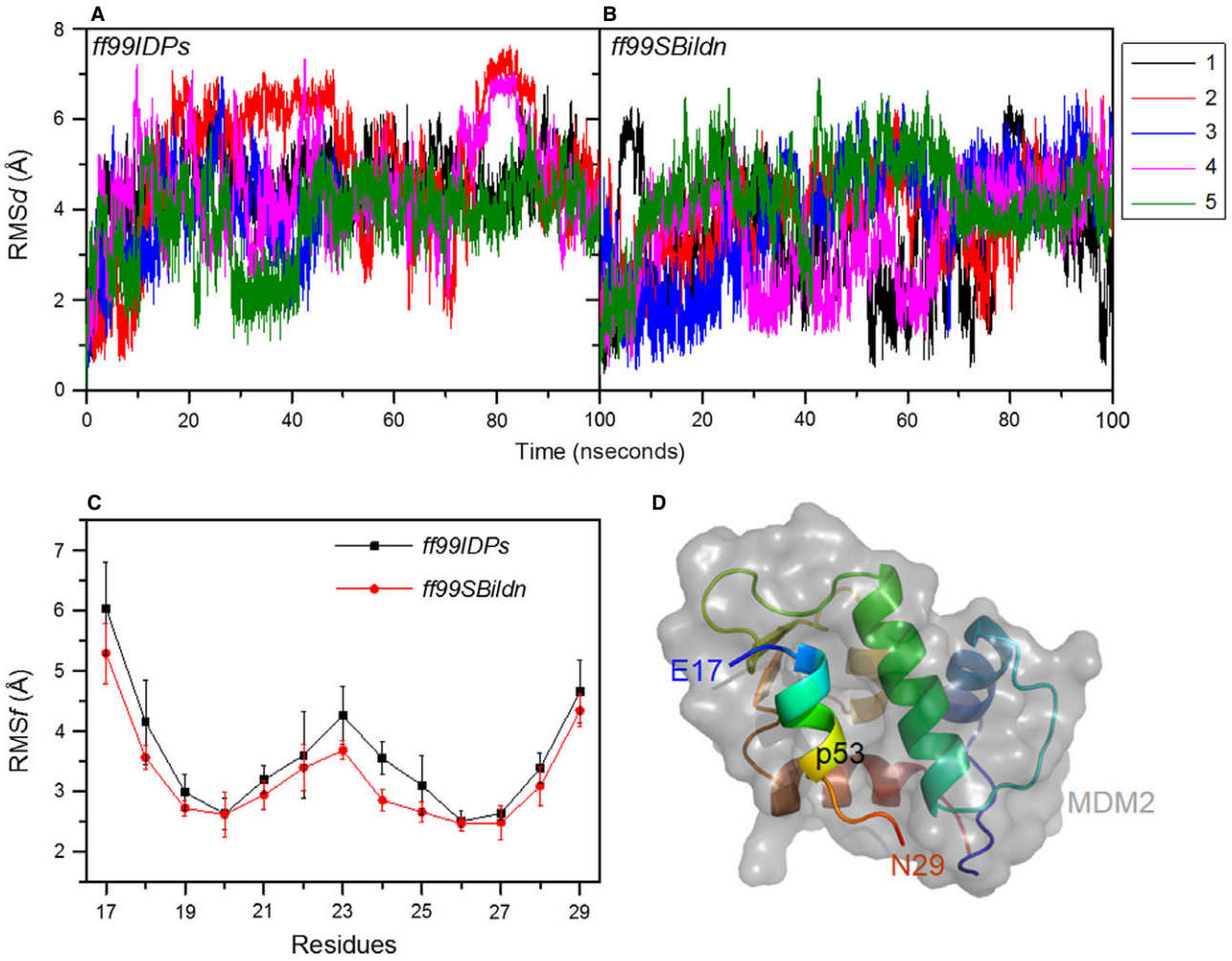


**Figure 8:** RMS*d* and RMS*f* of free *p53* under *ff99SBildn* and *ff99IDPs* force fields.

distribution, we conducted structural clustering based on $\varphi/\psi$ dihedrals. The top 8 clusters with their representative structures for *ff99IDPs* and *ff99SBildn* were gathered and are shown in Figure 5, respectively. The top 8 clusters of *ff99IDPs* simulations occupied 21.28% of all snapshots, while the top 8 clusters of *ff99SBildn* simulations represent 56.32% of all snapshots. This indicates that *ff99IDPs* protein models are more flexible and sample more conformational spaces than *ff99SBildn*, consistent with the IDPs characteristics of conformational heterogeneity (70). Distinct differences can be found between *ff99IDPs* and *ff99SBildn* conformations, especially at the N-terminus, which is in agreement with the previous observations that residues near the N-terminus in the IDPs present a low helical propensity (47,71). Also, the helicity for every residue can be found in Figure 6. The helicity of residues 484–492 modeled by *ff99IDPs* is lower than that by *ff99SBildn*. This is consistent with previous experiment that N-terminal residues have lower helicity than other residues (63,71).

To measure the influence of *ff99IDPs* to IDPs complex, we also performed simulation on bound MeV $N_{TAIL}$ with MeV P protein. RMS*d* and RMS*f* are shown in Figures 7 A and B, which suggest that bound MeV $N_{TAIL}$ is very stable. Furthermore, the secondary structure (shown in Figure 7C) indicates highly ordered structures of the IDPs complex under both force fields. After 100-ns simulation under IDPs force field, the RMS*d* is only 0.807 Å between bound $N_{TAIL}$ protein and crystal structure. This suggests that *ff99IDPs* can still be used to model IDPs complex (44,45).

## Explicit water simulations of *p53*

To further validate the feasibility of *ff99IDPs* on IDPs, we performed MD simulations on the free *p53* (residues 17–29) in *ff99SBildn* and *ff99IDPs*, respectively. RMS*d* and RMS*f* are illustrated in Figure 8. The RMS*d* plot for *ff99IDPs* trajectories demonstrates higher fluctuations than that of *ff99SBildn.* The Cα variations in *p53* in *ff99IDPs* are also slight higher than that in *ff99SBildn*.

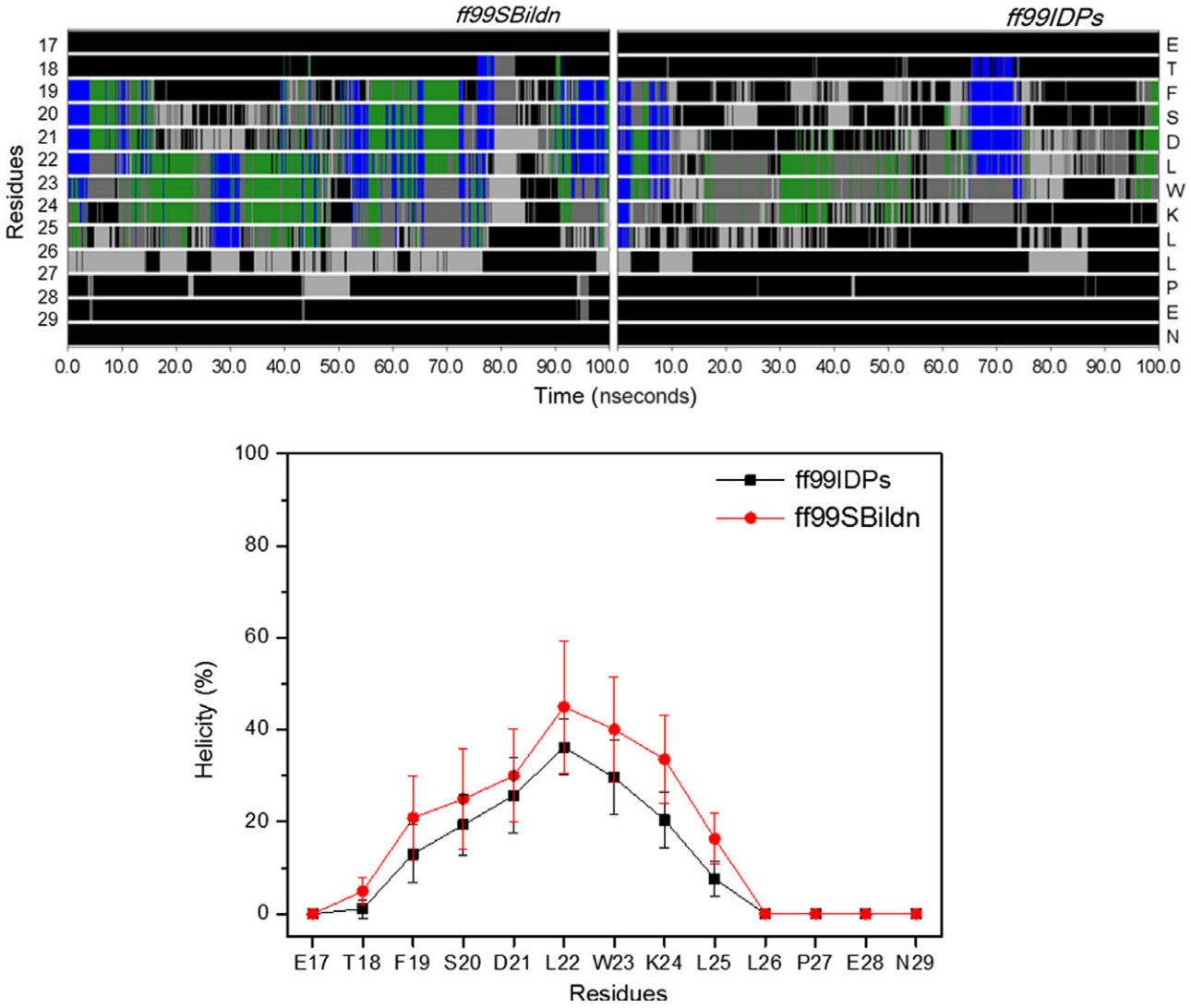


**Figure 9:** Secondary structure and helical propensity of free *p53* under *ff99SBildn* and *ff99IDPs* force fields.

The cumulative average disorder populations of free *p53* from *ff99IDPs* and *ff99SBildn* within 10-ns window are shown in Figure S7, respectively. Similar to MeV $N_{TAIL}$, the disorder populations reach dynamic equilibrium after 50 ns. This suggests that 100-ns simulation is sufficient to collect enough samples of structures for free *p53*.

To further understand the conformational change in p53 while free from MDM2, the secondary structure assignments calculated by DSSP program are shown in Figure 9. In the trajectory under *ff99SBildn* force field, *p53* remains as a stable helical secondary structure during the simulation, which the helical propensity of residue 19–24 is more than 40%. In contrast, the secondary structures dynamically change between helix and random coil for *ff99IDPs* (for more detailed information in Figure S8). This fully illustrates the instability of *p53* as partially intrinsically disordered protein. Interestingly, the helical population for residues 21–24 is nearly 30% that is in quantitative agreement with the experimental observation (48).

Structural cluster was also used to shed light on the heterogeneity of *p53* structures. The top 8 clusters for each simulation under *ff99SBildn* and *ff99IDPs* were extracted and are shown in Figure 10. The top 8 representative structures with *ff99SBildn* and *ff99IDPs* occupy 70.36% and 62.54% of the conformational ensemble, respectively. Similar to MeV $N_{TAIL}$, conformations in *ff99IDPs* are more heterogeneous than those in *ff99SBildn* (70). Five clusters of the structure have helical secondary structure near residues 21–24 for *ff99IDPs*, which is consistent with experimental data. However, the helical secondary structure is observed on the fragment for residues 19–25, which is slight longer than that observed in experiment.

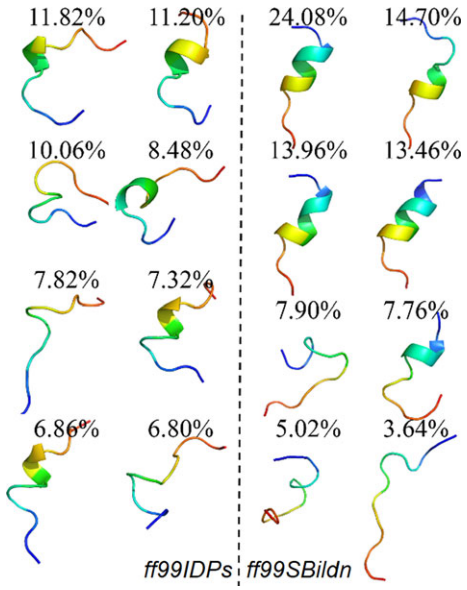


**Figure 10:** Structural representations of the top eight clusters of *p53* under *ff99IDPs* and *ff99SBildn* force fields.

## Comparison with experimental NMR chemical shift

To further evaluate the performance of *ff99IDPs*, average secondary chemical shifts of free and bound MeV $N_{TAIL}$ protein and free *p53* were calculated. The predicted data could be quantitatively compared with the experimental chemical shift.

The experimental and predicted secondary Cα chemical shift data of free and bound MeV $N_{TAIL}$ are shown in Figure 11. This figure suggests that the predicted chemical shift data from *ff99IDPs* simulation are in quantitative agreement with the experimental data with correlation coefficient ($R^2$) of 0.79 for the free state. However, the prediction data from *ff99SBildn* simulation are significantly different from experimental data with $R^2$ of 0.35 for the free state. For the bound state, the correlation between the simulation in *ff99IDPs* and experiment is almost the same to that observed in the free state, with $R^2$ of 0.76, while the predicted chemical shifts in the *ff99SBildn* simulation are also in good agreement with experiment, with $R^2$ of 0.61. This suggests that *ff99IDPs* can be used in both disordered and ordered states for IDPs while *ff99SBildn* only works well for the ordered bound state.

We also calculated the secondary chemical shift for free *p53* and the results are shown in Figure 11. Interestingly, the predictions derived from *ff99IDPs* show a similar agreement with experiment (with a correlation coefficient $R^2$ of 0.63, Figure 11H) to that of *ff99SBildn* ($R^2$ = 0.62, Figure 11I). This is consistent with the structural analyses above showing there are only ~4 residues are forming a helical turn as shown in Figure 10.

Overall, more consistent agreement with the NMR experiment can be observed when the simulations were conducted in *ff99IDPs*.

## Comparison with experimental residue dipolar couplings

Residue dipolar couplings were further used to evaluate the *ff99IDPs* force field. The flexible-meccano (FM)(68) software was utilized to calculate the $NH^N$ couplings of *p53* from the helical propensities in both *ff99IDPs* and *ff99SBildn* simulations. Calculated and experimental RDCs (48) of free *p53* are shown in Figure 12. RDCs values of residues 20–25 in the *ff99IDPs* simulation are positive, indicating that these residues tend to form α-helix, consistent with the experimental RDCs values. The linear fit between predicted RDCs from *ff99IDPs* simulation and experimental data shows a higher correlation ($R^2$ equal to 0.76) than from *ff99SBildn* (underestimated RDCs with $R^2$ of 0.18). Combining with the simulated secondary structure populations described before, our validation suggests that *ff99IDPs* may reproduce the structural properties of IDPs both in free and in bound states.

## Implicit water simulations of MeV $N_{TAIL}$ and *p53*

To test the compatibility of *ff99IDPs* in implicit solvent, we used the PB model to study the structural characters of free MeV $N_{TAIL}$ and *p53* with both *ff99IDPs* and *ff99SBildn*. As expected, similar results as in the explicit solvent model were observed and are shown in Figure S9. Cα chemical shift data of free MeV $N_{TAIL}$ protein under the *ff99IDPs*/PB model have a higher correlation coefficient $R^2$ (0.64) with experimental data than that of the *ff99SBildn*/PB model (0.36). Chemical shift data of free *p53* were also predicted close to experimental data for *ff99IDPs* with $R^2$ of 0.76 and 0.71 for *ff99SBildn*. These results indicate that *ff99IDPs* can also be used in the implicit water model for IDPs studies.

## Conclusion

Intrinsically disordered proteins play important biological function in cell signaling and cancer upon binding with multiple interaction partners. These IDPs are found in many diseases. However, widely used force fields could not accurately simulate the property of IDPs. In this study, we develop a specific force field to solve this problem. At first, the distributions of $\varphi/\psi$ dihedral for disordered residues are significant different from those of PDB structures. Then, we report an effort on improving the $\varphi/\psi$ dihedral terms with CMAP energy term in *ff99SBildn* energy function. The results of CMAP optimization indicate that *ff99IDPs* force field could reproduce the $\varphi/\psi$ dihedral distribution of 8 disordered residues, and these distributions are similar to those of disordered benchmark data. Finally, the test of two IDPs proteins confirms that *ff99IDPs* improves the conformer distribution of IDPs. The helical

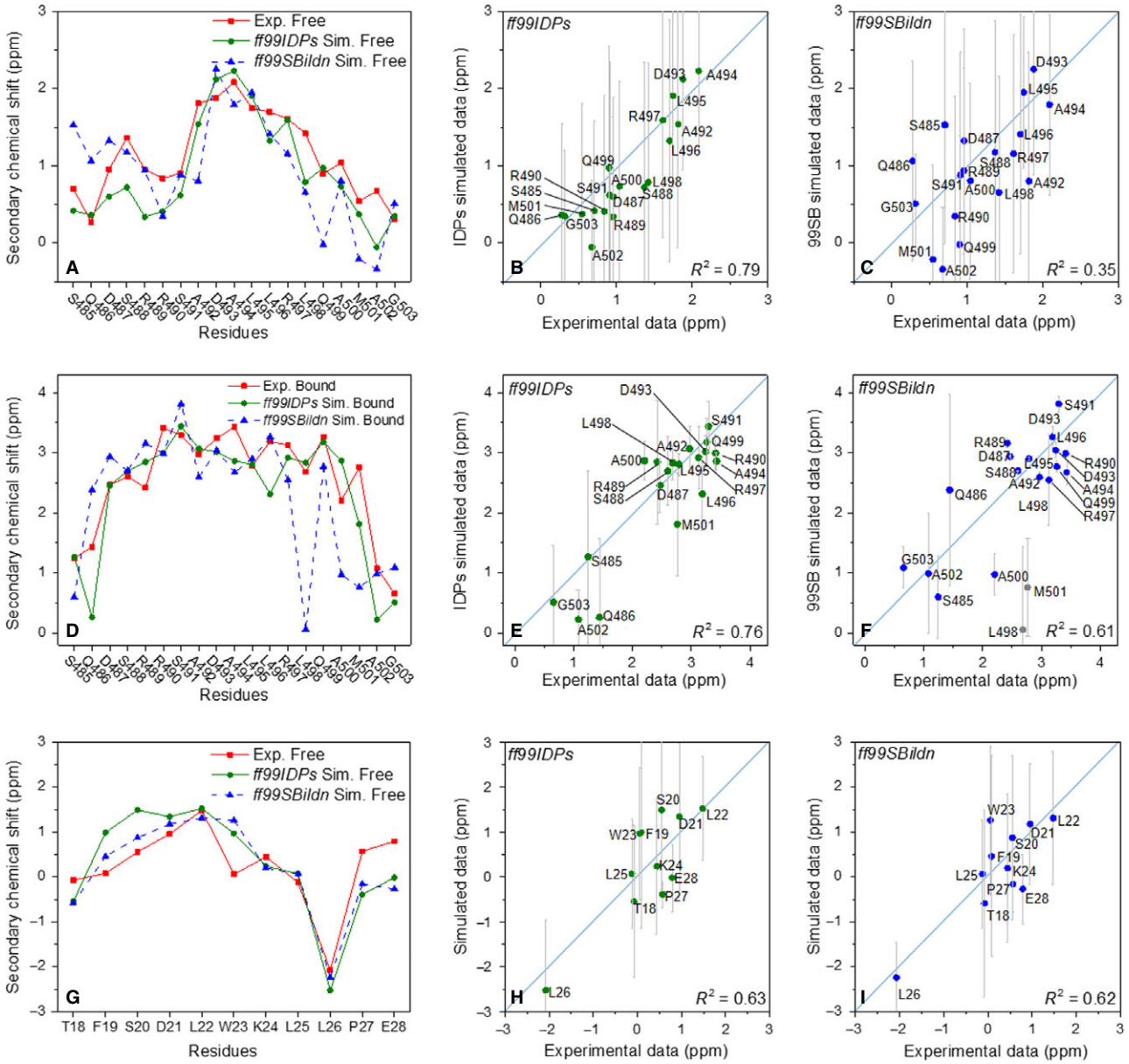


**Figure 11:** Secondary Cα chemical shift comparisons between predicted data and experimental data. (A) Comparisons for free MeV $N_{TAIL}$ protein. Red lines: experimental data; Green lines: predicted data under *ff99IDPs* force field; Blue dashes: predicted data under *ff99SBildn*. (B) Correlations between predicted secondary chemical shift for free MeV $N_{TAIL}$ protein under *ff99IDPs* force field driving MD simulations of free MeV $N_{TAIL}$ protein and the corresponding experimental data, with correlation coefficient ($R^2$) of 0.79. (C) Correlations between predicted secondary chemical shift for free MeV $N_{TAIL}$ protein under *ff99SBildn* force field and the corresponding experimental data, $R^2$ is 0.35. (D) Comparisons for bound MeV $N_{TAIL}$ protein. Red lines: experimental data; Green lines: predicted data under *ff99IDPs* force field; Blue dashes: predicted data under *ff99SBildn*. (E) Correlations between predicted secondary chemical shift for *ff99IDPs* force field driving MD simulations of bound MeV $N_{TAIL}$ protein and the corresponding experimental data, with correlation coefficient ($R^2$) of 0.76. (F) Correlations between predicted secondary chemical shift for bound MeV $N_{TAIL}$ protein under *ff99SBildn* force field and the corresponding experimental data, $R^2$ is 0.61 (L498 and M501 are outliers, colored in gray). (G) Comparisons for free *p53*. Red line: experimental data; Green line: predicted data under *ff99IDPs* force field; Blue dashes: predicted data under *ff99SBildn*. (H) Correlations between predicted secondary chemical shift data and experimental data under *ff99IDPs* force filed ($R^2$ = 0.63). (I) Correlations between predicted secondary chemical shift data and experimental data under *ff99SBildn* force filed ($R^2$ = 0.62).

location and helical population of simulated free *p53* are agreement with those of the experimental observation. Furthermore, the predicted secondary chemical shift data for MeV $N_{TAIL}$ and *p53* are in quantitative accord with experimental data. Finally, *ff99IDPs* can also be used in implicit water model for IDPs. In summary, *ff99IDPs* can

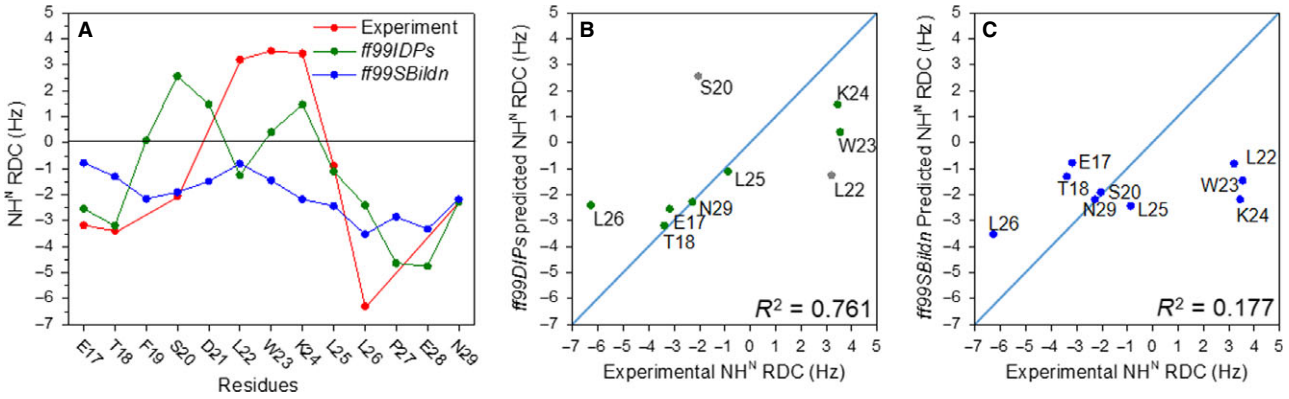


**Figure 12:** Comparison of free *p53* (17–29) $NH^N$ RDCs between predicted data and experimental data. (A) Red line: experimental data; Green line: predicted data under *ff99IDPs* force field; Blue line: predicted data under *ff99SBildn* force field. (B) Linear fit of predicted RDC under *ff99IDPs* (green solid circles), with $R^2$ = 0.715 (S20 and L22 are outliers). Predicted data under *ff99SBildn* were also plotted with blue empty circles, which have low correlation with experimental data. (C) Correlation for predicted data from *ff99SBildn* simulation with experimental data.

reproduce the conformation of IDPs or IDRs both in bound and in free states. More tests will be performed next.

## Acknowledgments

This work was supported by Center for HPC at Shanghai Jiao Tong University, grants from the Ministry of Science and Technology of China (2012CB721003), by the National High-tech R&D Program of China (863 Program) (2014AA021502), the National Natural Science Foundation of China (J1210047 and 31271403), by the Innovation Program of the Shanghai Education Committee (Grants No. 12ZZ023), and by Medical Engineering Cross Fund of Shanghai Jiaotong University (YG2013MS68).

## Conflict of Interest

There is no conflict of interest.

## References

1. Romero P., Obradovic Z., Li X., Garner E.C., Brown C.J., Dunker A.K. (2001) Sequence complexity of disordered protein. Proteins;42:38–48.
2. Dunker A.K., Obradovic Z., Romero P., Garner E.C., Brown C.J. (2000) Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform;11:161–171.
3. Dunker A.K., Brown C.J., Lawson J.D., Iakoucheva L.M., Obradovic Z. (2002) Intrinsic disorder and protein function. Biochemistry;41:6573–6582.
4. Jirgensons B. (1966) Classification of proteins according to conformation. Die Makromolekulare Chemie;91:74–86.
5. Wright P.E., Dyson H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol;293:321–331.
6. Dunker A.K., Lawson J.D., Brown C.J., Williams R.M., Romero P., Oh J.S., Oldfield C.J., Campen A.M., Ratliff C.M., Hipps K.W. (2001) Intrinsically disordered protein. J Mol Graph Model;19:26–59.
7. Uversky V.N., Oldfield C.J., Dunker A.K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys;37:215–246.
8. Dunker A.K., Babu M.M., Barbar E., Blackledge M., Bondos S.E., Dosztányi Z., Dyson H.J., Forman-Kay J., Fuxreiter M., Gsponer J. (2013) What's in a name? Why these proteins are intrinsically disordered. Intrinsically Disord Proteins;1:e24157.
9. Theillet F.-X., Kalmar L., Tompa P., Han K.-H., Selenko P., Dunker A.K., Daughdrill G.W., Uversky V.N. (2013) The alphabet of intrinsic disorder I. Act like a Pro: on the abundance and roles of proline residues in intrinsically disordered proteins. Intrinsically Disordered. Proteins;1:e24360.
10. Iakoucheva L.M., Brown C.J., Lawson J.D., Obradović Z., Dunker A.K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol;323:573–584.
11. Tsai C.-J., Ma B., Nussinov R. (2009) Protein–protein interaction networks: how can a hub protein bind so many different partners? Trends Biochem Sci;34:594–600.
12. Collavin L., Lunardi A., Del Sal G. (2010) p53-family proteins and their regulators: hubs and spokes in tumor suppression. Cell Death Differ;17:901–911.
13. Chouard T. (2011) Structural biology: breaking the protein rules. Nature;471:151–153.
14. Uversky V.N., Oldfield C.J., Midic U., Xie H., Xue B., Vucetic S., Iakoucheva L.M., Obradovic Z., Dunker A.K. (2009) Unfoldomics of human diseases: linking

protein intrinsic disorder with diseases. BMC Genomics;10(Suppl. 1): S7.
15. Wenning G.K., Jellinger K.A. (2005) The role of [alpha]-synuclein and tau in neurodegenerative movement disorders. Curr Opin Neurol;18:357–362.
16. Pawar A.P., Dubay K.F., Zurdo J., Chiti F., Vendruscolo M., Dobson C.M. (2005) Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. J Mol Biol;350:379–392.
17. Hampel H., Blennow K., Shaw L.M., Hoessler Y.C., Zetterberg H., Trojanowski J.Q. (2010) Total and phosphorylated tau protein as biological markers of Alzheimer's disease. Exp Gerontol;45:30–40.
18. Goedert M., Spillantini M., Jakes R., Rutherford D., Crowther R. (1989) Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. Neuron;3:519–526.
19. Gamblin T.C., Berry R.W., Binder L.I. (2003) Modeling tau polymerization in vitro: a review and synthesis. Biochemistry;42:15009–15017.
20. Cheng Y., LeGall T., Oldfield C.J., Dunker A.K., Uversky V.N. (2006) Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry;45:10448–10460.
21. Riek R., Hornemann S., Wider G., Glockshuber R., Wuthrich K. (1997) NMR characterization of the full-length recombinant murine prion protein, mPrP(23-231). FEBS Lett;413:282–288.
22. Klepeis J.L., Lindorff-Larsen K., Dror R.O., Shaw D.E. (2009) Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol;19:120–127.
23. Lindorff-Larsen K., Piana S., Palmo K., Maragakis P., Klepeis J.L., Dror R.O., Shaw D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins;78:1950–1958.
24. Ye W., Wang W., Jiang C., Yu Q., Chen H. (2013) Molecular dynamics simulations of amyloid fibrils: an in silico approach. Acta Biochim Biophys Sin (Shanghai);45:503–508.
25. Pearlman D.A., Case D.A., Caldwell J.W., Ross W.S., Cheatham T.E. III, DeBolt S., Ferguson D., Seibel G., Kollman P. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun;91:1–41.
26. Brooks B.R., Bruccoleri R.E., Olafson B.D., Swaminathan S., Karplus M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem;4:187–217.
27. Berendsen H.J., van der Spoel D., van Drunen R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. Comput Phys Commun;91:43–56.
28. Jorgensen W.L., Tirado-Rives J. (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc;110:1657–1666.
29. Bernstein F.C., Koetzle T.F., Williams G.J., Meyer E.F. Jr, Brice M.D., Rodgers J.R., Kennard O., Shimanouchi T., Tasumi M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol;112:535–542.
30. Allen F.H., Bellard S., Brice M., Cartwright B.A., Doubleday A., Higgs H., Hummelink T., Hummelink-Peters B., Kennard O., Motherwell W. (1979) The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. Acta Crystall B - Stru;35:2331–2339.
31. Wang Z.X., Zhang W., Wu C., Lei H., Cieplak P., Duan Y. (2006) Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. J Comput Chem;27:781–790.
32. Antila H.S., Salonen E. (2013) Polarizable force fields. In: Luca M., Emppu S., editors. Biomolecular Simulations. New York, NY: Humana Press; p. 215–241.
33. Halgren T.A., Damm W. (2001) Polarizable force fields. Curr Opin Struct Biol;11:236–242.
34. Ponder J.W., Wu C., Ren P., Pande V.S., Chodera J.D., Schnieders M.J., Haque I., Mobley D.L., Lambrecht D.S., DiStasio R.A. Jr (2010) Current status of the AMOEBA polarizable force field. J Phys Chem B;114:2549–2564.
35. Garcia A.E., Sanbonmatsu K.Y. (2002) Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. Proc Natl Acad Sci U S A;99:2782–2787.
36. Sorin E.J., Pande V.S. (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. Biophys J ;88:2472–2493.
37. Hornak V., Abel R., Okur A., Strockbine B., Roitberg A., Simmerling C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins;65:712–725.
38. Mathura V., Braun W., Garner E., Young J., Takayama S., Brown C., Dunker A. (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder. 89–100.
39. Midic U., Obradovic Z. (2011) Intrinsic disorder in putative protein sequences. Proteome Sci;10:S19.
40. MacKerell A.D., Feig M., Brooks C.L. (2004) Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc;126:698–699.
41. MacKerell A.D., Feig M., Brooks C.L. (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem;25:1400–1415.
42. Wright P.E., Dyson H.J. (2009) Linking folding and binding. Curr Opin Struct Biol;19:31–38.

43. Brown C.J., Johnson A.K., Dunker A.K., Daughdrill G.W. (2011) Evolution and disorder. Curr Opin Struct Biol;21:441–446.
44. Longhi S., Receveur-Bréchot V., Karlin D., Johansson K., Darbon H., Bhella D., Yeo R., Finet S., Canard B. (2003) The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J Biol Chem;278:18638–18648.
45. Bourhis J.-M., Johansson K., Receveur-Bréchot V., Oldfield C.J., Dunker K.A., Canard B., Longhi S. (2004) The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. Virus Res;99:157–167.
46. Bourhis J.M., Receveur-Bréchot V., Oglesbee M., Zhang X., Buccellato M., Darbon H., Canard B., Finet S., Longhi S. (2005) The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. Protein Sci;14:1975–1992.
47. Wang Y., Chu X., Longhi S., Roche P., Han W., Wang E., Wang J. (2013) Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. Proc Natl Acad Sci U S A;110:E3743–E3752.
48. Wells M., Tidow H., Rutherford T.J., Markwick P., Jensen M.R., Mylonas E., Svergun D.I., Blackledge M., Fersht A.R. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A;105:5762–5767.
49. Chen H.F., Luo R. (2007) Binding induced folding in p53-MDM2 complex. J Am Chem Soc;129:2930–2937.
50. Kabsch W., Sander C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers;22:2577–2637.
51. Joosten R.P., te Beek T.A., Krieger E., Hekkelman M.L., Hooft R.W., Schneider R., Sander C., Vriend G. (2011) A series of PDB related databases for everyday needs. Nucleic Acids Res;39:D411–D419.
52. MacKerell A.D., Bashford D., Bellott M., Dunbrack R., Evanseck J., Field M.J., Fischer S., Gao J., Guo H., Ha S. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B;102:3586–3616.
53. Case D.A., Cheatham T.E., Darden T., Gohlke H., Luo R., Merz K.M., Onufriev A., Simmerling C., Wang B., Woods R.J. (2005) The Amber biomolecular simulation programs. J Comput Chem;26:1668–1688.
54. Case D., Darden T., Cheatham T. III, Simmerling C., Wang J., Duke R., Luo R., Walker R., Zhang W., Merz K. (2012) AMBER 12. San Francisco: University of California.
55. Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W., Kollman P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc;117:5179–5197.
56. Wilcox R. (2005) Kolmogorov–Smirnov test. Encyclopedia of Biostatistics.
57. Berger V.W., Zhou Y. (2005) Kolmogorov–Smirnov tests. Encyclopedia of Statistics in Behavioral Science. 2:1023–1006.
58. Wlodarski T., Zagrovic B. (2009) Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. Proc Natl Acad Sci U S A;106:19346–19351.
59. Darden T., York D., Pedersen L. (1993) Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. J Chem Phys;98:10089–10092.
60. Goetz A., Salomon-Ferrer R., Poole D., Grand S., Walker R. (2013) Routine microsecond molecular dynamics simulations with AMBER. Part II: particle mesh Ewald. J Chem Theory Comput;9:3878–3888.
61. Götz A.W., Williamson M.J., Xu D., Poole D., Le Grand S., Walker R.C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. J Chem Theory Comput;8:1542.
62. Ryckaert J.P., Ciccotti G., Berendsen H.J.C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys;23:327–341.
63. Kingston R.L., Hamel D.J., Gay L.S., Dahlquist F.W., Matthews B.W. (2004) Structural basis for the attachment of a paramyxoviral polymerase to its template. Proc Natl Acad Sci U S A;101:8301–8306.
64. Kussie P.H., Gorina S., Marechal V., Elenbaas B., Moreau J., Levine A.J., Pavletich N.P. (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science;274:948–953.
65. Feig M., Karanicolas J., Brooks C.L. III (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model;22:377–395.
66. Ulrich E.L., Akutsu H., Doreleijers J.F., Harano Y., Ioannidis Y.E., Lin J., Livny M., Mading S., Maziuk D., Miller Z. (2008) BioMagResBank. Nucleic Acids Res;36:D402–D408.
67. Shen Y., Bax A. (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR;38:289–302.
68. Ozenne V., Bauer F., Salmon L., Huang J.-R., Jensen M.R., Segard S., Bernadó P., Charavay C., Blackledge M.. (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics;28:1463–1470.

69. Jensen M.R., Communie G., Ribeiro E.A. Jr, Martinez N., Desfosses A., Salmon L., Mollica L., Gabel F., Jamin M., Longhi S., Ruigrok R.W., Blackledge M. (2011) Intrinsic disorder in measles virus nucleocapsids. Proc Natl Acad Sci U S A;108:9839–9844.
70. Mao A.H., Crick S.L., Vitalis A., Chicoine C.L., Pappu R.V. (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc Natl Acad Sci U S A;107:8183–8188.
71. Jensen M.R., Communie G., Ribeiro E.A., Martinez N., Desfosses A., Salmon L., Mollica L., Gabel F., Jamin M., Longhi S. (2011) Intrinsic disorder in measles virus nucleocapsids. Proc Natl Acad Sci U S A;108:9839–9844.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Figure S1**. Histograms of the populations of the four principal regions of $\varphi/\psi$ dihedrals for different loop length limits (labeled as different colors).

**Figure S2**. Simulation condition for CMAP and parameter optimization.

**Figure S3**. $\varphi/\psi$ distributions of disorder benchmark, PDB, *ff99SBildn*, and CMAP minimizations of 5 iteration steps for the 8 disorder-promoting residues.

**Figure S4**. Cumulative population for disorder benchmark, PDB, *ff99SBildn*, and the final step of CMAP optimization for 8 di-residues.

**Figure S5.** The cumulative average disorder population of free MeV $N_{TAIL}$ within 10-ns window from *ff99IDPs* and *ff99SBildn* simulations.

**Figure S6**. Secondary structures along simulation time for free MeV $N_{TAIL}$ under *ff99IDPs* and *ff99SBildn*.

**Figure S7.** The cumulative average disorder population of free p53 within 10-ns window from *ff99IDPs* and *ff99SBildn*.

**Figure S8**. Secondary structures along simulation time for free p53 under *ff99IDPs* and *ff99SBildn*.

**Figure S9**. Secondary Cα chemical shift comparisons between implicit MD predicted data and experimental data.