UC Irvine
UC Irvine Previously Published Works

Title
Well-Balanced Force Field ff03CMAP for Folded and Disordered Proteins.

Permalink
https://escholarship.org/uc/item/2hf2d3b6

Journal
Journal of Chemical Theory and Computation, 15(12)

ISSN
1549-9618

Authors
Zhang, Yangpeng
Liu, Hao
Yang, Sheng
et al.

Publication Date
2019-12-10

DOI
10.1021/acs.jctc.9b00623

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2020 May 03.

Published in final edited form as:

J Chem Theory Comput. 2019 December 10; 15(12): 6769–6780. doi:10.1021/acs.jctc.9b00623.

A Well-balanced Force Field *ff03CMAP* for Folded and Disordered Proteins

Yangpeng Zhang[1], **Hao Liu**[1], **Sheng Yang**[1], **Ray Luo**[2,*], **Hai-Feng Chen**[1,3,*]

[1]State Key Laboratory of Microbial metabolism, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, China

[2]Departments of Molecular Biology and Biochemistry, Chemical and Molecular Engineering, and Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine, USA

[3]Shanghai Center for Bioinformation Technology, Shanghai, 200235, China

Abstract

Molecular dynamics simulation as an important complement of experiment is widely used to study protein structures and functions. However, previous studies indicate that current force fields cannot, simultaneously, provide accurate descriptions of folded proteins and intrinsically disordered proteins (IDPs). Therefore, a CMAP optimized force field based on the Amber *ff03* force field (termed *ff03CMAP* herein) was developed for balanced sampling of folded proteins and IDPs. Extensive validations of short peptides, folded proteins, disordered proteins, and fast-folding proteins show that simulated chemical shifts, J-coupling constants, order parameters, and residual dipolar couplings with the *ff03CMAP* force field are in very good agreement with NMR measurements and are more accurate than other *ff03*-series force fields. The influence of solvent models was also investigated. It was found that the combination of *ff03CMAP*/TIP4P-Ew is suitable for folded proteins and that of *ff03CMAP*/TIP4PD is better for disordered proteins. These findings confirm that the newly developed force field *ff03CMAP* can improve the balance of conformer sampling between folded proteins and IDPs.

Introduction

Proteins can exist in three states, folded, molten globule, and random coil.[1] Folded proteins are easier to study because they are ordered and stable. But disordered proteins also need exploring. In eukaryotes, more than 30 percent of proteins contain disordered regions with more than 50 consecutive residues.[1] Proteins with disordered regions or overall intrinsically disordered proteins (IDPs) have been proved to have important biological functions, such as molecular recognition, molecular assembly, protein modification, and so on.[2] Furthermore, IDPs are associated with many human diseases, such as Alzheimer's disease, Parkinson's disease, Huntington's disease, cancer, and cardiovascular disease, to

*Corresponding authors: haifengchen@sjtu.edu.cn; ray.luo@rayluolab.org, Tel: 86-21-34204348, Fax: 86-21-34204348.

The authors declare that there is no conflict of interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

name a few.[3–4] IDPs are more flexible and unstable with little secondary structures than structured or ordered proteins. “Intrinsically disordered” implies a sequence-dependent nature in IDPs that trend to be lack of ordered structures.[5] Many experimental methods have been utilized to study IDPs, such as electron paramagnetic resonance (EPR), X-ray diffraction, Nuclear Magnetic Resonance (NMR), and small-angle X-ray scattering (SAXS).[6–8]

Because of their important biological functions, IDPs have become common topics in molecular dynamics studies in recent years. Due to limited accuracy in standard protein force fields, a set of special-purpose force fields have been developed for simulating IDPs, such as *ff99IDPs*, *ff14IDPs*, *ff14IDPSFF*, *ff03ws*, RSFF2, a99SB-disp, CHARMM36IDPSFF and so on.[9–15] In addition, the D. E. Shaw group also modified the dispersion interaction of TIP4P water model (TIP4P-D) to improve the quality of IDPs simulations.[16] However, it remains elusive to reach a good balance between ordered and disordered states with either standard or special-purpose force fields.

ff03 is a new-generation Amber force field that has become widely used in biomolecular simulation studies.[17] Based on *ff03*, the Best group modified backbone dihedral potentials in the context of the TIP4P/2005 water model. Their efforts lead to three *ff03* variants: *ff03**, *ff03w*, and *ff03ws*.[12, 18–19] These modifications were shown to partially improve the performance of conformer sampling of IDPs and folded proteins. The effect of solvent model was found to be important in the sampling of IDPs in their studies.[12, 18–19] TIP3P is the most commonly used water model in earlier force fields.[20] Improvement of water models has always been concurrent with protein force field developments. Besides TIP4P-D, TIP4P-Ew and TIP4P/2005 are also two commonly used four-site water models and should be investigated in any force field improvement efforts.[21–22] These four-site water models have been found to reproduce well the hydrophobic effect and water density in a wide temperature range.

In this development, we systematically analyzed the original *ff03* force field, its published variants, and explored a new variant based on the CMAP method, *ff03CMAP*. The combinations of force fields and corresponding solvent models are *ff03* with TIP3P, *ff03** with TIP3P, *ff03w* with TIP4P/2005, *ff03ws* with TIP4P/2005 as in the published efforts. Based on our analyses, we recommend the combinations of *ff03CMAP* with TIP4P-Ew and TIP4P-D, respectively, suitable for folded proteins and IDPs.[12, 17–19] In order to evaluate the performance of these force fields, 16 proteins were simulated to probe the quality of various force field/water model combinations in reproducing experimental measurables. The tested short peptides and proteins are shown in Figure 1.

Material and Methods

Molecular Dynamics Simulations.

Initial structures were built by the LEaP module in the AMBER 14 suite if not available,[38] which was also used to conduct MD simulations.[39] All systems were neutralized and solvated in boxes of different water models.[20] All bonds involving hydrogen atom were constrained with the SHAKE algorithm.[43] Particle Mesh Ewald (PME) algorithm was used

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

to calculate long-range electrostatic interactions.[41] Initial structures were relaxed with 10000 steps of minimization, then subject to heating up for 20ps, and equilibration for 10ps in the NPT ensemble with PMEMD. The CUDA version of PMEMD was used to accelerate the simulations.[42] The simulation temperature and ion strength were set according to their respective experimental conditions. The number of replica-exchange molecular dynamics (REMD) replicas and temperatures were set by an online temperature predictor for parallel tempering simulations.[40] All simulation conditions are shown in Table S1.

Insert a table for all tested force field water combinations.

Benchmark of PDB Coil Structures.

Coil database was built and extracted from PDB. The DSSP program was utilized to classify the secondary structures and extract main chain dihedrals from these structures.[44–45] 2,611,450 amino acids without secondary structure were collected. The counts of amino acids in the coiled database are shown in Figure S1. The Ramachandran plots for the database were used as the benchmark for the optimization of dihedral distribution.

CMAP Method.

Grid-based energy correction maps (CMAP) is a useful method for automatically correcting dihedral distribution for the additive force field, which is based on the backbone dihedral distribution and has been used to develop IDP-specific force fields.[10, 46] CMAP was first published to modify the CHARMM force field and was transferred into Amber software.[9–11, 47–50] We minimized the main-chain dihedral distribution differences between the MD simulation and the benchmark for each of the 20 amino acids. A 576-(24 × 24) grid was used to cover the phi/psi map. The tetrapeptide models (Nme-Ala-X-Ala-Ace, where X represents one of 20 naturally amino acids, Nme for aminomethyl, and Ace for acetyl) were utilized in the CMAP optimization *via* MD simulations in the TIP4PEw water model. Ten cycles of CMAP optimization were conducted to minimize the distribution differences between the MD simulation and the benchmark. In each cycle, the solvated tetrapeptides were simulated for 200 nanoseconds. After the CMAP optimization, we added an additional structural factor which is the partial dihedral energy distribution of 'S' fragments predicted by DSSP in the PDB coil database to avoid overestimation of disordered state when using the new force field. Root-mean-square deviations of population (RMSp) is calculated to compare the difference between the MD simulation and the benchmark with equation (1).

$$RMSp = \sqrt{\frac{\sum_{i=1}^{576}\left(P_i^{DB} - P_i^{MD}\right)^2}{576}} \quad (1)$$

where P_i^{DB} is the population of the *i*-th grid in database benchmark and P_i^{MD} is the population of the *i*-th grid in the MD simulation.

Quantification in the Evaluation for Force Fields.

To quantitatively compare different *ff03* variants for folded proteins and disordered proteins with experimental measurements, we utilized the normalized force-field score.[14] For folded

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

proteins, equation (2) was used to calculate averaged normalized RMSD from each class of experimental data as

$$Folded\ Protein\ FF_{Score} = \frac{1}{N}\sum_{i=1}^{N}\frac{FF_{rmsd}}{rmsd_{Norm}} \quad (2)$$

where N is the number of classes of experimental measurements, FF_{rmsd} is the RMSD of the i-th class for simulated and experimental values, and $rmsd_{Norm}$ is the lowest RMSD of i-th class in all force fields. According to this metric, FF_{Score} is always greater than or equal to 1, and 1 is the best score theoretically which means this force field perfectly reproduced the experimental data. For disordered protein, we divided the experimental measurements into two groups because there are fewer experimental measurements than the folded proteins, for which there are chemical shifts and other NMR measurements. If experimental data for both chemical shifts and NMR measurements are available, FF_{Score} is calculated with equation (3).

$$Disordered\ Protein\ FF_{Score} = \frac{CS_{Score} + NMR_{Score}}{2} \quad (3)$$

and if there is only chemical shift, the FF_{Score} is calculated with equation (4).

$$Disordered\ Protein\ FF_{Score} = CS_{Score} \quad (4)$$

Where CS_{Score} and NMR_{Score} are calculated as same as the score of class in folded protein FF_{Score}.

Calculation of Experimental Observables.

Backbone chemical shifts were calculated by SHIFTX2 for Cα, Cβ, C, N, Hα and HN atom types.[51] Backbone scalar coupling constants were calculated using published Karplus relations for $^{3}J_{HNH\alpha}$, $^{3}J_{H\alpha C}$, $^{3}J_{HNC}$, $^{3}J_{CC}$, $^{3}J_{HNC\alpha}$, $^{3}J_{HNC\beta}$, $^{2}J_{C\alpha N}$, $^{1}J_{C\alpha N}$, $^{1}J_{H\alpha C\alpha}$, and $^{1}J_{C\alpha C\beta}$[52–60] and side-chain scalar coupling constants with Karplus relations for $^{3}J_{CC\gamma}$ and $^{3}J_{NC\gamma}$.[61] Backbone RDCs were calculated using PALES with a local alignment window of 15 residues.[62–63] Backbone amide and side-chain methyl axis S^2 order parameters were calculated with the direct method described in *Trbovic* et al.[64] Small angle X-ray scattering (SAXS) curves were calculated using the FoXS package.[65] Cα RMSD and radius of gyrations (Rg) were calculated using CPPTRAJ in AmberTools.[38] Conformational clustering was performed with the kClust program in the MMTSB tool.[66] MDTraj, a python package, was also used for miscillenous calculations.[67] The PyMOL molecular visualization system was used to show 3D structures for all proteins.[68] All experimental measurements are listed in Table S2.

Biphasic exponential decay model was used to analyze the IDPs sampling convergence for force fields. The equations (5) and (6) were used to calculate the fitting half-time.[71]

$$\Delta C\alpha\ Chemical\ Shift(N_t) = A_1 e^{-\left(\frac{x}{\tau_1}\right)} + A_2 e^{-\left(\frac{x}{\tau_2}\right)} + N_0 \quad (5)$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$t_{1/2} = \tau ln(2) \quad (6)$$

where N_0 is the plateau of an observable, $t_{1/2}$ is the half-life time, A and τ are constants. In this model, the decay consists of two stages for fast stage and slow stage. The slower τ_2 value was utilized to calculate the slow stage half-life and evaluate the convergence rate of IDPs simulation.

Results and Discussion

CMAP Optimization.

Ten cycles of CMAP optimization were performed for each amino acid. In the first cycle (CMAP0), the initial φ/ψ distribution was obtained from the standard *ff03* force field, where the lowest RMSp is 0.234% among 20 amino acids. In contrast, after 10 cycles of optimization, the lowest RMSp is less than 0.064%, as shown in Figure S2. Comparison of the distributions of CMAP0 and the benchmark database, we found that there is almost no left-handed helix distribution except MET, GLY and LEU. In addition, an obviously energy barrier exists between the β-sheet region and the α-helix region, so it would be difficult to sample both types of structures. After optimization, these limitations no longer present. The parameters for the best RMSp for each amino acid were selected as the final CMAP values. These parameters and structural factors were integrated with standard *ff03* force field to generate the new force field *ff03CMAP*.

Evaluation of *ff03*-series Force Fields.

We assessed the performance of *ff03*-series force fields in reproducing experimental data. The same conditions were used in all MD simulation among all tested force fields. The FF scores for short peptides, IDPs, and fold proteins are shown in Figure 2 and specific values are listed in supplementary Table S3. Figure 2 suggests that combination *ff03CMAP/TIP4PD* agrees the best with experiment for short peptides and IDPs. In addition, combination *ff03CMAP/TIP4PEw* leads to the best agreement with experiment for folded proteins. In summary, use of *ff03CMAP* can yield more accurate sampling the conformers for short peptide, IDPs and folded proteins.

Short Peptide Ala_5.

Table 1 shows the RMSD's of secondary chemical shifts, J-coupling constants and force field score for Ala_5. There are 6 types of secondary chemical shifts and 7 types of J-coupling constants. For the CS score, the performance of *ff03CMAP/TIP4PEw* are much better than all other force fields. However, the performance of *ff03*/TIP3P* is the best for the NMR score. If we combine CS and NMR scores (i.e. overall FF score), *ff03*-drived force fields are significantly improved over the origin *ff03* and the *ff03CMAP/TIP4PEw* is the best. The detail RMSD's of secondary chemical shifts and J-coupling constants are shown in Figures S3–S4.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Folded Proteins.

In order to evaluate the stability of folded proteins when modeled with *ff03CMAP*, three representative folded proteins were simulated: CspTm (all-β), ubiquitin (α/β) and SPR17 (all-α). The initial structures are extracted from PDB and the simulation time is 1 μs for each system.

Table 2 shows the FF scores of the three tested proteins for all *ff03*-series force fields. It is obviously that the FF score for combination *ff03CMAP/TIP4P-Ew* is the best among all tested force fields and the value close to 1. This suggests that *ff03CMAP/TIP4PEw* indeed can be used to simulate folded proteins. And we also found that the original *ff03* force field performs better performance than other revised *ff03* force fields. It is no surprise that the performance of *ff03CMAP/TIP4PD* is a little worse than that of *ff03CMAP/TIP4PEw* because TIP4P-D water model would destabilize the folded states of proteins as reported.[16] The details of the FF score composition for the three proteins are shown in Tables S4–S6.

To quantify the fluctuation in simulations, Cα RMSD's of the folded proteins are shown in Figure 3 [Can you only compute the RMSDs of the secondary structures/i.e. stable portions of the proteins?]. For CspTm and ubiquitin, the RMSDs in the *ff03CMAP/TIP4PEw* simulations are small and stable, which is consistent with the FF score. However, the RMSD's for SPR17 rise over 4Å after 700ns in the *ff03CMAP/TIP4PEw* simulation. The RMSDs of *ff03ws* for all three folded proteins fluctuate quite significantly, implying less stable folded states for the tested proteins.

More detailed analyses were conducted for ubiquitin to compare the performances of these force fields. Secondary chemical shifts and backbone scalar coupling constants of ubiquitin (Figures S6–S7) suggest that *ff03CMAP/TIP4PEw* simulation agree the best with experimental data. The same can be said for the side-chain scalar coupling constants as shown in Tables S16–S17. RDCs of backbone N-HN, Cα-Hα, Cα-C, C-N and C-HN were also calculated shown in Figure S8. Similar to chemical shifts and scalar coupling constants, the performance of *ff03CMAP/TIP4PEw* simulation also agrees among the best, along with the *ff03* simulation, while the *ff03ws* simulation agrees the worse. The order parameters for backbone amide and side-chain methyl groups are shown in Figure 4 and Table S18, respectively. The order parameters in the *ff03CMAP/TIP4PEw* simulation are also in good agreement with experimental data. The figure further indicates that the order parameters of the loop regions in the *ff03CMAP/TIP4PD* simulation are much lower than those from other force fields. [What is the point of the following statement?]Only *ff03CMAP/TIP4PEw* exhibited similar behavior to experiment for side-chain order parameters (Table S18?).

The RMSDs of secondary chemical shift, J-coupling, order parameters, and RDCs are gathered in Table 3. The summary indicates that *ff03CMAP/TIP4PEw* performs excellently in in reproducing all available experimental measurements and its FF score is very close to 1. While the FF score of *ff03CMAP/TIP4PD* is more than 1.5, which suggests TIP4P-D water model unsuitable for the simulation of folded proteins.

In order to further evaluate the stability of *ff03CMAP* for folded proteins, the dominant conformers of ubiquitin from six *ff03*-series force fields are retrieved and shown in Figure 5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

It was found that top 3 clusters in the *ff03* simulation occupy 100.00% of the snapshots. All the conformers include high percentage of helical structures. Top 5 clusters in the *ff03** simulation also occupy 100.00% of the snapshots with partially non-helical structures. In the *ff03w* simulation, only 2 clusters were found, and the conformers are highly structured. In the *ff03ws* simulation, top 8 clusters only occupy 78.70% of the snapshots. In the *ff03CMAP/TIP4PEw* simulation there is only one cluster and 4 clusters in the *ff03CMAP/TIP4PD* simulation. Additional conformation clustering was also conducted for the CspTm and SPR17 simulations (supplementary Figures S24–S25). These conformer clusters indicate that *ff03CMAP/TIP4PEw*, *ff03*, and *ff03w* may be the better *ff03* choices for folded proteins MD simulation.

Intrinsically Disordered Proteins.

We tested 9 typical disordered proteins with 19 to 124 residues. The FF scores of the six *ff03*-series force fields are listed in Table 4. Except for HEWL19 and HIVRev, the FF scores of *ff03CMAP/TIP4P-D* simulations are the lowest, and most of them are very close to 1, indicating very good agreement with experiment. Although the FF scores of *ff03CMAP/TIP4PD* for HEWL19 and HIVRev are not the best, the differences with the best performing force fields are not significant. This suggests that *ff03CMAP/TIP4P-D* can reproduce well conformers of tested IDPs. The details of the FF score composition for each tested system are shown in Tables S7–S15.

It is interesting to note that all *ff03* revisions improve over the original *ff03* in IDP simulations, as they are all designed to reproduce the properties of IDPs. As expected, the CMAP method can provide accurate descriptions of IDPs as in previous developments.[9–11, 15, 72] In addition, the TIP4P-D water model is demonstrated again to be suitable for IDP simulations.[16] In order to understand the influence of solvent models, three IDPs were simulated with *ff03/TIP4P-D* (Tables S10, S12–S13). The results show that TIP4P-D water model indeed partly improves the performance of tested IDPs conformers. However, the results are still much worse than those of *ff03CMAP/TIP4P-D*, suggesting that the CMAP improvement in *ff03CMAP/TIP4P-D* simulations plays a key role in reproducing the IDPs conformers.

To further illustrate the property of *ff03CMAP* force fields, we calculated the average RMSDs for different experimental measurements of IDPs (Table 5). We found that the performance of *ff03CMAP/TIP4PD* is the best for all experimental observables, and *ff03CMAP/TIP4PEw* also performs reasonably well. It is noticeable that *ff03CMAP* significantly improved the quality of simulated Cα and N secondary chemical shifts and $^{3}J_{HNH\alpha}$ scalar coupling constant, which have are closely related to backbone dihedrals. It is clear that the CMAP method can be used to correct the dihedral distributions and the *TIP4P-D* water model further refines the interactions between protein and water, leading to excellent observed performance in the *ff03CMAP/TIP4PEw* simulations for IDPs.

For IDP simulations, underestimation of Rg is a common limitation for generic protein force fields.[14, 16] We calculated Rg distributions for all tested IDPs (Figure S23). The force field with four-site water models [Which one, be specific] can sample a wider range of Rg distributions and larger Rg mean values, especially for *ff03CMAP/TIP4P-D* and *ff03ws*.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Conformers are more compact in force fields in the TIP3P water model such as *ff03* and *ff03**. We compared the experimental Rg and simulated Rg of three IDPs. The analysis indicates that *ff03CMAP* force field and *ff03ws* overestimate the Rg of Aβ40 and RS. The average RMSDs[What is this RMSD? For a single Rg value?] of *ff03CMAP/TIP4PEw* are also very small and close to experimental value. For ACTR, only the Rg in the *ff03CMAP/TIP4PD* simulation is located within the experimental range and other force fields significantly underestimate it.

Besides the above overall assessments, we next use a classical example of IDPs, RS, in the following discussion to illustrate the performance of these force fields. To evaluate the backbone and side-chain sampling for RS, we compared secondary chemical shifts and scalar coupling constants. Figure 6 shows the secondary chemical shifts and backbone scalar coupling constants for six *ff03*-series force fields. Tables S19–S20 list detailed data used in analysis. The chemical shifts and scalar coupling constants calculated from all revised *ff03* force fields are much closer to experimental data than those from the original *ff03* force field, and the RMSDs between simulated and experimental values from *ff03CMAP/TIP4P-Ew* and *ff03CMAP/TIP4P-D* combinations are smaller than those from other *ff03*-series force fields (Table 4). We also calculated the backbone N-HN, Cα-Hα and Cα-C RDCs with a local alignment window of 15 residues (Figure 7), whose performance is similar to that of chemical shifts and coupling constants, with both *ff03CMAP/TIP4PEw* and *ff03CMAP/TIP4PD* giving lower RDC(Q, what does Q mean? Why lower value is better) with smaller standard deviations (why smaller sd is better?) than other tested force fields.

Next FF scores were used to compare the performance of all force fields (listed in Table 6) in RS simulations. The RMSDs of the original *ff03* force field are the largest and FF score is larger than 4. *ff03CMAP/TIP4PD* combination gives the best agreement with all experimental measurements and the FF scores are around 1.1.

Finally we computed the ensemble-averaged SAXS curves for RS and fitted with the experimental curve (Figure 8). The χ^2 value was used to evaluate the quality of the fitted result to the given experimental SAXS profile as shown in the literature.[73] Our analysis shows that the χ^2 of *ff03CMAP/TIP4PEw* is the smallest among 6 tested *ff03*-series force fields. This suggests that *ff03CMAP/TIP4PEw* can reproduce the SAXS property for RS, while the *ff03CMAP/TIP4PD* combination leads to conformers that are too expansive.

To further illustrate the conformer sampling efficiency, kClust was used to cluster conformers according to φ angle and Cα RMSD. Representative conformers and their occupations are shown in Figures S26–S34. The results indicate that both *ff03CMAP* and *ff03ws* can sample more flexible and diverse disordered conformers, while the representative conformers in the *ff03* simulation contain several short helixes with tight packing. Convergence of conformer sampling is another important issue for IDP simulations. We used biphasic decay model to evaluate the convergence time scales for IDP simulations. It is interesting to note that *ff03CMAP* simulations have smaller decay half times, which suggests *ff03CMAP* simulations converge earlier than other *ff03*-series simulations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Ab Initio Folding of Fast-Folding Proteins.

We performed REMD for 3 typical fast-folding proteins, such as 16-residue two β-sheets GB1, small β-hairpin-forming protein CLN025, and helical 15-mer AAQAA3.[34, 36–37] The melting curves in the *ff03CMAP/TIP4PEw* simulations are shown in Figure 9. The melting curves show that GB1 and CLN025 can be ab initio folded when modeled with *ff03CMAP/TIP4PEw*. However, a few folded structures were observed in the REMD simulation of AAQAA3. To study whether the ab initio folding of helical structures can be improved by modifying CMAP parameters, we updated a new set of CMAP parameters by only decreasing the parameters in α_h region with the revised force field termed as *ff03CMAP2*. Our REMD simulation shows that *ff03CMAP2/TIP4PEw* performs significantly better in helix folding. In the meantime, *ff03CMAP2/TIP4PEw* can maintain almost same melting curves for sheet and hairpin fast-folding proteins. [You should say whether the IDP performance can still be maintained.]

Conclusion

The backbone dihedral term for all 20 amino acids was optimized to improve the performance of the current force field. TIP4P-Ew and TIP4P-D are combined with newly developed force field *ff03CMAP* to simulate different type proteins. Extensive tests of typical short peptide, folded proteins, disordered proteins, and fast-folding proteins show that the simulated chemical shifts, J-coupling, order parameters, and RDC with the *ff03CMAP* force fields are in quantitative agreement with those from NMR experiment and are more accurate than other *ff03*-series force fields. The influences of solvent models were also investigated. The results indicate that *ff03CMAP/TIP4PEw* for folded proteins and *ff03CMAP* combined with TIP4P-D was suitable for disordered proteins (*ff03CMAP/TIP4PEw* also shows good performance in IDPs). Therefore, these findings confirm that the newly developed force field *ff03CMAP* can improve the balance and efficiency of conformer sampling between intrinsically folded proteins and disordered proteins. Although *ff03CMAP* force field has the limitation of folding helix structures, this can be improved by adjusting CMAP parameters which is the *ff03CMAP2*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

This work was supported by Center for HPC at Shanghai Jiao Tong University, the National Natural Science Foundation of China (31770771, 21977068, and 31620103901), the National Key Research and Development Program of China (2018YFC0310803 and 2017YFE0103300), Medical Engineering Cross Fund of Shanghai Jiao Tong University (YG2017MS08). and National Institutes of Health/NIGMS (GM093040 & GM079383).

References

[1]. Dunker AK; Lawson JD; Brown CJ; Williams RM; Romero P; Oh JS; Oldfield CJ; Campen AM; Ratliff CM; Hipps KW; Ausio J; Nissen MS; Reeves R; Kang C; Kissinger CR; Bailey RW; Griswold MD; Chiu W; Garner EC; Obradovic Z, Intrinsically disordered protein. Journal of molecular graphics & modelling 2001, 19 (1), 26–59. [PubMed: 11381529]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[2]. Dunker AK; Brown CJ; Lawson JD; Iakoucheva LM; Obradovic Z, Intrinsic disorder and protein function. Biochemistry 2002, 41 (21), 6573–82. [PubMed: 12022860]

[3]. Babu MM, The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. Biochemical Society transactions 2016, 44 (5), 1185–1200. [PubMed: 27911701]

[4]. Cheng Y; LeGall T; Oldfield CJ; Dunker AK; Uversky VN, Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry 2006, 45 (35), 10448–60. [PubMed: 16939197]

[5]. Oldfield CJ; Dunker AK, Intrinsically disordered proteins and intrinsically disordered protein regions. Annual review of biochemistry 2014, 83, 553–84.

[6]. Bernado P; Bertoncini CW; Griesinger C; Zweckstetter M; Blackledge M, Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. Journal of the American Chemical Society 2005, 127 (51), 17968–9. [PubMed: 16366524]

[7]. Bernado P; Mylonas E; Petoukhov MV; Blackledge M; Svergun DI, Structural characterization of flexible proteins using small-angle X-ray scattering. Journal of the American Chemical Society 2007, 129 (17), 5656–64. [PubMed: 17411046]

[8]. Schneider R; Huang JR; Yao M; Communie G; Ozenne V; Mollica L; Salmon L; Jensen MR; Blackledge M, Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. Molecular bioSystems 2012, 8 (1), 58–68. [PubMed: 21874206]

[9]. Ye W; Ji D; Wang W; Luo R; Chen HF, Test and Evaluation of ff99IDPs Force Field for Intrinsically Disordered Proteins. J Chem Inf Model 2015, 55 (5), 1021–9. [PubMed: 25919886]

[10]. Song D; Wang W; Ye W; Ji D; Luo R; Chen HF, ff14IDPs force field improving the conformation sampling of intrinsically disordered proteins. Chem Biol Drug Des 2017, 89 (1), 5–15. [PubMed: 27484738]

[11]. Song D; Luo R; Chen HF, The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. J Chem Inf Model 2017, 57 (5), 1166–1178. [PubMed: 28448138]

[12]. Best RB; Zheng W; Mittal J, Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. Journal of chemical theory and computation 2014, 10 (11), 5113–5124. [PubMed: 25400522]

[13]. Zhou CY; Jiang F; Wu YD, Residue-specific force field based on protein coil library. RSFF2: modification of AMBER ff99SB. The journal of physical chemistry. B 2015, 119 (3), 1035–47. [PubMed: 25358113]

[14]. 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. [PubMed: 29735687]

[15]. Liu H; Song D; Lu H; Luo R; Chen HF, Intrinsically disordered protein-specific force field CHARMM36IDPSFF. Chem Biol Drug Des 2018, 92 (4), 1722–1735. [PubMed: 29808548]

[16]. Piana S; Donchev AG; Robustelli P; Shaw DE, Water dispersion interactions strongly influence simulated structural properties of disordered protein states. The journal of physical chemistry. B 2015, 119 (16), 5113–23. [PubMed: 25764013]

[17]. Duan Y; Wu C; Chowdhury S; Lee MC; Xiong G; Zhang W; Yang R; Cieplak P; Luo R; Lee T; Caldwell J; Wang J; Kollman P, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 2003, 24 (16), 1999–2012. [PubMed: 14531054]

[18]. Best RB; Hummer G, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. The journal of physical chemistry. B 2009, 113 (26), 9004–15. [PubMed: 19514729]

[19]. Best RB; Mittal J, Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. The journal of physical chemistry. B 2010, 114 (46), 14916–23. [PubMed: 21038907]

[20]. Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML, Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics 1983, 79 (2), 926–935.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[21]. Horn HW; Swope WC; Pitera JW; Madura JD; Dick TJ; Hura GL; Head-Gordon T, Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys 2004, 120 (20), 9665–78. [PubMed: 15267980]

[22]. Abascal JL; Vega C, A general purpose model for the condensed phases of water: TIP4P/2005. J Chem Phys 2005, 123 (23), 234505. [PubMed: 16392929]

[23]. Welker C; Bohm G; Schurig H; Jaenicke R, Cloning, overexpression, purification, and physicochemical characterization of a cold shock protein homolog from the hyperthermophilic bacterium Thermotoga maritima. Protein Sci 1999, 8 (2), 394–403. [PubMed: 10048332]

[24]. Vijay-Kumar S; Bugg CE; Cook WJ, Structure of ubiquitin refined at 1.8 A resolution. J Mol Biol 1987, 194 (3), 531–44. [PubMed: 3041007]

[25]. Bennett V; Gilligan DM, The spectrin-based membrane skeleton and micron-scale organization of the plasma membrane. Annu Rev Cell Biol 1993, 9, 27–66. [PubMed: 8280463]

[26]. Graf J; Nguyen PH; Stock G; Schwalbe H, Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. Journal of the American Chemical Society 2007, 129 (5), 1179–89. [PubMed: 17263399]

[27]. Xiang S; Gapsys V; Kim HY; Bessonov S; Hsiao HH; Mohlmann S; Klaukien V; Ficner R; Becker S; Urlaub H; Luhrmann R; de Groot B; Zweckstetter M, Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. Structure 2013, 21 (12), 2162–74. [PubMed: 24183573]

[28]. Tan R; Chen L; Buettner JA; Hudson D; Frankel AD, RNA recognition by an isolated α helix. Cell 1993, 73 (5), 1031–1040. [PubMed: 7684657]

[29]. Sgourakis NG; Yan Y; McCallum SA; Wang C; Garcia AE, The Alzheimer's peptides Abeta40 and 42 adopt distinct conformations in water: a combined MD / NMR study. J Mol Biol 2007, 368 (5), 1448–57. [PubMed: 17397862]

[30]. Iesmantavicius V; Jensen MR; Ozenne V; Blackledge M; Poulsen FM; Kjaergaard M, Modulation of the intrinsic helix propensity of an intrinsically disordered protein reveals long-range helix-helix interactions. Journal of the American Chemical Society 2013, 135 (27), 10155–63. [PubMed: 23758617]

[31]. Li M; Phylip LH; Lees WE; Winther JR; Dunn BM; Wlodawer A; Kay J; Gustchina A, The aspartic proteinase from Saccharomyces cerevisiae folds its own inhibitor into a helix. Nat Struct Biol 2000, 7 (2), 113–7. [PubMed: 10655612]

[32]. Wells M; Tidow H; Rutherford TJ; Markwick P; Jensen MR; Mylonas E; Svergun DI; Blackledge M; Fersht AR, Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A 2008, 105 (15), 5762–7. [PubMed: 18391200]

[33]. Sillen A; Barbier P; Landrieu I; Lefebvre S; Wieruszeski JM; Leroy A; Peyrot V; Lippens G, NMR investigation of the interaction between the neuronal protein tau and the microtubules. Biochemistry 2007, 46 (11), 3055–64. [PubMed: 17311412]

[34]. Shalongo W; Dugad L; Stellwagen E, Distribution of helicity within the model peptide acetyl (AAQAA) 3amide. Journal of the American Chemical Society 1994, 116 (18), 8288–8293.

[35]. Blanco FJ; Rivas G; Serrano L, A short linear peptide that folds into a native stable β-hairpin in aqueous solution. Nature structural biology 1994, 1 (9), 584. [PubMed: 7634098]

[36]. Munoz V; Thompson PA; Hofrichter J; Eaton WA, Folding dynamics and mechanism of β-hairpin formation. Nature 1997, 390 (6656), 196. [PubMed: 9367160]

[37]. Honda S; Akiba T; Kato YS; Sawada Y; Sekijima M; Ishimura M; Ooishi A; Watanabe H; Odahara T; Harata K, Crystal structure of a ten-amino acid protein. Journal of the American Chemical Society 2008, 130 (46), 15327–31. [PubMed: 18950166]

[38]. Case DA; Babin V; Berryman J; Betz R; Cai Q; Cerutti D; Cheatham Iii T; Darden T; Duke R; Gohlke H, Amber 14 2014.

[39]. Lee TS; Cerutti DS; Mermelstein D; Lin C; LeGrand S; Giese TJ; Roitberg A; Case DA; Walker RC; York DM, GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. J Chem Inf Model 2018, 58 (10), 2043–2050. [PubMed: 30199633]

[40]. Patriksson A; van der Spoel D, A temperature predictor for parallel tempering simulations. Phys Chem Chem Phys 2008, 10 (15), 2073–7. [PubMed: 18688361]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[41]. Darden T; York D; Pedersen L, Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. The Journal of chemical physics 1993, 98 (12), 10089–10092.

[42]. Salomon-Ferrer R; Gotz AW; Poole D; Le Grand S; Walker RC, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. Journal of chemical theory and computation 2013, 9 (9), 3878–88. [PubMed: 26592383]

[43]. Ryckaert J-P; Ciccotti G; Berendsen HJ, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of computational physics 1977, 23 (3), 327–341.

[44]. Kabsch W; Sander C, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983, 22 (12), 2577–637. [PubMed: 6667333]

[45]. Joosten RP; te Beek TA; Krieger E; Hekkelman ML; Hooft RW; Schneider R; Sander C; Vriend G, A series of PDB related databases for everyday needs. Nucleic Acids Res 2011, 39 (Database issue), D411–9. [PubMed: 21071423]

[46]. Wang W; Ye W; Jiang C; Luo R; Chen HF, New force field on modeling intrinsically disordered proteins. Chem Biol Drug Des 2014, 84 (3), 253–69. [PubMed: 24589355]

[47]. MacKerell AD; Bashford D; Bellott M; Dunbrack RL; Evanseck JD; Field MJ; Fischer S; Gao J; Guo H; Ha S; Joseph-McCarthy D; Kuchnir L; Kuczera K; Lau FT; Mattos C; Michnick S; Ngo T; Nguyen DT; Prodhom B; Reiher WE; Roux B; Schlenkrich M; Smith JC; Stote R; Straub J; Watanabe M; Wiorkiewicz-Kuczera J; Yin D; Karplus M, All-atom empirical potential for molecular modeling and dynamics studies of proteins. The journal of physical chemistry. B 1998, 102 (18), 3586–616. [PubMed: 24889800]

[48]. MacKerell AD Jr.; Feig M; Brooks CL 3rd, Improved treatment of the protein backbone in empirical force fields. Journal of the American Chemical Society 2004, 126 (3), 698–9. [PubMed: 14733527]

[49]. Mackerell AD Jr.; Feig M; Brooks CL 3rd, 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 2004, 25 (11), 1400–15. [PubMed: 15185334]

[50]. Huang J; Rauscher S; Nawrocki G; Ran T; Feig M; de Groot BL; Grubmuller H; MacKerell AD Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 2017, 14 (1), 71–73. [PubMed: 27819658]

[51]. Han B; Liu Y; Ginzinger SW; Wishart DS, SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 2011, 50 (1), 43–57. [PubMed: 21448735]

[52]. Vogeli B; Ying J; Grishaev A; Bax A, Limits on variations in protein backbone dynamics from precise measurements of scalar couplings. Journal of the American Chemical Society 2007, 129 (30), 9377–85. [PubMed: 17608477]

[53]. Hu J-S; Bax A, Determination of ϕ and $\chi 1$ Angles in Proteins from 13C– 13C Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond? Journal of the American Chemical Society 1997, 119 (27), 6360–6368.

[54]. Hennig M; Bermel W; Schwalbe H; Griesinger C, Determination of ψ torsion angle restraints from 3 J (Cα., Cα) and 3 J (Cα., HN) coupling constants in proteins. Journal of the American Chemical Society 2000, 122 (26), 6268–6277.

[55]. Lee JH; Li F; Grishaev A; Bax A, Quantitative residue-specific protein backbone torsion angle dynamics from concerted measurement of 3J couplings. Journal of the American Chemical Society 2015, 137 (4), 1432–5. [PubMed: 25590347]

[56]. Ding K; Gronenborn AM, Protein Backbone 1H(N)-13Calpha and 15N-13Calpha residual dipolar and J couplings: new constraints for NMR structure determination. Journal of the American Chemical Society 2004, 126 (20), 6232–3. [PubMed: 15149211]

[57]. Wirmer J; Schwalbe H, Angular dependence of 1J(Ni,Calphai) and 2J(Ni,Calpha(i-1)) coupling constants measured in J-modulated HSQCs. J Biomol NMR 2002, 23 (1), 47–55. [PubMed: 12061717]

[58]. Mantsyzov AB; Maltsev AS; Ying J; Shen Y; Hummer G; Bax A, A maximum entropy approach to the study of residue-specific backbone angle distributions in alpha-synuclein, an intrinsically disordered protein. Protein Sci 2014, 23 (9), 1275–90. [PubMed: 24976112]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[59]. Vuister GW; Delaglio F; Bax A, The use of 1 J CαHα coupling constants as a probe for protein backbone conformation. Journal of biomolecular NMR 1993, 3 (1), 67–80. [PubMed: 8448436]

[60]. Schmidt JM; Howard MJ; Maestre-Martinez M; Perez CS; Lohr F, Variation in protein C(alpha)-related one-bond J couplings. Magn Reson Chem 2009, 47 (1), 16–30. [PubMed: 18853398]

[61]. Schmidt JM, Asymmetric Karplus curves for the protein side-chain 3J couplings. J Biomol NMR 2007, 37 (4), 287–301. [PubMed: 17333486]

[62]. Zweckstetter M, NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 2008, 3 (4), 679–90. [PubMed: 18388951]

[63]. Marsh JA; Baker JM; Tollinger M; Forman-Kay JD, Calculation of residual dipolar couplings from disordered state ensembles using local alignment. Journal of the American Chemical Society 2008, 130 (25), 7804–5. [PubMed: 18512919]

[64]. Trbovic N; Kim B; Friesner RA; Palmer AG 3rd, Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation. Proteins 2008, 71 (2), 684–94. [PubMed: 17975832]

[65]. Schneidman-Duhovny D; Hammel M; Tainer JA; Sali A, Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys J 2013, 105 (4), 962–74. [PubMed: 23972848]

[66]. Feig M; Karanicolas J; Brooks CL 3rd, MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. Journal of molecular graphics & modelling 2004, 22 (5), 377–95. [PubMed: 15099834]

[67]. McGibbon RT; Beauchamp KA; Harrigan MP; Klein C; Swails JM; Hernandez CX; Schwantes CR; Wang LP; Lane TJ; Pande VS, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys J 2015, 109 (8), 1528–32. [PubMed: 26488642]

[68]. Schrodinger LLC, The PyMOL Molecular Graphics System, Version 1.8. 2015.

[69]. Cerf O, Tailing of survival curves of bacterial spores. J Appl Bacteriol 1977, 42 (1), 1–19. [PubMed: 323208]

[70]. Brouwer AF; Eisenberg MC; Remais JV; Collender PA; Meza R; Eisenberg JN, Modeling Biphasic Environmental Decay of Pathogens and Implications for Risk Analysis. Environ Sci Technol 2017, 51 (4), 2186–2196. [PubMed: 28112914]

[71]. Duong VT; Chen Z; Thapa MT; Luo R, Computational Studies of Intrinsically Disordered Proteins. The journal of physical chemistry. B 2018, 122 (46), 10455–10469. [PubMed: 30372613]

[72]. Guo X; Han J; Luo R; Chen HF, Conformation Dynamics of the Intrinsically Disordered Protein c-Myb with the ff99IDPs Force Field. RSC Adv 2017, 7 (47), 29713–29721. [PubMed: 29104751]

[73]. Schneidman-Duhovny D; Hammel M; Sali A, FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic acids research 2010, 38 (suppl_2), W540–W544. [PubMed: 20507903]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

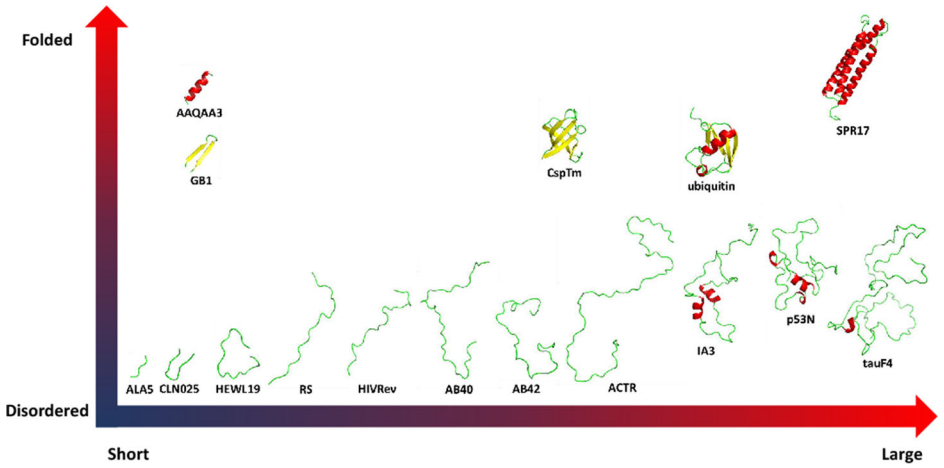

Figure 1.

Test system from short peptides to proteins. Three representative folded proteins such as cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima* (CspTm, all-β),[23] ubiquitin of human (α/β)[24] and chicken brain alpha spectrin repeat 17 (SPR17, all-α).[25] 9 typical disordered protein including 19 length peptide of hen egg-white lysozyme (HEWL19),[26] phosphorylated SRSF1 (RS),[27] HIV-1 Rev ARM peptide (HIVRev),[28] 40 length amyloid-beta-peptides (AB40),[29] 42 length amyloid-beta-peptides (AB42),[29] activation domain of the nuclear hormone receptor coactivator (ACTR),[30] an aspartic proteinase inhibitor for *Saccharomyces cerevisiae* (IA3),[31] p53 N-terminal transactivation domain (p53N)[32] and tau protein fragment (TauF4).[33] 3 fast-folding proteins including 15-residue helix-forming peptide Ac-$(AAQAA)_3$-NH_2 (AAQAA3),[34] β-hairpin B1 domain of protein G (GB1)[35–36] and Chignolin, a 10 residue folded peptide designed by segment statistics (CLN025).[37]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

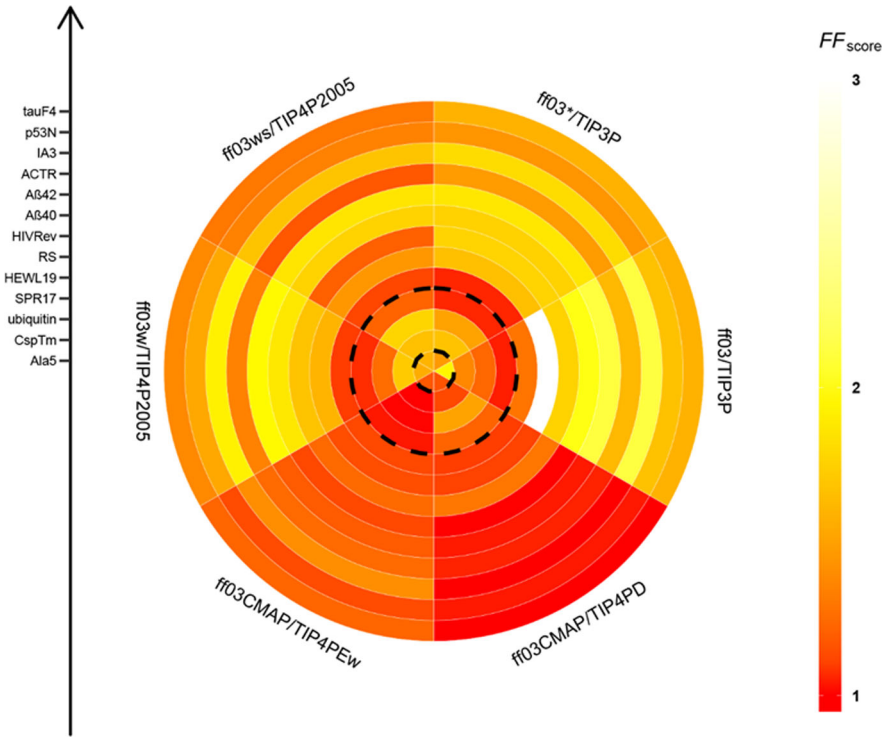

Figure 2.

FF scores for short peptide, disordered proteins, and folded proteins. The plot is divided into six sectors which represent six *ff03*-series force fields. From inner circle to outer circle, there are 13 simulated system for 1 short peptide, 9 disordered proteins, and 3 folded proteins sorted by the residue length in each category.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

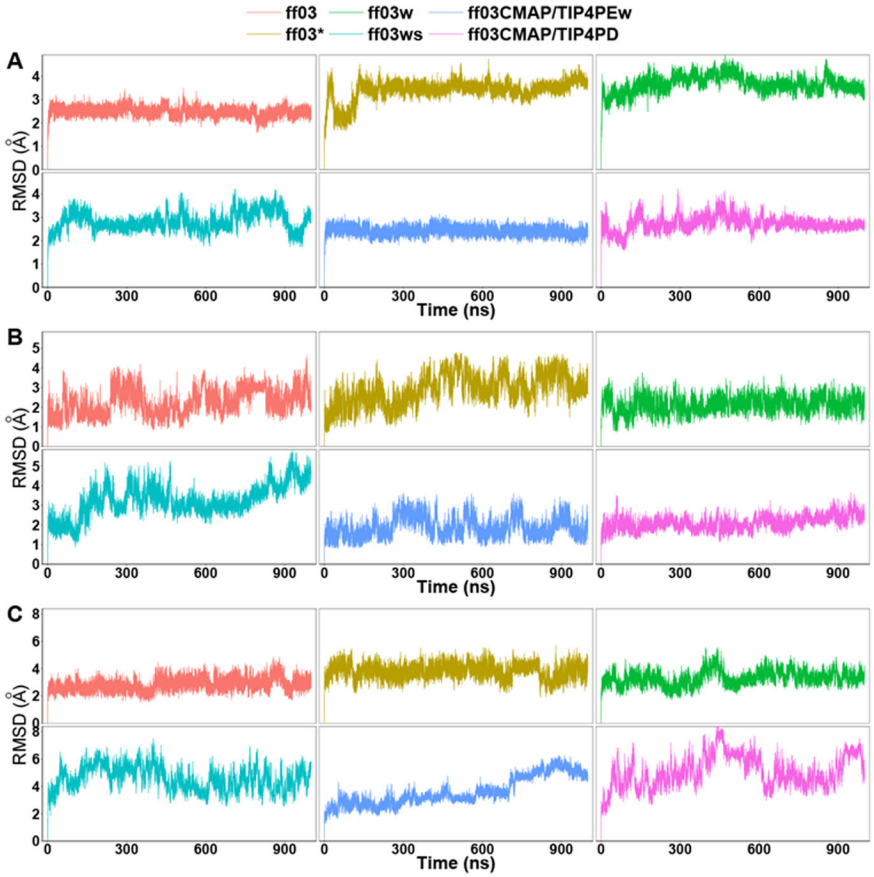

Figure 3.

Cα RMSD of six *ff03*-series force fields. A: CspTm. B: ubiquitin. C: SPR17.[See my comments]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

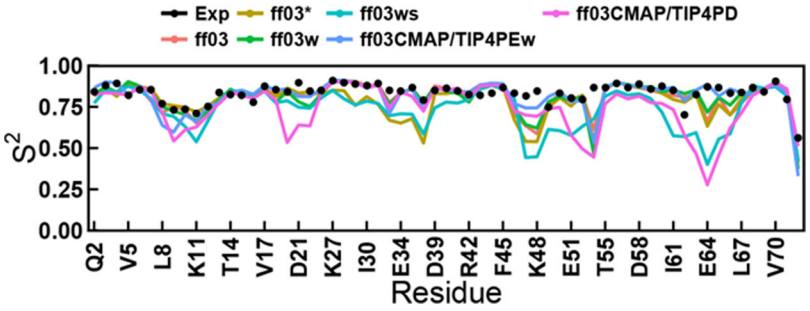

Figure 4.

Backbone amide order parameters of six *ff03*-series force fields for ubiquitin. The RMSDs between simulation and experiment for *ff03*, *ff03**, *ff03w*, *ff03ws*, *ff03CMAP/TIP4PEw* and *ff03CMAP/TIP4PD* are 0.074, 0.104, 0.081, 0.149, 0.069 and 0.154.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

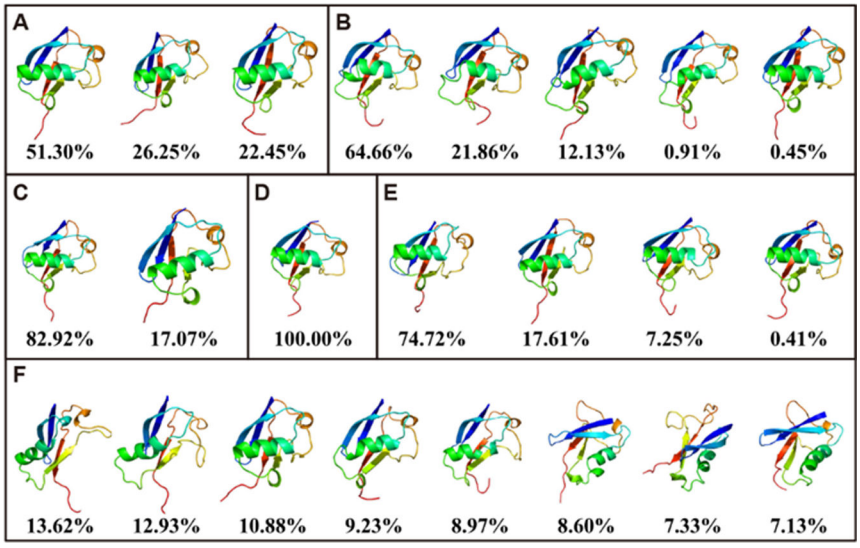

Figure 5.

Conformation clustering of simulation for ubiquitin. The figure shows the top 8 clusters at most with dominant conformations and percentage for *ff03* (A), *ff03** (B), *ff03w* (C), *ff03CMAP/TIP4PEw* (D), *ff03CMAP/TIP4PD* (E), and *ff03ws* (F).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

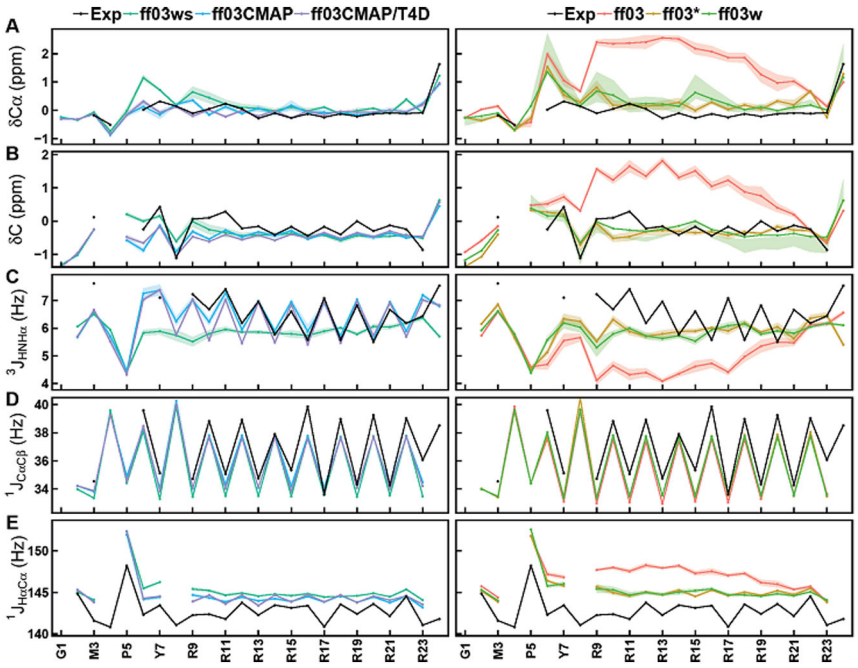

Figure 6.

Secondary chemical shifts and backbone scalar coupling constants of simulation and experimental data for RS. Simulated and experimental secondary chemical shifts for (A) Cα and (B) C, backbone scalar coupling constants for (C) $^3J_{HNH\alpha}$, (D) $^1J_{C\alpha C\beta}$ and (E) $^1J_{H\alpha C\alpha}$. Simulated values are shown for *ff03* (red), *ff03** (brown), *ff03w* (light green), *ff03ws* (cyan), *ff03CMAP* (blue), and *ff03CMAP/T4D* (violet). Experimental values are displayed as black lines. The shadow means the stand error of mean.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

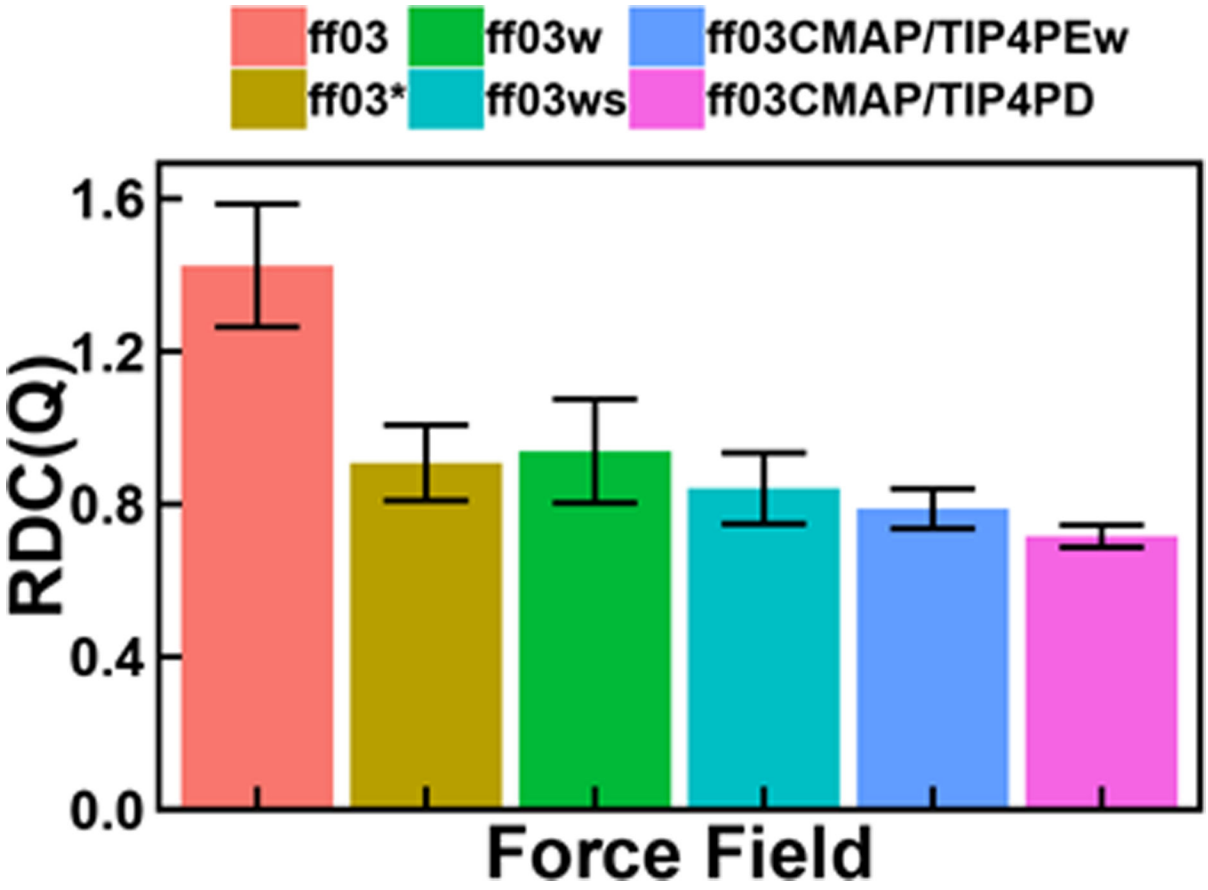

Figure 7.

Backbone RDCs of simulation with six *ff03*-series force fields for RS. Error bar means the standard deviations of calculation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

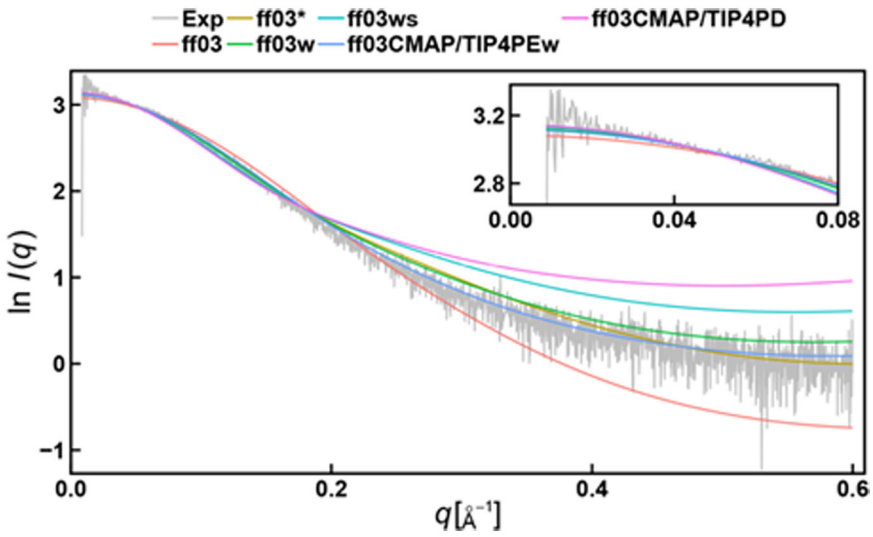

Figure 8.

SAXS profiles of simulation and experimental data for RS. Experimental values are displayed as grey drawing. The χ^2 of *ff03*, *ff03**, *ff03w*, *ff03ws*, *ff03CMAP/TIP4PEw* and *ff03CMAP/TIP4PD* are 5.118, 2.217, 2.418, 10.708, 1.319 and 23.658.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

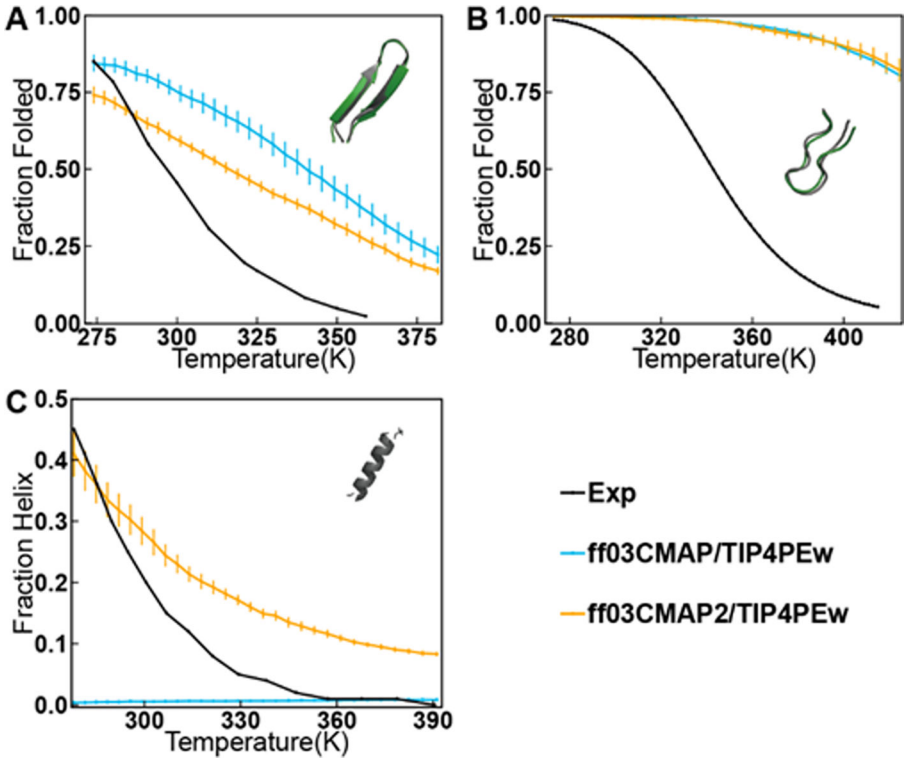

Figure 9.

Melting curves for *ff03CMAP/TIP4PEw* and *ff03CMAP2/TIP4PEw* with experimental values. (A) Temperature-dependent folded population of GB1. (B) Temperature-dependent folded population of CLN025. (C) Temperature-dependent helix formation of AAQAA3 (The definition of helicity calculation is mention in SI). Simulated values are shown for *ff03CMAP/TIP4PEw* are in blue lines, *ff03CMAP2/TIP4PEw* are in orange line, and experimental melting curves are shown in black lines. All simulations initiated from fully unfolded structures and 100 ns (for CLN025 and AAQAA3) and 300 ns (for GB1) for equilibration. Folded structures obtained from REMD simulations (gray) starting from completely unfolded state are compared with the native structure from the PDB database (green), and the RMSD between folded structure and native structure is 0.728Å for GB1 and 0.993Å for CLN025.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

RMSD's of secondary chemical shifts, J-coupling constants and force field scores of *ff03*-series force fields for Ala5.[a]

	ff03	*ff03**	*ff03w*	*ff03ws*	*ff03CMAP/ TIP4PEw*	*ff03CMAP/TIP4PD*
Cα	0.321	0.182	0.192	0.173	0.162	0.183
Cβ	0.372	0.474	0.499	0.514	0.192	0.204
C	0.297	0.453	0.471	0.490	0.346	0.348
N	1.633	1.022	1.112	1.177	0.417	0.415
Hα	0.045	0.057	0.055	0.057	0.024	0.021
HN	0.352	0.375	0.381	0.384	0.346	0.348
$^{3}J_{HNH\alpha}$	0.443	0.292	0.281	0.297	0.319	0.259
$^{3}J_{HNC\alpha}$	0.104	0.047	0.046	0.047	0.047	0.042
$^{3}J_{HNC\beta}$	0.374	0.234	0.264	0.275	0.359	0.296
$^{3}J_{H\alpha C}$	0.120	0.131	0.135	0.137	0.124	0.154
$^{3}J_{HNC}$	0.224	0.133	0.168	0.179	0.188	0.160
$^{1}J_{C\alpha N}$	0.265	0.092	0.141	0.136	0.166	0.179
$^{2}J_{C\alpha N}$	0.435	0.217	0.222	0.217	0.222	0.228
CS_{score}	1.995	1.888	1.950	1.997	1.049	1.061
NMR_{score}	1.905	1.047	1.178	1.202	1.307	1.250
FF_{score}	**1.950**	**1.468**	**1.564**	**1.599**	**1.178**	**1.155**

[a]Chemical shifts are in ppm, J-coupling constants are in Hz and the scores are unitless.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

FF scores of three folded proteins for six *ff03*-series force fields

Protein	*ff03*	*ff03**	*ff03w*	*ff03ws*	*ff03CMAP/TIP4PEw*	*ff03CMAP/TIP4PD*
CspTm	1.376[a]	1.674	1.718	1.663	1.000	1.123
Ubiquitin	1.231	1.506	1.249	1.769	1.003	1.509
SPR17	1.026	1.041	1.052	1.219	1.019	1.220

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3.

RMSDs of secondary chemical shifts, J-coupling constants, RDCs, S^2 parameters and FF scores of ubiquitin for six *ff03*-series force fields.[a]

	ff03	*ff03**	*ff03w*	*ff03ws*	*ff03CMAP/TIP4PEw*	*ff03CMAP/TIP4PD*
Cα	0.481	0.716	0.507	0.723	0.485	0.575
Cβ	0.816	0.885	0.825	1.069	0.770	0.860
C	0.607	0.702	0.611	0.789	0.609	0.650
N	2.050	2.450	2.188	2.625	2.142	2.573
HA	0.145	0.185	0.141	0.234	0.106	0.165
HN	0.329	0.349	0.326	0.407	0.298	0.293
$^3J_{HNH\alpha}$	1.341	1.487	1.484	1.597	1.196	1.319
$^3J_{H\alpha C}$	0.684	0.725	0.623	0.986	0.410	0.609
$^3J_{HNC\beta}$	0.643	0.734	0.671	0.701	0.633	0.612
$^3J_{HNC}$	0.681	0.734	0.701	0.852	0.564	0.662
$^1J_{C\alpha N}$	0.465	0.567	0.494	0.588	0.415	0.469
$^1J_{H\alpha C\alpha}$	2.669	2.717	2.628	3.030	2.225	2.476
$^1J_{C\alpha C\beta}$	0.984	1.056	0.977	1.229	0.856	0.966
$^2J_{C\alpha N}$	0.464	0.555	0.437	0.682	0.361	0.477
$^2J_{CC\gamma}$	0.650	0.973	0.717	0.936	0.419	0.775
$^3J_{NC\gamma}$	0.490	0.592	0.501	0.557	0.428	0.496
S^2_{NH}	0.074	0.104	0.081	0.149	0.069	0.154
S^2_{axis}	0.279	0.260	0.247	0.289	0.174	0.279
RDC	0.174	0.249	0.187	0.325	0.166	0.222
CS_{score}	1.092	1.321	1.107	1.512	1.012	1.200
Backbone ^3J	1.225	1.359	1.228	1.562	1.004	1.182
Sidechain ^3J	1.348	1.853	1.440	1.768	1.000	1.504
Backbone S^2	1.070	1.510	1.175	2.156	1.000	2.226
Sidechain S^2	1.603	1.493	1.418	1.657	1.000	1.603
Backbone RDC	1.048	1.500	1.127	1.958	1.000	1.337
FF_{score}	**1.231**	**1.506**	**1.249**	**1.769**	**1.003**	**1.509**

[a]Chemical shifts are in ppm, J-coupling constants and RDC are in Hz, and the scores and S^2 parameters are unitless.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

FF scores of six *ff03*-series force field for 9 disordered proteins

Protein	*ff03*	*ff03**	*ff03w*	*ff03ws*	*ff03CMAP/TIP4PEw*	*ff03CMAP/TIP4PD*
HEWL19	1.290[a]	1.039	1.089	1.117	1.125	1.092
RS	4.453	1.657	1.600	1.454	1.140	1.105
HIVRev	1.761	1.763	1.677	1.208	1.244	1.295
AB40	2.016	1.756	1.872	1.758	1.124	1.001
AB42	2.106	1.866	1.972	1.858	1.184	1.021
ACTR	1.618	1.474	1.354	1.172	1.245	1.033
IA3	2.111	1.810	1.923	1.677	1.404	1.000
p53N	1.668	1.441	1.537	1.351	1.133	1.019
tauF4	1.587	1.590	1.384	1.301	1.226	1.000

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5.

Average of RMSDs for experimental observables of IDPs[a]

Exp. Type	*ff03*	*ff03**	*ff03w*	*ff03ws*	*ff03CMAP/TIP4PEw*	*ff03CMAP/TIP4PD*
Cα	1.332	0.986	1.001	0.849	0.641	0.535
Cβ	0.545	0.607	0.559	0.566	0.461	0.405
C	0.853	0.745	0.649	0.613	0.553	0.544
N	2.545	2.153	2.275	1.958	1.575	1.208
Hα	0.146	0.132	0.137	0.126	0.118	0.114
HN	0.334	0.311	0.320	0.297	0.262	0.237
$^{3}J_{HNH\alpha}$	1.322	0.996	1.026	0.920	0.475	0.462
RDC	0.928	0.788	0.764	0.737	0.764	0.722

[a]The experimental measurements are calculated in 4 different IDP systems at least. Chemical shifts are in ppm, J-coupling constants and RDCs are in Hz.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6.

RMSD of secondary chemical shifts, J-coupling constants, RDC and FF score of RS for six *ff03*-series force fields.[a]

	ff03	*ff03**	*ff03w*	*ff03ws*	*ff03CMAP/TIP4PEw*	*ff03CMAP/TIP4PD*
Cα	1.813	0.502	0.484	0.385	0.239	0.193
C	1.199	0.333	0.296	0.307	0.354	0.373
$^{3}J_{HNH\alpha}$	1.777	0.987	0.974	0.981	0.407	0.491
$^{1}J_{C\alpha C\beta}$	1.625	1.407	1.420	1.438	1.084	1.195
$^{1}J_{H\alpha C\alpha}$	4.190	2.462	2.511	2.395	1.774	1.824
$^{3}J_{CC\gamma}$	0.302	0.314	0.271	0.278	0.330	0.311
$^{3}J_{NC\gamma}$	0.386	0.254	0.272	0.233	0.230	0.217
RDC	1.425	0.908	0.939	0.841	0.788	0.716
CS_{score}	6.720	1.862	1.753	1.514	1.217	1.128
NMR_{score}	2.186	1.451	1.448	1.394	1.063	1.081
FF_{score}	**4.453**	**1.657**	**1.600**	**1.454**	**1.140**	**1.105**

[a]Chemical shifts are in ppm, J-coupling constants and RDCs are in Hz.