UC Irvine
UC Irvine Previously Published Works

Title

Test and Evaluation of ff99IDPs Force Field for Intrinsically Disordered Proteins

Permalink

https://escholarship.org/uc/item/78m9v9h4

Journal

Journal of Chemical Information and Modeling, 55(5)

ISSN

1549-9596

Authors

Ye, Wei
Ji, Dingjue
Wang, Wei
et al.

Publication Date

2015-05-26

DOI

10.1021/acs.jcim.5b00043

Peer reviewed

Test and Evaluation of *ff99IDPs* Force Field for Intrinsically Disordered Proteins

Wei Ye[1], **Dingjue Ji**[1], **Wei Wang**[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, Chemical Engineering and Materials Science, Biomedical Engineering, University of California, Irvine, California 92697-3900, USA

[3]Shanghai Center for Bioinformation Technology, 1278 Keyuan Road, Shanghai, 200235, China

Abstract

Over 40% eukaryotic proteomic sequences have been predicted as intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) and confirmed to be associated with many diseases. However, widely used force fields could not well reproduce the conformers of IDPs. A previously *ff99IDPs* force field was released with CMAP energy corrections for the 8 disorder promoting residues to simulate IDPs. In order to further confirm the performance of *ff99IDPs*, three representative IDPs systems (arginine-rich HIV-1 Rev, aspartic proteinase inhibitor IA$_3$, and α-Synuclein) were used to test and evaluate the simulation results. For free disordered proteins, the results show that the chemical shifts from the *ff99IDPs* simulations are in quantitative agreement with those from reported NMR measurements and better than those from *ff99SBildn*. Then, *ff99IDPs* can sample more clusters of disordered conformer than *ff99SBildn*. For structural proteins, both *ff99IDPs* and *ff99SBildn* can reproduce the conformations. In general, *ff99IDPs* can success in simulating the conformation of IDPs or IDRs both in bound and free states. However, relative errors could still be found at the boundaries of the scattering order-disorder promoting residues. Therefore, polarizable force fields might be one of possibility ways to further improve the performance on IDPs.
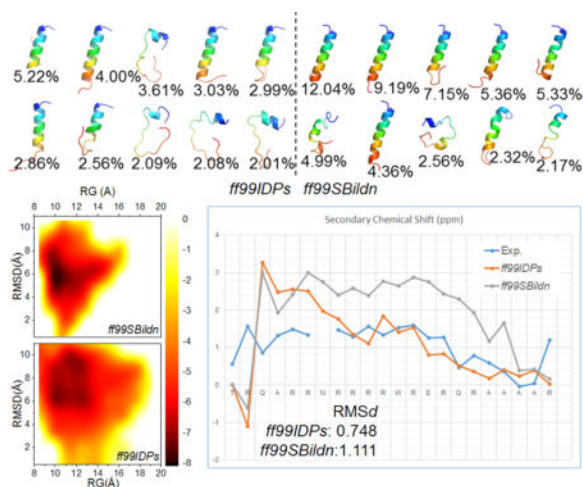
TOC image

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

Supporting Information

Figure S1 plots the RMSd time evolution for all the test systems; Figure S2 shows the top 10 clusters' representative structures of HIVRev, respectively in the first 100 ns, 70 ns, 80 ns, and 90 ns; Figure S3 shows calculated J-coupling of apo- and bound-HIVRev; Figure S4 shows comparisons of the calculated helicities of HIVRev between DSSP and STRIDE; Figures S5–S12 show all the secondary structure time evolution for all the test systems; Figure S13 represents structural clusters of bound-HIVRev; Figure S14 shows the secondary chemical shift of apo- and bound-IA3 and their comparison with experimental data. This material is available free of charge via the Internet at http://pubs.acs.org.

Note:

The authors declare that there is no conflict of interest.

Introduction

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) in structured proteins have been confirmed to play key roles in protein function. Although they cannot automatically fold into stable and ordered structures, the structural flexibilities have important functions in many cellular processes, such as molecular recognition, assembly, and post-translational modifications[1–5]. Over 40 % eukaryotic proteomic sequences have been predicted as disorder with more than 40 consecutive amino acids, indicating IDPs' wide spreading[6, 7]. Recently, IDPs are intensively studied on their association with many diseases: such as cancer[8], cardiovascular diseases[9], and neurodegenerations[10–14]. Thus, IDPs obviously provide exciting opportunities and bring huge challenges to the understanding of the protein structure-function paradigm.

Interestingly, IDPs can rapidly fold into structural proteins upon binding with their partners. Typical observations could be found in tumor suppressor p53, which acts as a central hub in multiple signaling pathways[8, 15–17], through interacting with hundreds of partners. Since the significant conformational adjustments from disordered structures to ordered, essential challenges are also brought to molecular dynamics (MD) simulations on IDPs.

Indeed, MD simulation becomes a powerful tool to investigate the dynamic motions in macromolecules at the long time scale (~ns-μs)[18–22]. However, accuracy of force field remains to be an issue in applications of MD simulations[23, 24]. In the case of IDPs, the issue is that most protein force fields are often too stable to model the unstructured proteins based on our tests. To overcome this issue, we developed an AMBER[25] force field, *ff99IDPs*[26], with the addition of grid-based energy correction maps (CMAP)[27, 28] on the standard *ff99SBildn* force field[29, 30] to reproduce the main chain torsional distributions of 8 disorder promoting residues (A, G, P, R, Q, S, E, and K) which were reported in the literature[31]. Tests

on the apo- and bound-state of measles virus nucleoprotein (MeV N_{TAIL}) and p53 show that *ff99IDPs* could reproduce IDPs better than the widely-used *ff99SBildn*. In addition, the rigid secondary structures in bound state could be well maintained under *ff99IDPs*.

In this study, three IDPs were used to test and evaluate the performance of *ff99IDPs* in both disordered apo-state and ordered bound-state. The arginine-rich HIV-1 Rev is a RNA-binding protein and regulates the HIV-1 replication cycle[32–34]. Another test was conducted on IA$_3$, the aspartic proteinase inhibitor. Aspartic proteinase has been confirmed to express in many human infectious agents life cycle[35–37]. Among its few natural inhibitors, IA$_3$ is a typical IDP which could fold into highly rigid helix upon binding with aspartic proteinase[38, 39]. The third test was α-Synuclein. It is an IDP on the cell membrane in substantia nigra pars compacta (SNc), which is crucial in the pathogenesis of Parkinson's disease. It may misfold into highly ordered cross-β fibrils and form Lewis bodies[18, 40–42]. Our tests show that *ff99IDPs* can reproduce highly disordered conformations of three tested IDPs, consistent with the experimental observations. Furthermore, the performance of *ff99IDPs* is better than that of *ff99SBildn* to sample the conformers of IDPs. For bound-state, both force fields have reproduced the conformers with stable secondary structures.

As a reference, we also tested the applicability of *ff99IDPs* on structured proteins, lysozyme and ubiquitin[30]. Our data show that the newly developed force field behaves similar to that of *ff99SBildn* in the tested structured proteins. However, our test data show that disagreement with experiments still exist for the ordered residues when they are scattered in long disorder-promoting sequences. The possible causes are discussed and remedies with the deployment of polarizable force fields are also proposed.

Materials and Methods

Selection of test systems

In order to quantitatively validate the performance of the newly developed AMBER force field *ff99IDPs*, intrinsically disordered proteins were searched in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Biological Magnetic Resonance Data Bank (BMRB)[43] with the following criteria: 1) The proteins which are exactly described as IDPs; 2) Thermodynamic data, such as chemical shift and order parameter (S^2), are available. Under these criteria, three representative systems: arginine-rich motif of HIV-1 Rev (termed **HIVRev**, PDB code: **1ETF**)[44], the aspartic proteinase inhibitor IA$_3$ (termed **IA$_3$**, PDB code: **1DP5**)[39], and the disordered region of micelle bound α-synuclein (termed **αSyn**, PDB code: **2KKW**)[45], were selected to evaluate the performance of *ff99IDPs*. HIVRev has a very high proportion of disorder-promoting residues, with only 4 order-promoting ones out of 21 residues. As a distinct contrast, IA$_3$ has a high proportion of order-promoting residues, with 14 order-promoting residues out of 31-mer polypeptide. Both HIVRev and IA$_3$ could fold into rigid α-helices in bound-state. αSyn is another typical IDP that consists of a long structured region and a long disordered loop. For HIVRev and IA$_3$, both disordered state and ordered state were tested. Furthermore, we also tested two widely-used proteins for the validation of previous force fields[30], hen egg white lysozyme (HEWL, PDB code: **6LYT**)[46] and ubiquitin (PDB code: **1UBQ**)[47], to check if the IDP-specific force field could be used for structural proteins. All the simulations were done under *ff99IDPs* and *ff99SBildn*.

Overview of *ff99IDPs* Implementation

The total energy of IDP specific force field *ff99IDPs* is the summation of *ff99SBildn* total energy and the dihedral energy correction term with parameters for the 8 disorder-promoting residues, as shown in Eq. 1[26]. All tested structures with PDB file format were first converted into AMBER topology files, using *ff99SBildn* force field. CMAP parameters were then added using the in-house PERL script[26].

$$E_{ff99IDPs} = E_{ff99SBildn} + E_{CMAP} \quad (1)$$

All other energy terms of *ff99IDPs* except the dihedral energy term remain the same as a chosen base additive force field of *ff99SBildn*. Furthermore, only the backbone dihedral parameters for the 8 disorder-promoting residues were optimized while the parameters for the other 12 residues remain the same to minimize the perturbation to folded structure distributions.

CMAP is a matrix of corrections on dihedral-grids with the corrections between grid points calculated with a two-dimensional bicubic interpolation method.[28] 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 ($G_{i,j}$) converted from φ/ψ distributions from the disordered protein structures to compute the correction matrix with Eq. 2.

$$\Delta G_{i,j} = -RT \ln (N_{i,j}/N_0) \quad (2)$$

where $N_{i,j}$ is the population of φ/ψ dihedral bin (i, j), and N_0 is the population of the most-populated bin. In this equation, 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 Eq. 3.

$$E_{i,j}^{CMAP} = \Delta G_{i,j}^{DB} - \Delta G_{i,j}^{MD} \quad (3)$$

where $\Delta G_{i,j}^{DB}$ and $\Delta G_{i,j}^{MD}$ are database and MD simulation converted free energies for φ/ψ dihedral bin (i, j), respectively. The iteration starts with a CMAP correction matrix initialized as zero, so the initial $\Delta G_{i,j}^{MD}$ 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 simulation was added to the base force field *ff99SBildn*. Root mean square deviations of population (termed RMS*p*) among all bins were calculated to quantitatively measure the difference between MD and database populations until RMS*p* was less than 0.15%.

Molecular dynamics simulations

All chosen structures were first minimized in SYBYL®-X 2.1.1[48] to eliminate any possible overlaps or clashes. All simulations and most analyzing procedures were performed using the AMBER12 software package[25]. Hydrogen atoms were added using the LEaP module of AMBER12. 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)[49] was employed to treat long-range electrostatic interactions with the default setting in AMBER12. The newly developed *ff99IDPs* was added as described in the previous literature[26] based on *ff99SBildn*, which was also taken as the benchmark to compare with the experimental data. All the MD simulations were accelerated with the CUDA version of PMEMD[50, 51] in GPU cores of NVIDIA® Tesla K20. The SHAKE algorithm[52] was used to constrain bonds involving hydrogen atoms. Up to 20,000-step steepest descent minimization was performed to relieve any structural clash in the solvated systems. This was followed by a 400-ps' heating up and a 200-ps' equilibration in the NVT ensemble at 298K with PMEMD of AMBER12. Langevin dynamics with a time step of 2 fs was used in the heating and equilibration runs with a friction constant of 1 ps^{-1}. To evaluate the performance of *ff99IDPs* and compare with *ff99SBildn*, five independent trajectories of 100 ns each were simulated for apo- and bound-HIVRev, apo- and bound-IA$_3$, and apo-αSyn, under *ff99IDPs* and *ff99SBildn*, respectively. To show the compatibility of *ff99IDPs* on normal proteins, five 100ns' trajectories were also conducted on HEWL and ubiquitin under both force fields. A total of 6.6 μs simulations was collected at 298K, taking about 3,500 GPU hours. Detailed simulation conditions are listed in Table 1.

Data Analyses

Root mean square deviations (RMSD) and fluctuations (RMSF) in MD trajectories were calculated with the PTRAJ module in AMBER12 and AmberTools13[25]. Structural cluster was conducted with the kclust program in MMTSB toolset[53]. Secondary structures of all the snapshots were identified with DSSP[54, 55]. Disorder population was calculated as the disordered (predicted by DSSP as "disordered loop" and "bend") population during the production runs. Disorder population within every 10 ns' period along all the IDP trajectories under *ff99IDPs* and *ff99SBildn* were plotted to test if both force fields can converge to sample the disorder conformers. STRIDE[56, 57] was also used for HIVRev to compare with DSSP. Experimental Cα chemical shift data for all models and N-H order parameters (S^2) for HEWL and ubiquitin were retrieved from the Biological Magnetic Resonance Data Bank[43]. Accession numbers could also be found in Table 1. PMF free energy landscapes were mapped by calculating normalized probability from a histogram analysis, and plotted with Origin 8.5. For each simulation, sampling was conducted every 50 ps (10000 snapshots for 5×100 ns' simulations). Radius of gyration (RG) and RMSD were both separated into 8 bins. The energy landscape was plotted among these 64 (8×8) bins. The secondary chemical shift data for the simulated structures were calculated with SPARTA version 1.01[58]. N-H J-coupling data of free and bound-HIVRev were calculated with Karplus Equations[59–61]. All the structural visualizations were represented with PyMOL 1.7[62].

Results and Discussion

We performed MD simulations on the models mentioned above. Then, the evaluations were performed through conformation sampling, structural clustering, helicity evolution, secondary Cα chemical shift, and other structural or thermodynamic indices. For every model, RMSD and Cα fluctuation (RMSF) were first calculated to ensure that *ff99IDPs* could sample rational conformations. Comparing to the initial structures, RMSD of each trajectory under both *ff99IDPs* and *ff99SBildn* (supplementary Figure S1) indicates that 100 ns simulations are sufficient to become dynamics equilibration at room temperature. Furthermore, RMSDs of free IDPs under *ff99IDPs* were mostly higher than those under *ff99SBildn*. This indicates that *ff99IDPs* might sample more flexible conformations than *ff99SBildn*.

In order to test the convergence of conformation sampling, the disorder populations within every 10 ns' simulation time were shown in Figure 1. For all the IDP systems, the disorder population became stable before 100 ns, indicating 100 ns simulation is sufficient for the convergence of conformation sampling. Taking HIVRev as an example, we also compared the representative structures and their occupation in the total conformation within first 70 ns, 80 ns, 90 ns, and 100ns, as shown in supplementary Figure S2. Most representative structures found in 100 ns could also be found in the other 3 time durations, with similar occupation ratios. This indicates that 100 ns simulations are sufficient for the conformation sampling.

HIV Rev ARM (HIVRev)

Secondary Cα chemical shift, clustering, and helicity of apo-HIVRev under *ff99IDPs* and *ff99SBildn* were shown in Figure 2. Apparent differences could be found in the structural clustering (Figures 2A and 2B). Top 10 clusters under *ff99IDPs* occupy 30.45 % of the total conformations (top 58 for 70 %), many of which have high ratio of population for disordered structures. However, under *ff99SBildn*, top 10 clusters occupy up to 55.47 % of the total conformations (top 20 for 70 %), only 2 of whose conformers are highly disordered. The PMF free energy landscapes with the reaction coordinates of RMSD and the radius of gyration (RG) (Figures 2C and 2D) show that the distribution of conformers from *ff99IDPs* is between RMSD at 1~11 Å and RG at 9 Å ~19 Å, and between RMSD at 1~10 Å and RG at 9 Å ~16 Å for *ff99SBildn*. This indicates that *ff99IDPs* could sample more flexible conformers than *ff99SBildn*, which is in agreement with the structural clustering. Using the clustered representative structures which occupy no less than 70% conformations and their occupancy, secondary Cα chemical shifts were calculated and compared with the experimental data (Figure 2E). The full length RMS*d* was 0.748 ppm for *ff99IDPs* and 1.111 ppm for *ff99SBildn* between predicted chemical shifts and experimental data, respectively. This suggests that the predicted chemical shifts under *ff99IDPs* are more approaching to experimental values than those under *ff99SBildn*. Calculated J-coupling values are shown in supplementary Figure S3. Because of the lack for the exact experimental data, the full length RMSd was not calculated. However, similar to that of chemical shift, the predicted J-coupling values of *ff99IDPs* are better than those of *ff99SBildn*. The helicity of HIVRev is shown in Figures 2F and 2G. These figures indicate that *ff99IDPs* reproduces significantly

lower helical conformations than *ff99SBildn*, both in different statistics strategies. We also used STRIDE to yield the helical population for HIVRev, as shown in supplementary Figure S4. The results are similar to those of DSSP which indicates the reliability and reproducibility of DSSP algorithm. Detailed representations of secondary structures along simulation time could be found in Figures S5–S12 in supporting information.

Our tests so far show that *ff99IDPs* performs better than *ff99SBildn* for IDPs simulation. The clustering analysis shows a higher structural flexibility derived by *ff99IDPs*. Structure trinity of IDPs, i.e. ordered, molten globular, and disordered structures[3], could also be observed in the *ff99IDPs* derived structure representatives, especially in the middle region and C-terminus. In the N-terminus, the performance of *ff99IDPs* is similar to *ff99SBildn* for chemical shift or J-coupling prediction. We detected the internal interactions in apo-HIVRev under both force fields. Long-range interactions, other than short-range hydrogen bonds and hydrophobic interactions, play key roles in the internal contacts. Under *ff99IDPs*, hydrogen bonds and electrostatic interactions were only found in the N-terminus (A-T-R-Q), as the threonine, arginine and glutamine have long side chains which can tangle with the charge-charge interactions; while under *ff99SBildn*, these interactions were also found in the middle region. The over-stability in the N-terminus under *ff99IDPs* may be caused by the only consideration of correction on the backbone dihedrals, while the charges and electrostatic interactions should be considered as another important aspect in the IDPs modeling. This could be implemented to improve the force field in the future.

For DNA-bound HIVRev, both *ff99IDPs* and *ff99SBildn* have good performance to reproduce the conformation of structured HIVRev. Calculated secondary Cα chemical shift of bound-HIVRev under both force fields is shown in Figure 3 and compared with experimental data. The RMSD is 0.698 ppm under *ff99IDPs* and 0.604 ppm under *ff99SBildn*, respectively. In Figure S13, helical structures of HIVRev were very stable upon DNA binding, indicating that *ff99IDPs* could also be used to ordered proteins. Time evolutions of secondary structure for bound-HIVRev under both force fields are shown in Figure S6.

Aspartic Proteinase Inhibitor (IA$_3$)

IA$_3$ is another intrinsically disordered protein, which folds rigidly upon binding with aspartic proteinase. Similar to HIVRev, the structural cluster also indicates that *ff99IDPs* reproduces more disordered and flexible conformations than *ff99SBildn* (Figures 4A and 4B). Top 10 clusters under *ff99IDPs* occupy only 14.09 % of the total conformations (top 172 for 70 %); meanwhile under *ff99SBildn*, top 10 clusters occupy 23.63 % of the total conformations (top 85 for 70 %). Therefore, the disordered level of conformer from *ff99IDPs* was also higher than that from *ff99SBildn*. The full length RMSd between predicted secondary Cα chemical shifts and experimental data is 0.835 ppm for *ff99IDPs* and 1.186 ppm for *ff99SBildn*, respectively. Time evolutions of secondary structure for apo-IA$_3$ under both force fields are shown in Figure S7. This also shows that the performance of *ff99IDPs* is better than that of *ff99SBildn* to simulate IDPs. However, both force fields could not well reproduce the disorder conformers for the middle region of apo-IA$_3$ because of over stabilized α-helical structures. CMAP corrections are only added onto disorder-promoting

residues, which might cause an energy gap between order- and disorder-promoting residues. As a result, most significant disparities in IA_3 were found at the boundaries between order- and disorder-promoting residues. This might be eliminated by the polarizable charge model.

The structures of aspartic proteinase bound-IA_3 modeled in both force fields are similar. Because of the lack of experimental data, we cannot quantitatively evaluate the performance. Chemical shift for bound-IA_3 could be found in supplementary Figure S14.

α-Synuclein

The NMR structures of αSyn have two long α-helices linked with a short turn (Res. 1–94, Figure 5A). In the crystal experiment, this region binds with vesicle and micelle[45]. In this test, we focused on the disordered region (Res. 95–140) and compared the thermodynamics data with the experiments. Therefore, all analysis results are focused on this IDPs region. Structural clustering is shown in Figures 5B and 5C and secondary structure evolutions are listed in supplementary Figure S9. The conformers of top 10 clusters occupies of 20.05 % (*ff99IDPs*) and 23.77 % (*ff99SBildn*), respectively. As shown in Figure 5, the whole conformers are highly disordered under ff99IDPs; however, the conformers evolve some helical structure under *ff99SBildn*. This shows that the disordered magnitude is higher under *ff99IDPs* than under *ff99SBildn*. The full length RMSd between predicted secondary Cα chemical shifts and experimental data is 0.461 ppm for *ff99IDPs* and 0.588 ppm for *ff99SBildn*, respectively. This also shows that the performance of *ff99IDPs* is better than that of *ff99SBildn* to simulate IDPs.

Lysozyme and Ubiquitin

Lysozyme and ubiquitin have been tested for the evaluation of previous force field many times[30, 63]. Here these two intensively studied proteins were also employed to evaluate the quality of *ff99IDPs*. Figures 6 and 7 show the secondary Cα chemical shift and the order parameter (S^2) derived by *ff99IDPs* and *ff99SBildn* and their comparison with the experimental data. Both force fields have good performance, showing the newly development *ff99IDPs* is also suitable for the normal protein. Calculated chemical shift results from both force fields are similar. It is interesting to note that *ff99IDPs* leads to more disordered properties in the loop and turn regions, as shown in S^2 values, which is consistent with the destabilization of CMAP energy term. Their secondary structure evolutions are shown in Figures S10 and S11.

Conclusion and Perspectives

In this study, we validated the quality of newly developed ff99IDPs force field with three intrinsically disordered proteins (HIVRev, IA_3, and αSyn) and two ordered proteins (lysozyme and ubiquitin) with extensive all-atom molecular dynamics simulations. Thermodynamic data were calculated and compared with experimental data. Overall, *ff99IDPs* leads to more consistent chemical shift than *ff99SBildn*, and reproduces more flexible disorder conformations for IDPs. Thus, the validation data indeed show that *ff99IDPs* has better performance on IDPs than *ff99SBildn*. At the same time, *ff99IDPs* can also be used to simulate structural proteins.

Nevertheless, some limitations do exist. *ff99IDPs* cannot perfectly reproduce the full length of the intrinsically disordered proteins, e.g. the N-terminus of HIVRev and the middle region of IA$_3$. The stable secondary structures (α-helix in most time) in these regions were observed in simulations, which may be the cause of the high Cα chemical shift. The electrostatic and hydrogen bonding interactions in these regions might play key roles in the structural stabilization, rather than the dihedral energy correction in CMAP energy term. Thus, the next step for improving the performance and accuracy of the IDP specific force field might be focused on the polar interactions. Indeed, the charge distribution of a residue should be perturbed by its neighboring residues, which is the principle of polarizable force fields. Therefore, our effort next step will be to explore how to improve the accuracy of the inter-residues' interactions in the disordered regions with strategies in polarizable force fields and incorporate these to improve the IDPs force fields.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Dunker AK, Babu MM, Barbar E, Blackledge M, Bondos SE, Dosztányi Z, Dyson HJ, Forman-Kay J, Fuxreiter M, Gsponer J. What's in a Name? Why These Proteins Are Intrinsically Disordered. Intrinsically Disordered Proteins. 2013; 1:0–4.

2. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. Annu Rev Biophys. 2008; 37:215–246. [PubMed: 18573080]

3. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW. Intrinsically Disordered Protein. J Mol Graph Model. 2001; 19:26–59. [PubMed: 11381529]

4. Gough J, Dunker AK. Sequences and Topology: Disorder, Modularity, and Post/Pre Translation Modification. Curr Opin Struct Biol. 2013; 23:417–419. [PubMed: 23721748]

5. Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. Annu Rev Biochem. 2014

6. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic Protein Disorder in Complete Genomes. Genome Inform Ser. 2000:161–171.

7. Dunker A, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic ZK. Intrinsic Disorder and Protein Function. Biochemistry-US. 2002; 41:6573–6582.

8. Iakoucheva LM, Brown CJ, Lawson JD, Obradovi Z, Dunker AK. Intrinsic Disorder in Cell-Signaling and Cancer-Associated Proteins. J Mol Biol. 2002; 323:573–584. [PubMed: 12381310]

9. Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN. Abundance of Intrinsic Disorder in Protein Associated with Cardiovascular Disease. Biochemistry-US. 2006; 45:10448–10460.

10. Wenning GK, Jellinger KA. The Role of [alpha]-Synuclein and Tau in Neurodegenerative Movement Disorders. Curr Opin Neurol. 2005; 18:357–362. [PubMed: 16003109]

11. Pawar A, Dubay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CMP. Prediction of "Aggregation-Prone" and "Aggregation-Susceptible" Regions in Proteins Associated with Neurodegenerative Diseases. J Mol Biol. 2005; 350:379–392. [PubMed: 15925383]

12. Hampel H, Blennow K, Shaw LM, Hoessler YC, Zetterberg H, Trojanowski JQ. Total and Phosphorylated Tau Protein as Biological Markers of Alzheimer's Disease. Exp Gerontol. 2010; 45:30–40. [PubMed: 19853650]

13. Goedert M, Spillantini M, Jakes R, Rutherford D, Crowther R. Multiple Isoforms of Human Microtubule-Associated Protein Tau: Sequences and Localization in Neurofibrillary Tangles of Alzheimer's Disease. Neuron. 1989; 3:519–526. [PubMed: 2484340]

14. Gamblin TC, Berry RW, Binder LI. Modeling Tau Polymerization in Vitro: a Review and Synthesis. Biochemistry-US. 2003; 42:15009–15017.

15. Tsai C-J, Ma B, Nussinov R. Protein–Protein Interaction Networks: How Can a Hub Protein Bind So Many Different Partners? Trends Biochem Sci. 2009; 34:594–600. [PubMed: 19837592]

16. Collavin L, Lunardi A, Del Sal G. P53-Family Proteins and Their Regulators: Hubs and Spokes in Tumor Suppression. Cell Death Differ. 2010; 17:901–911. [PubMed: 20379196]

17. Chouard T. Structural Biology: Breaking the Protein Rules. Nature. 2011; 471:151–3. [PubMed: 21390105]

18. Ye W, Wang W, Jiang C, Yu Q, Chen H. Molecular Dynamics Simulations of Amyloid Fibrils: an in silico Approach. Acta Bioch Bioph Sin. 2013; 45:503–508.

19. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE III, DeBolt S, Ferguson D, Seibel G, Kollman P. 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. 1995; 91:1–41.

20. Brooks BR, Bruccoleri RE, Olafson BD, Swaminathan S, Karplus M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comput Chem. 1983; 4:187–217.

21. Berendsen HJ, van der Spoel D, van Drunen R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. Comput Phys Commun. 1995; 91:43–56.

22. Jorgensen WL, Tirado-Rives J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. J Am Chem Soc. 1988; 110:1657–1666. [PubMed: 27557051]

23. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. Long-Timescale Molecular Dynamics Simulations of Protein Structure and Function. Curr Opin Struct Biol. 2009; 19:120–127. [PubMed: 19361980]

24. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. Proteins: Struct Funct Bioinfo. 2010; 78:1950–1958.

25. Case, D., Darden, T., Cheatham, T., III, Simmerling, C., Wang, J., Duke, R., Luo, R., Walker, R., Zhang, W., Merz, K. AMBER 12. University of California; San Francisco: 2012.

26. Wang W, Ye W, Jiang C, Luo R, Chen HF. New Force Field on Modeling Intrinsically Disordered Proteins. Chem Biol Drug Des. 2014; 84:253–69. [PubMed: 24589355]

27. MacKerell AD, Feig M, Brooks CL. Improved Treatment of the Protein Backbone in Empirical Force Fields. J Am Chem Soc. 2004; 126:698–699. [PubMed: 14733527]

28. MacKerell AD, Feig M, Brooks CL. 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; 25:1400–1415. [PubMed: 15185334]

29. Sorin EJ, Pande VS. Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. Biophys J. 2005; 88:2472–93. [PubMed: 15665128]

30. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins: Struct Funct Bioinfo. 2006; 65:712–725.

31. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence Complexity of Disordered Protein. Proteins: Struct Funct Bioinfo. 2001; 42:38–48.

32. Tan R, Chen L, Buettner JA, Hudson D, Frankel AD. RNA Recognition by an Isolated α Helix. Cell. 1993; 73:1031–1040. [PubMed: 7684657]

33. Calnan BJ, Biancalana S, Hudson D, Frankel AD. Analysis of Arginine-Rich Peptides from the HIV Tat Protein Reveals Unusual Features of RNA-Protein Recognition. Gene Dev. 1991; 5:201–210. [PubMed: 1899841]

34. Lazinski D, Grzadzielska E, Das A. Sequence-Specific Recognition of RNA Hairpins by Bacteriophage Antiterminators Requires a Conserved Arginine-Rich Motif. Cell. 1989; 59:207–218. [PubMed: 2477156]

35. Craig C, Race E, Sheldon J, Whittaker L, Gilbert S, Moffatt A, Rose J, Dissanayeke S, Chirn G-W, Duncan IB. HIV Protease Genotype and Viral Sensitivity to HIV Protease Inhibitors Following Ssaquinavir Therapy. AIDS. 1998; 12:1611–1618. [PubMed: 9764779]

36. Monod M, Togni G, Hube B, Sanglard D. Multiplicity of Genes Encoding Secreted Aspartic Proteinases in Candida Species. Mol Microbiol. 1994; 13:357–368. [PubMed: 7984113]

37. Dame JB, Reddy GR, Yowell CA, Dunn BM, Kay J, Berry C. Sequence, Expression and Modeled Structure of an Aspartic Proteinase from the Human Malaria Parasite *Plasmodium falciparum*. Mol Biochem Parasit. 1994; 64:177–190.

38. Dreyer T, Valler M, Kay J, Charlton P, Dunn B. The Selectivity of Action of the Aspartic-Proteinase Inhibitor IA3 from Yeast (*Saccharomyces cerevisiae*). Biochem J. 1985; 231:777. [PubMed: 3907626]

39. 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 Mol Biol. 2000; 7:113–117.

40. Auluck PK, Caraveo G, Lindquist S. alpha-Synuclein: Membrane Interactions and Toxicity in Parkinson's Disease. Annu Rev Cell Dev Biol. 2010; 26:211–233. [PubMed: 20500090]

41. Goldberg MS, Lansbury PT Jr. Is There a Cause-and-Effect Relationship between α-Synuclein Fibrillization and Parkinson's Disease? Nat Cell Biol. 2000; 2:E115–E119. [PubMed: 10878819]

42. Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC. Molecular Basis for Amyloid Fibril Formation and Stability. Proc Natl Acad Sci USA. 2005; 102:315–320. [PubMed: 15630094]

43. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z. BioMagResBank. Nucleic Acids Res. 2008; 36:D402–D408. [PubMed: 17984079]

44. Casu F, Duggan BM, Hennig M. The Arginine-Rich RNA-Binding Motif of HIV-1 Rev Is Intrinsically Disordered and Folds upon RRE Binding. Biophys J. 2013; 105:1004–1017. [PubMed: 23972852]

45. Rao JN, Jao CC, Hegde BG, Langen R, Ulmer TS. A Combinatorial NMR and EPR Approach for Evaluating the Structural Ensemble of Partially Folded Proteins. J Am Chem Soc. 2010; 132:8657–8668. [PubMed: 20524659]

46. Young A, Dewan JC, Nave C, Tilton R. Comparison of Radiation-Induced Decay and Structure Refinement from X-Ray Data Collected From Lysozyme Crystals at Low and Ambient Temperatures. J Appl Crystallo. 1993; 26:309–319.

47. Vijay-Kumar S, Bugg CE, Cook WJ. Structure of Ubiquitin Refined at 1.8 Å Resolution. J Mol Biol. 1987; 194:531–544. [PubMed: 3041007]

48. Certara. SYBYL-X Suite, Version 2. 2012; 1:1.

49. Darden T, York D, Pedersen L. Particle Mesh Ewald: an N.log(N) Method for Ewald Sums in Large Systems. J Chem Phys. 1993; 98:10089–92.

50. Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. J Biol Chem. 2012; 8:1542.

51. Götz AW, Salomon-Ferrer R, Poole D, Grand S, Walker R. Routine Microsecond Molecular Dynamics Simulations with AMBER. Part II: Particle Mesh Ewald. J Chem Theory Comput. 2013; 9:3878–3888. [PubMed: 26592383]

52. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. J Chem Phys. 1977; 23:327–41.

53. Feig M, Karanicolas J, Brooks CL III. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. J Mol Graph Model. 2004; 22:377–395. [PubMed: 15099834]

54. Kabsch W, Sander C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. Biopolymers. 1983; 22:2577–2637. [PubMed: 6667333]

55. 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:D411–D419. [PubMed: 21071423]

56. Frishman D, Argos P. Knowledge-Based Protein Secondary Structure Assignment. Proteins. 1995; 23:566–79. [PubMed: 8749853]

57. Heinig M, Frishman D. STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. Nucleic Acids Res. 2004; 32:W500–W502. [PubMed: 15215436]

58. Shen Y, Bax A. Protein Backbone Chemical Shifts Predicted from Searching a Database for Torsion Angle and Sequence Homology. J Biomol NMR. 2007; 38:289–302. [PubMed: 17610132]

59. Bystrov VF. Spin-Spin Coupling and the Conformational States of Peptide Systems. Prog Nucl Mag Res Sp. 1976; 10:41–82.

60. Pardi A, Billeter M, Wüthrich K. Calibration of the Angular Dependence of the Amide Proton-C Alpha Proton Coupling Constants, 3JHN alpha, in a Globular Protein. Use of 3JHN alpha for Identification of Helical Secondary Structure. J Mol Biol. 1984; 180:741–751. [PubMed: 6084720]

61. Ludvigsen S, Andersen KV, Poulsen FM. Accurate Measurements of Coupling Constants from Two-Dimensional Nuclear Magnetic Resonance Spectra of Proteins and Determination of φ-Angles. J Mol Biol. 1991; 217:731–736. [PubMed: 2005622]

62. Schrodinger, LLC, In; 2014

63. Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD Jr. Importance of the CMAP Correction to the CHARMM22 Protein Force Field: Dynamics of Hen Lysozyme. Biophys J. 2006; 90:L36–L38. [PubMed: 16361340]
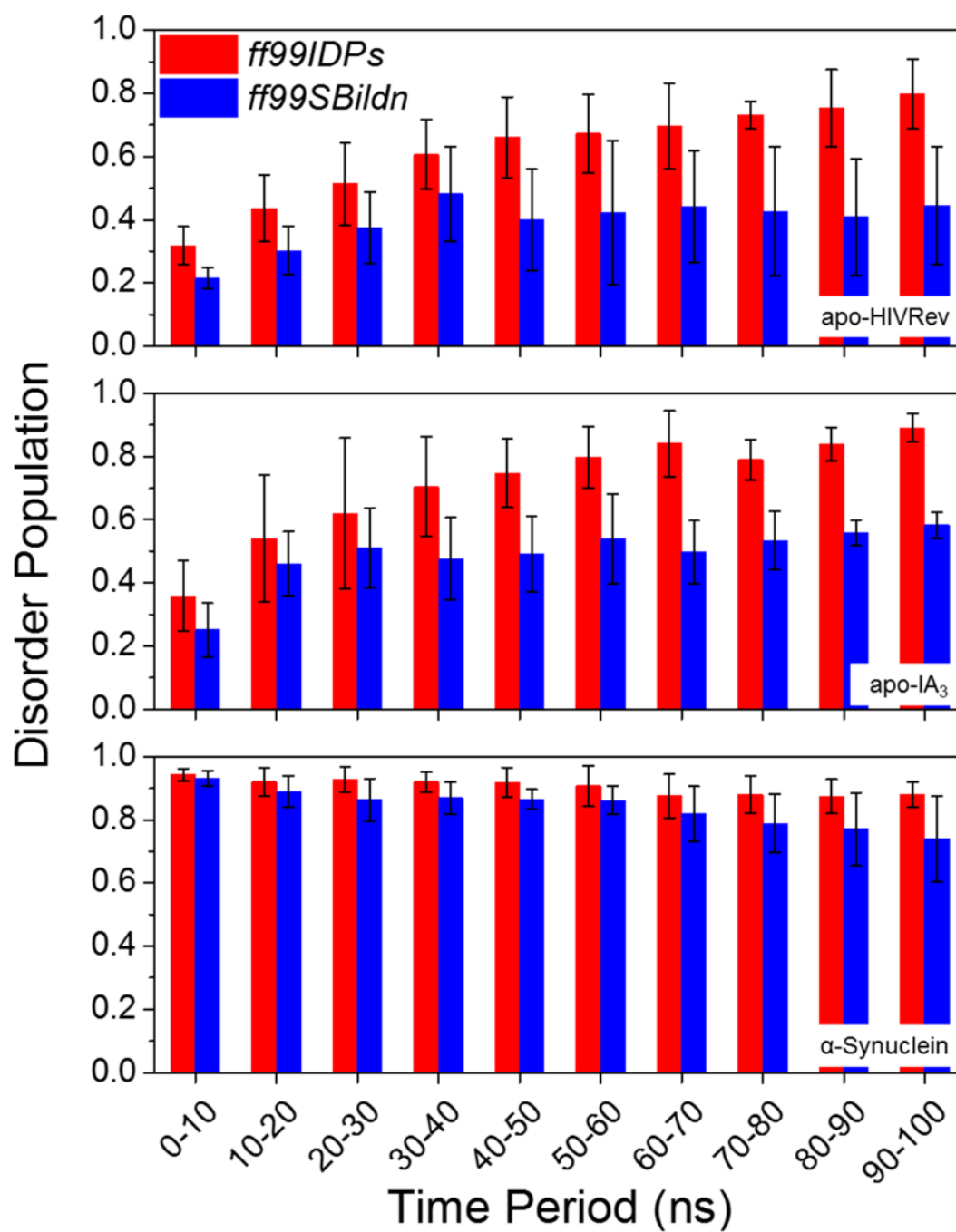
Figure 1.
Disorder population within 10 ns' time period along all the trajectories for three test IDPs, under *ff99IDPs* (red) and *ff99SBildn* (blue).
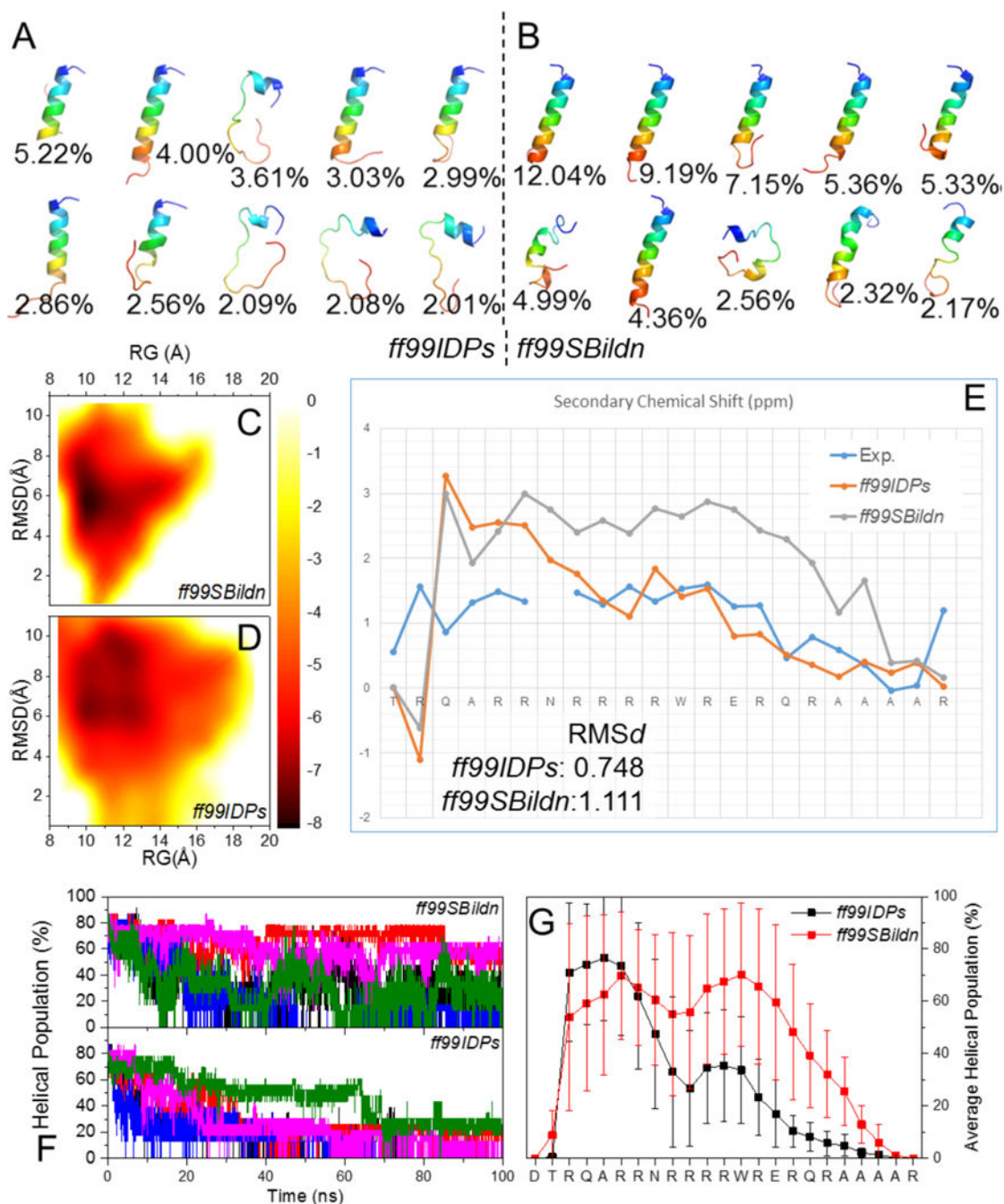
Figure 2.
Simulation and thermodynamic data derived from *ff99IDPs* and *ff99SBildn* for apo-HIVRev. (A)(B) Representative structures of top 10 clusters and their occupations. (C)(D) PMF free energy landscape on 2D space of radius gyration (RG) and RMSD, showing ff99IDPs could sample wider and more flexible conformation space. (E) Comparison of the secondary chemical shift data. (F)(G) Comparison of the average helicity under both force fields, with residual averaging (F) and time averaging (G), respectively.
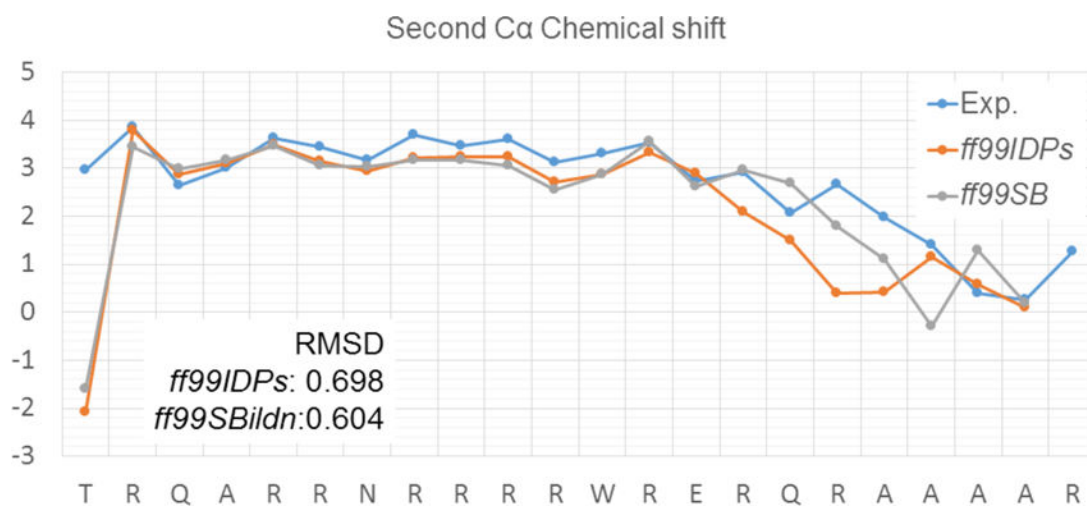
Figure 3.
Comparison of the secondary chemical shift data of bound-HIVRev under *ff99IDPs* and
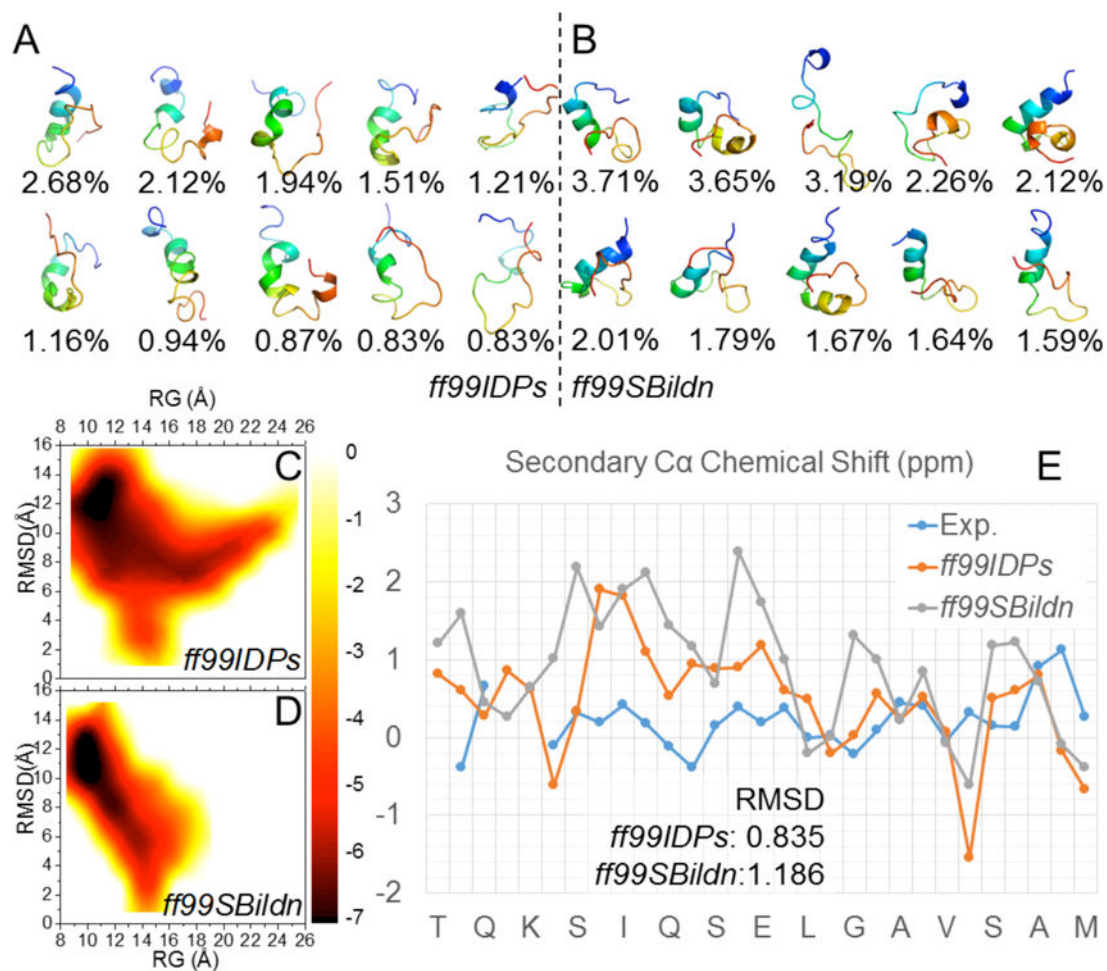*ff99SBildn.* RMS*d* values are calculated with the outlier terminal threonine omitted.

Figure 4.
Simulation and thermodynamic data derived from *ff99IDPs* and *ff99SBildn* for apo-IA$_3$. (A)
Representative structures of top 10 clusters and their occupations from ff99IDPs. (B)
Representative structures of top 10 clusters and their occupations from ff99SBildn. (C)(D)
PMF free energy landscape on 2D space of radius gyration (RG) and RMSD. (E)
Comparison of the secondary chemical shift data.
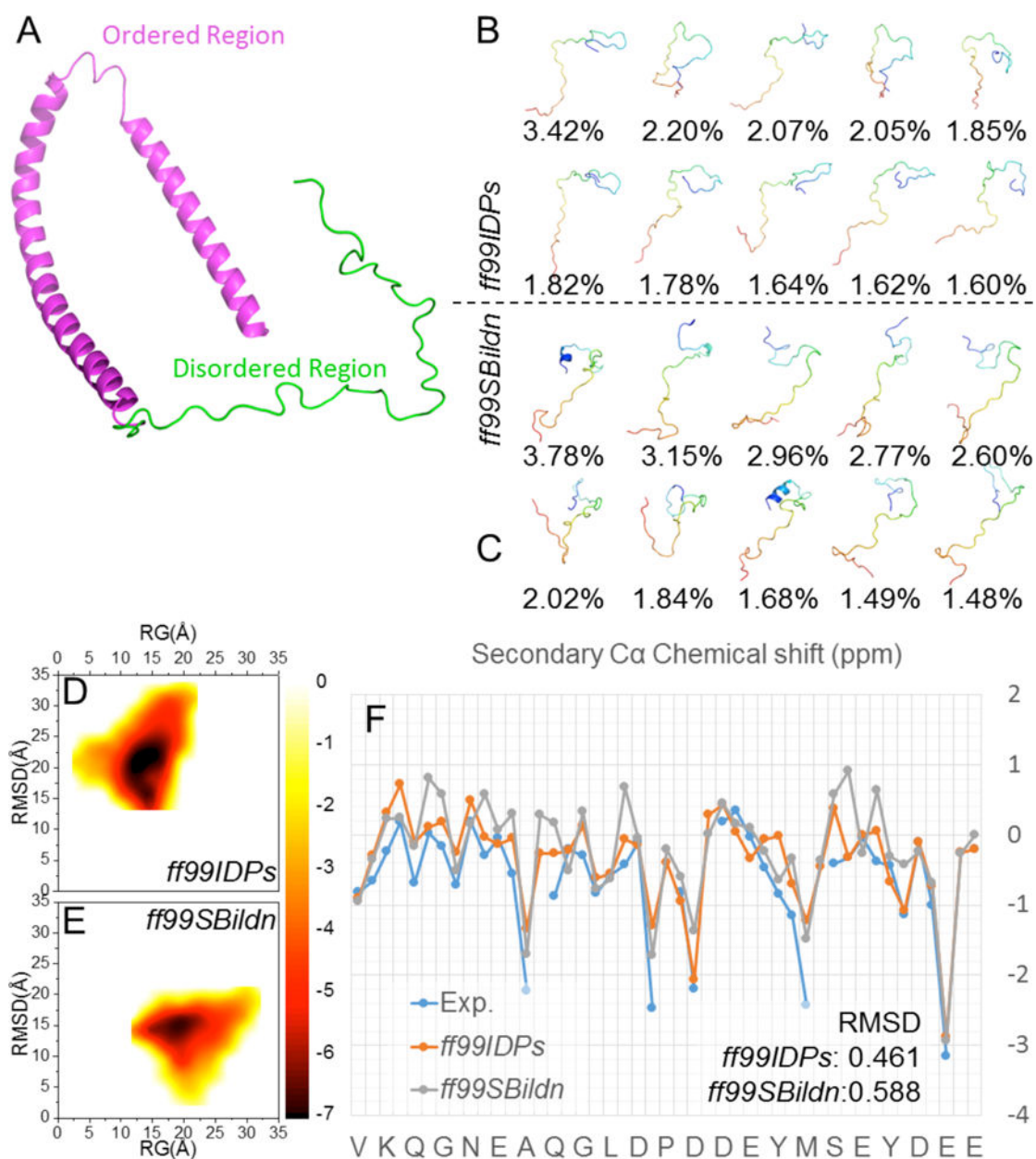
Figure 5.
Simulation and thermodynamic data derived from *ff99IDPs* and *ff99SBildn* for apo-αSyn.
(A) Cartoon representative of α-Synuclein, with the ordered and disordered parts labelled.
(B) Representative structures of top 10 clusters and their occupations from ff99IDps. (C)
Representative structures of top 10 clusters and their occupations from ff99SBildn. (D)(E)
PMF free energy landscape on 2D space of radius gyration (RG) and RMSD. (F)
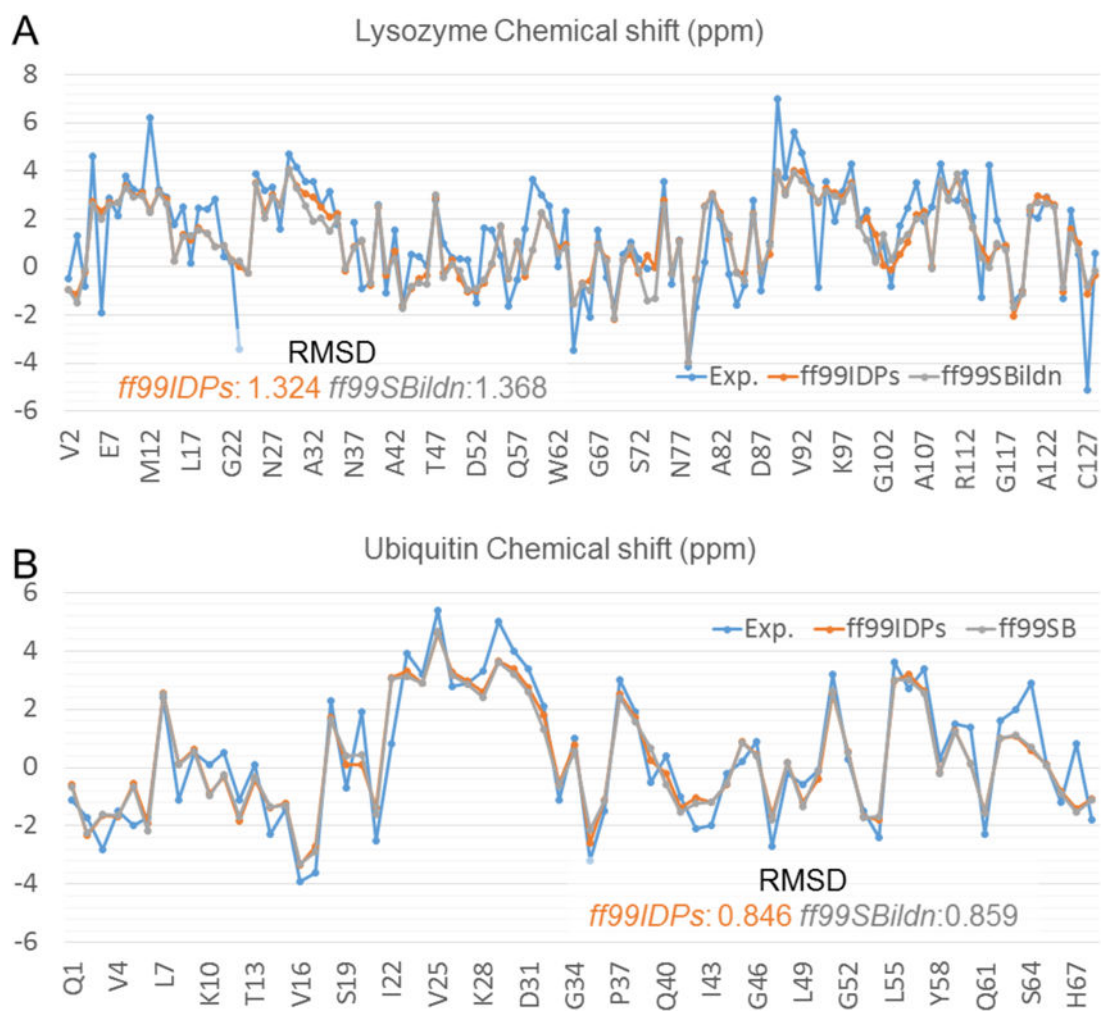Comparison of the secondary chemical shift data.

Figure 6.
Secondary chemical shift data comparison for lysozyme (A) and ubiquitin (B).
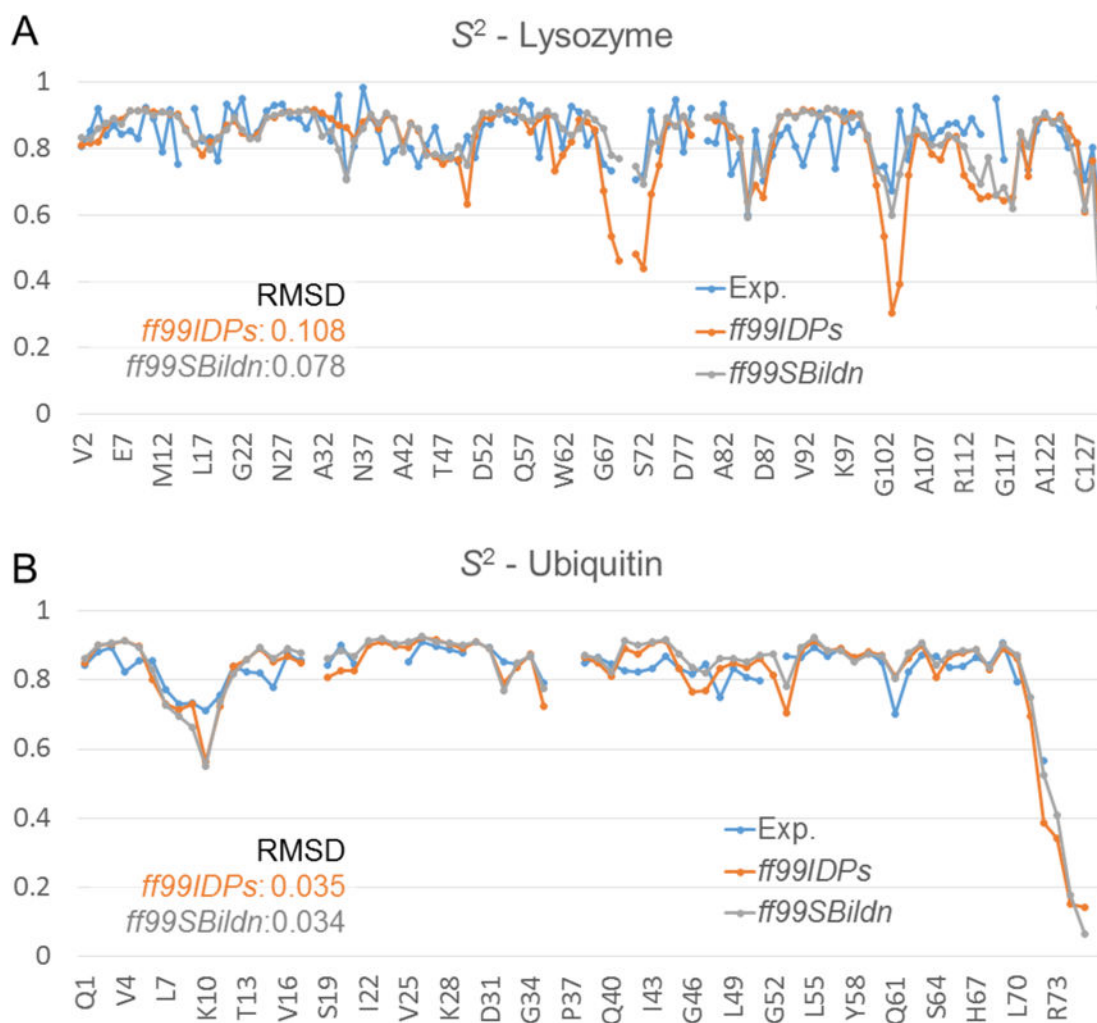
Figure 7.
Order parameter data (S^2) comparison for lysozyme (A) and ubiquitin (B).

Table 1

Simulation conditions for all models.

System	Force Field	Traj. Num.	Simulation Time (ns)	Ions	Waters	BMRB Accession No.
Apo-HIVRev	*ff99IDPs*	5	100	9 Cl⁻	4409	18851
	ff99SBildn	5	100			
Bound-HIVRev	*ff99IDPs*	5	100	23 Na⁺	12480	18852
	ff99SBildn	5	100			
Apo-IA₃	*ff99IDPs*	5	100	1 Na⁺	6434	6078
	ff99SBildn	5	100			
Bound-IA₃	*ff99IDPs*	3	100	24 Na⁺	11781	–
	ff99SBildn	3	100			
αSyn	*ff99IDPs*	5	100	9 Na⁺	47408	16302
	ff99SBildn	5	100			
HEWL	*ff99IDPs*	5	100	8 Cl⁻	5799	4562, 18304(S^2)
	ff99SBildn	5	100			
Ubiquitin	*ff99IDPs*	5	100	–	4669	5387, 6470(S^2)
	ff99SBildn	5	100			