Supporting Information: Refined Parameterization of Non-Bonded Interactions Improves Conformational Sampling and Kinetics of Protein Folding Simulations

Jejoong Yoo and Aleksei Aksimentiev*

Center for the Physics of Living Cells, Department of Physics, University of Illinois at Urbana–Champaign, 1110 West Green Street, Urbana, Illinois 61801

E-mail: aksiment@illinois.edu

Table of Contents

Supporting Methods

SM1: General MD simulation protocol

All simulations were carried out using the Gromacs 5.0.4 package[1] on CPU-only XE nodes of the Blue Waters (UIUC) supercomputer. Temperature was kept constant using the Nosé-Hoover scheme.[2,3] Pressure was controlled using the Parrinello-Rahman scheme.[4] Electrostatic forces were computed using the Particle-mesh Ewald (PME) summation scheme[5] with 12-Å real-space cutoff and 1.2 Å grid spacing. Lennard-Jones (LJ) interactions were computed using a switching scheme with 10–12 Å switching window. SETTLE[6] and LINCS[7] algorithms were used to constrain bonds with hydrogen in water and protein, respectively. Integration time step was 2 fs.

SM2: Force fields

Amino acids and proteins were described using either the AMBER ff99sb-ildn-phi[8–10] or ff03ws[11,12] parameter sets.

To perform simulations using the ff99sb-ildn-phi force field, we added the phi modification[10] to the AMBER ff99sb-ildn implementation included in the Gromacs 5.0.4 package. Our CUFIX corrections to ion–ion, ion–carboxylate, amine–carboxylate, and aliphatic carbon–carbon interactions[13–15] for the ff99sb-ildn-phi parameter set were implemented by modifying the nonbond_params section of the Gromacs parameter file. For water, we employed the original TIP3P model.[16] For ions, we employed the ion parameters developed by the Cheatham group.[17] The final parameter set including all CUFIX corrections (ff99sb-ildn-phi-CUFIX) can be downloaded from http://bionano.physics.illinois.edu/CUFIX.

To perform simulations using the ff03ws parameter set,[11,12] we used the Gromacs implementation of Dr. Best (ff03ws.tgz) (See footnote [1]). For water, we employed the TIP4P-2005 model[18] along with modified water–solute LJ parameters taken from Ref. 12. For ions, we

[1] http://www.gromacs.org/Downloads/User_contributions/Force_fields

employed the default ion parameters of the ff03 parameter set,[11] a choice consistent with Ref. 12.

For simulations using the charmm22* parameter set,[19] we used the Gromacs implementation (charmm22star.ff) downloaded from the Gromacs website (See footnote [1]). For water, we employed the CHARMM-modified TIP3P model.

SM3: Calibration of LJ parameters for aliphatic carbon atoms (type CT) of AMBER ff99

The calibration simulations were performed using the same setup as described in Refs. 20 and 21. 32 N-acetyl-leucine-methyl-amide (NALMA) and 1760 TIP3P water molecules were confined to a cubic box 4-nm on a side and simulated under periodic boundary condition at 298 K and 1 bar. The initial configuration containing randomly distributed NALMA molecules was created using the genbox program of the Gromacs package. The simulations were carried out using the Gromacs 5.0.4 package and the AMBER ff99sb-ildn-phi force field. Corrections to the CT–CT pair σ parameter were specified in the nonbond_param section of the Gromacs parameter file. Note that we did not modify the default LJ parameters of the CT type.

For each correction to the CT–CT pair σ, we performed a 20 ns equilibration of the NALMA system. The last 10 ns fragment of each trajectory was used to compute the radial distribution function (RDF) of inter-solute carbon–carbon atoms for each NALMA monomer. The final RDF was computed by averaging over the 32 RDF curves. For each carbon–carbon distance, the error bar was computed as the standard error of the 32 RDF values.

SM4: Replica-exchange MD (REMD) simulations of WW domain and HP35

The amino acid sequence of the WW domain protein was that of the GTT mutant:[22] EEKLPPGWEKRMSAD–GRVYYFNHITGTTQWERPSG; residue index 4 to 39 of PDB 2F21.[23] Proton of His27 was assigned to its NE2 atom using pdb2gmx program of the Gromacs package. The sequence and the structure (PDB 2F4K) of the HP35 protein were those of the Lys65Nle/Lys70Nle double mutant:[24] LSDEDFKAVFGMTRSAFANLPLWJQQHLJKEKGLF; residue index 42 to 76 of 2F4K. In the above sequence, J indicates a norleucine (Nle) residue. To be consistent with Ref. 24, His68 was protonated. Because WW domain and HP35 are of the same size, the same setup was used for REMD simulations of both proteins.

The initial coordinates of the protein were taken from the respective crystal structure. The protein was solvated in a rhombic dodecahedron unit cell ($a = b = c =\sim 53$ Å, $\alpha = \beta = 60°$, $\gamma = 90°$); the center-of-mass distance between the periodic images of the protein was 53 Å. The final system contained 3,200 water molecules. Na^+ and Cl^- ions were randomly added to a concentration of 100 mM. The simulation of each system began with a 1,000-step minimization, followed by a 100-ps equilibration having each non-hydrogen atom harmonically restrained to its initial coordinates using a harmonic constant $k = 1{,}000$ kJ/(mol·nm^2). Then, the system was equilibrated in a constant-pressure/constant-temperature ensemble for 50 ns at 300 K. Starting from the last snapshot of the equilibration trajectory, the system was equilibrated for 20 ns at 1000 K to produce an unfolded conformation. The fraction of native contacts in the unfolded conformations of the WW domain and HP35 were 0.02 and 0.06, respectively.

REMD simulations of protein folding[25,26] were performed starting from the unfolded conformations. Each simulation employed 56 replicas with temperature ranging from 284 to 453 K. The following list of replica temperatures was determined using the temperature prediction server[27] for the target exchange probability of 0.3: 284.00, 286.50, 289.02, 291.57, 294.12, 296.71, 299.31, 301.92, 304.55, 307.19, 309.87, 312.57, 315.28, 318.02, 320.78, 323.55,

326.35, 329.17, 332.01, 334.87, 337.75, 340.65, 343.58, 346.53, 349.51, 352.50, 355.51, 358.55, 361.61, 364.69, 367.79, 370.92, 374.07, 377.25, 380.44, 383.67, 386.91, 390.18, 393.47, 396.79, 400.15, 403.52, 406.91, 410.33, 413.77, 417.24, 420.74, 424.27, 427.81, 431.39, 434.99, 438.62, 442.29, 445.98, 449.70, 453.43 K. Production REMD simulations were performed in the constant-temperature/constant volume ensemble; replica exchange attempt frequency was 4 ps (2,000 steps). During the REMD runs, the trajectory in each replica was saved every 20 ps.

SM5: Calculation of the fraction of native contacts

We computed the fraction of native contacts, Q, using Eq. 13 of Ref. 28 with a cutoff of 8 Å excluding $i-j <= 4$ where i and j are residue indices.[28] For completeness, we reproduce Eq. 13 below:

$$Q = \frac{\sum_{i<j-1} \theta(r_c - r_{ij}^N) \exp\left[-\frac{(r_{ij}-r_{ij}^N)^2}{2\sigma_{ij}^2}\right]}{\sum_{i<j-1} \theta(r_c - r_{ij}^N)}. \quad (1)$$

In the above expression, r_{ij}^N and r_{ij} denote distances between Cα atoms of residues i and j in the native (N) and current structures, respectively; θ is the Heaviside step function; $r_c = 8$ Å; $\sigma_{ij} = |i-j|^{0.15}$ Å.

SM6: Calculation of the free energy landscapes

REMD method enhances conformational sampling efficiency such that the thermodynamic equilibrium can be reached much faster than in a conventional MD simulation.[25] For example, in our REMD simulation of WW domain folding performed using ff99cufix, the equilibrium was reached after $\sim$2 μ from the beginning of the simulation, Fig. S2 and Fig. S3 . Once the equilibrium ensemble is reached, the free energy landscape can be computed by counting the number of microscopic configurations having the prescribed value of the reaction coordinate.

The free-energy landscapes of WW domain and HP35 proteins were obtained by analyzing 25,000 microscopic configurations per each temperature replica recorded during the last

0.5 μs of the respective REMD simulation. The 25,000 configurations were assigned to one of the Q value bins, yielding the number of microscopic states having the specified Q, $n(Q)$. The free energy landscape, $\Delta G(Q)$, was computed by taking the Boltzmann inversion of $n(Q)$:[29]

$$\Delta G(Q) = -k_{\mathrm{B}}T \log n(Q), \tag{2}$$

where k_{B} is the Boltzmann constant. The free energy landscapes for other reaction coordinates or higher dimensions were obtained using a similar method.

SM7: Calculation of position-dependent diffusion coefficients

The REMD trajectory of each replica at a given temperature is unsuitable for the calculation of the diffusion coefficients $D(Q)$ because frequent (every tens of picoseconds) swapping of the coordinates with other replicas makes the trajectory at a given temprerature discontinuous. To compute the diffusion constant, we reorganized the 56 discontinuous, same temperature trajectories into 56 continuous, variable temperature trajectories using the demux.pl program of the Gromacs package.[26] Using these continuous trajectories, we computed the dependence of the mean square displacement (MSD) of Q on the time differential Δt. During the calculation of MSD, we discarded the fragments of the trajectories at temperatures exceeding 334 K, limiting our analysis to experimentally relevant conditions; the full range of temperatures explored in REMD simulations was 284 to 453 K. Then, $D(Q)$ was obtained by first determining the slope of the MSD dependence on Δt within the 500 ps to 1 ns range and then dividing the slope by 2 (MSD = $2D\Delta t$).

SM8: Brownian dynamics (BD) simulation

To compute folding rates, we performed Brownian dynamics (BD) simulations along one-dimensional reaction coordinate Q using the algorithm of Ermak and McCammon:[30–32]

$$Q(t + \Delta t) = Q(t) + \frac{D}{kT} F \Delta t + S, \tag{3}$$

where $F(Q)$ and $D(Q)$ are the thermodynamic force and the diffusion coefficient, respectively, extracted from REMD simulations and $S(Q)$ a random number generated from a normal distribution having the width of $2D(Q)\Delta t$.[31] The force $F(Q)$ was computed as a numerical derivative of the 1 D free energy landscape $G(Q)$ of the 297 K replica: $F(Q) = -dG(Q)/dQ$. For each system, we performed 1 million independent BD simulations starting from $Q = 0$. For each BD simulation, the folding time was estimated as the time elapsed from the beginning of the simulation until the Q value exceeded 0.9 for the first time.

SM9: MD simulations of Ala_5 and Val_3 peptides

To create 50 mM solutions of Ala_5 or Val_3 peptides, 30 copies of either Ala_5 or Val_3 peptides were randomly places in a 10 nm cubic box. Each peptide system was solvated with 30,000 TIP3P water molecules. Note that the reference experimental data show that the J-couplings are not affected by the peptide concentration in the 0.02 to 88 mM range.[33] The simulations were carried out using the Gromacs 5.0.4 package and the AMBER ff99sb-ildn-phi force field. ϕ and ψ angles were computed using g_rama program of the Gromacs package. J-couplings based on the Karplus equation[34] were computed using the parameters presented in Table S2 of the reference 33.

SM10: MD simulations of ubiquitin

The initial coordinates of a 76 amino acid ubiquitin protein were taken from the PDB database (PDB: 1UBQ).[35] For the only histidine residue, His68, a proton was assigned to its

NE2 atom upon examination of the hydrogen bond patterns using the pdb2gmx program of the Gromacs package.[1] The ubiquitin structure was solvated in a rhombic dodecahedron unit cell; the distance between the image copies of the protein was 55 Å ($a = b = c =\sim 55$ Å, $\alpha = \beta = 60°$, $\gamma = 90°$). Na^+ and Cl^- ions were randomly added to a concentration of 150 mM. The final system contained 3422 water molecules and 10 Na^+ and 10 Cl^- ions. The simulation of each system began with a 1,000 step minimization, followed by a 500 ps equilibration with all heavy atoms restrained to the initial coordinates using a harmonic constant of $k = 1{,}000$ kJ/(mol·nm^2). Production simulations were performed at constant pressure of 1 bar and constant temperature of 300 K. During the production runs, trajectories were saved every 2 ps.

Supporting Figures

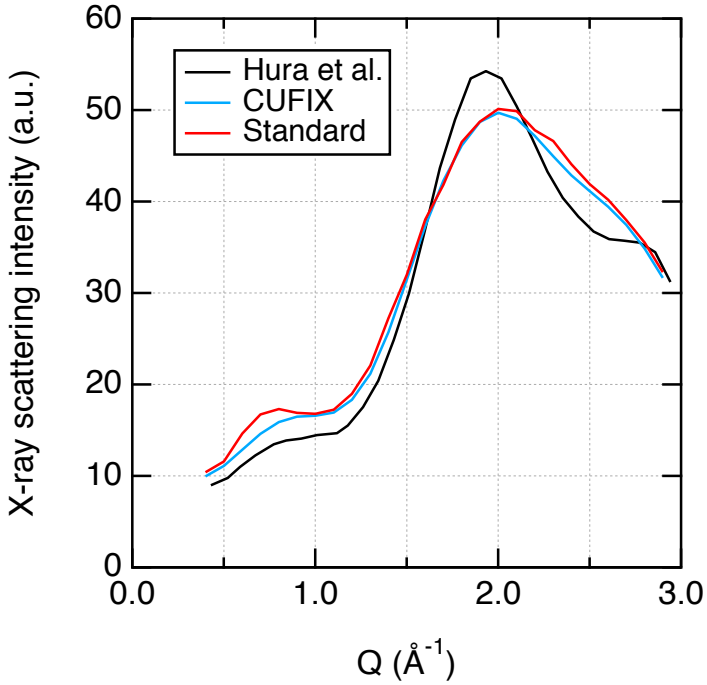

Figure S1: **The effect of CUFIX corrections on simulated X-ray scattering intensity of a 1 M NALMA solution.** X-ray scattering intensity curves computed from MD trajectories of 1 M NALMA solution simulated with (blue) and without (red) the CUFIX corrections for the CT atom types. The corresponding experimental data (black) were taken from Hura et al.[36] The X-ray scattering intensity curves were computed using Eq. 2 from Ref. 36. According to Hura et al.,[36] the peak at $Q = 0.8$ Å observed for the standard model indicates artificial aggregations of NALMA molecules and is inconsistent with experiment.[36]

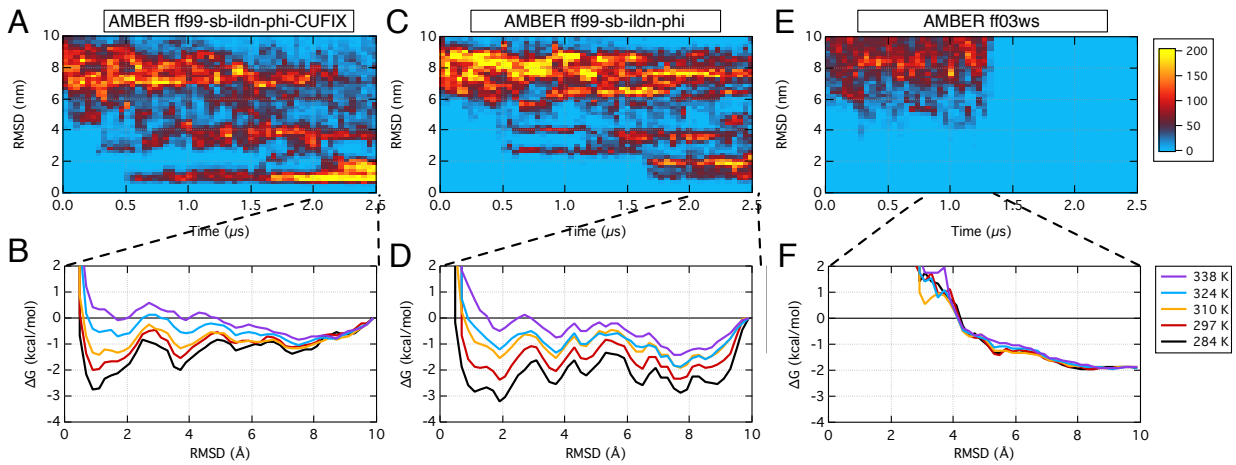

Figure S2: **REMD simulation of WW domain folding.** (A,B) REMD simulation of WW domain folding performed using ff99cufix. (A) The number of microscopic states observed in the 297 K replica (heat map) having the prescribed RMSD from the native structure as a function of the simulation time. As the simulation progresses, the folded conformations (RMSD < 2 Å) become dominant. The heatmap was made by splitting the 20 ps sampled trajectory of the 297 K replica into 30 ns intervals each containing 1,500 microscopic configurations and binning the configurations according to their RMSD values with a 0.05 interval. The RMSD value was computed using coordinates of the C_α atoms excluding the two residues at each terminal of the protein. (B) The free energy landscapes of the WW domain *versus* RMSD from the native structure at five representative temperatures. The free energy landscape was obtained by binning 25,000 microscopic configurations observed during the last 0.5 μs of the REMD simulation according to their RMSD values with a 0.02 interval. (C,D) Same as in panels A and B, respectively, for the REMD simulation of WW domain folding performed using ff99. (E,F) Same as in panels A and B, respectively, for the REMD simulation of WW domain folding performed using ff03ws.

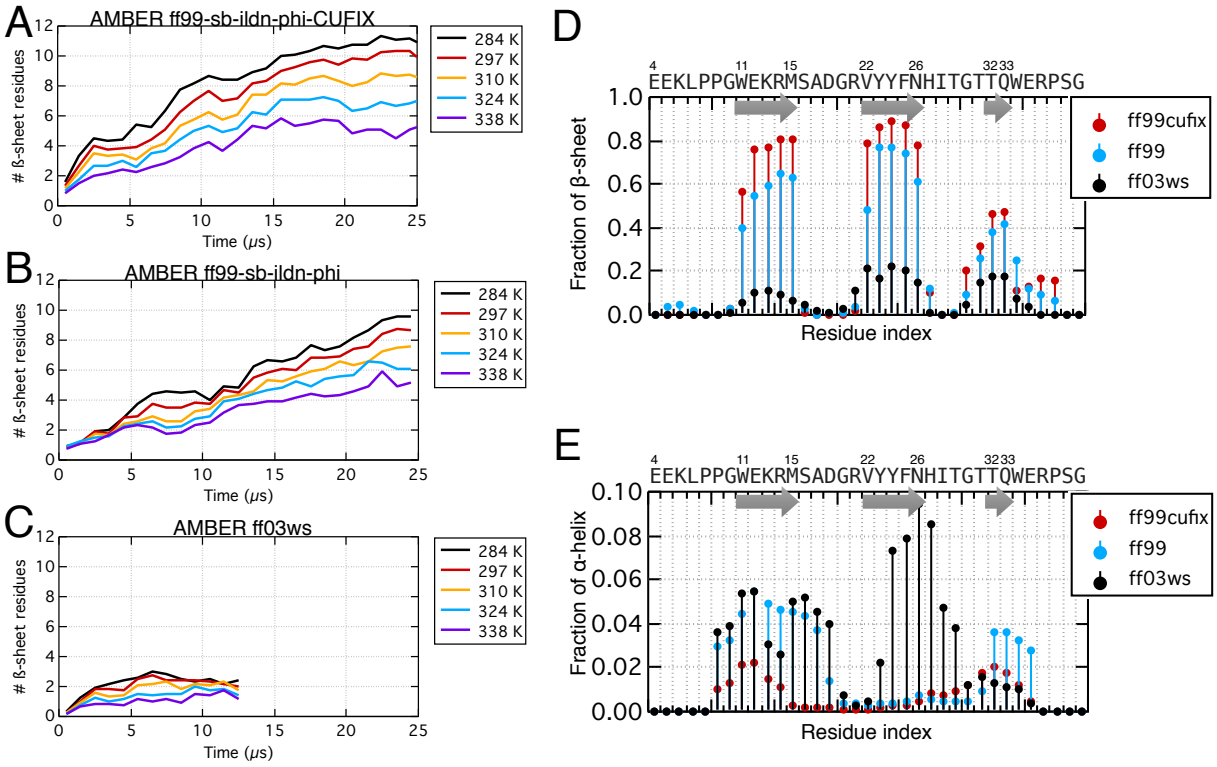

Figure S3: **Secondary structure analysis of WW domain folding.** (A–C) The number of β-sheet residues *versus* simulation time at five representative temperatures of the WW domain system obtained using ff99cufix (A), ff99 (B), and ff03ws (C). The secondary structure type of each residue was determined using the database of secondary structure assignments (DSSP) program.[37] The plots show 2 ns block averages of 20 ps sampled trajectories. Note that the number of β-sheet residues in the experimental structure[23] is 12. (D,E) The fraction of β-sheet (D) and α-helix (E) of each residue in the 297 K replica averaged over the 2.0 to 2.5 μ fragment of the REMD trajectory. Arrows indicate the residues that are part of a β-sheet fold in the native protein structure. The amino acid sequence and the residue index of the protein are shown at the top of the plots.

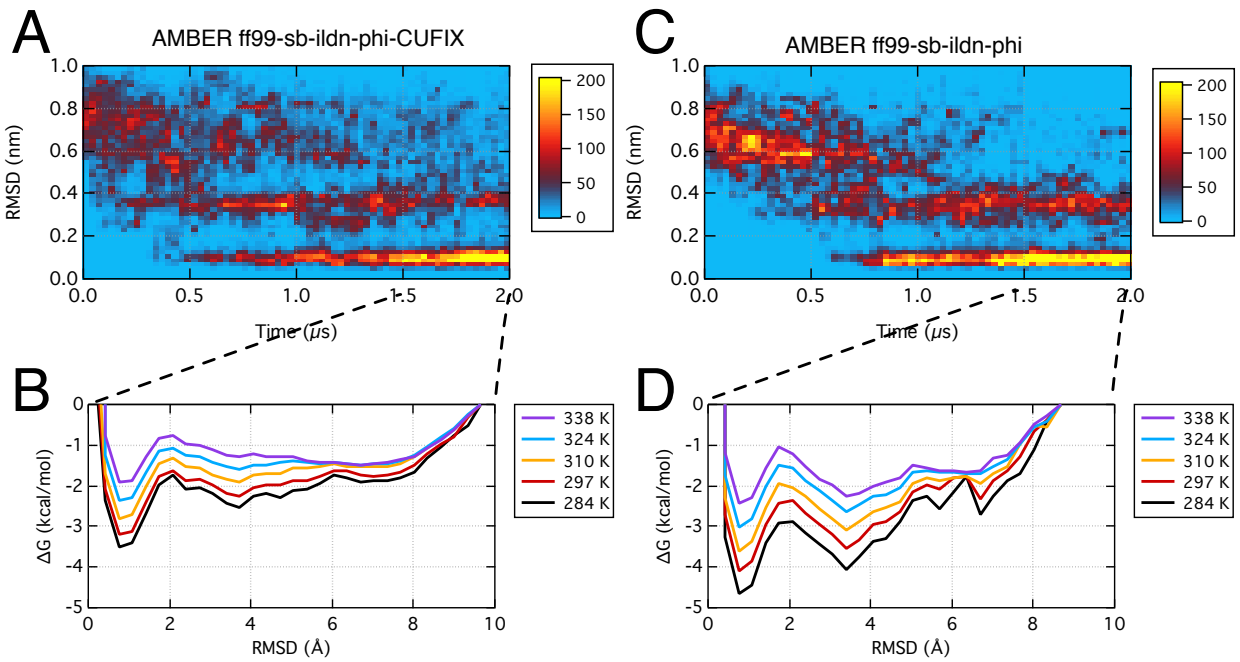

Figure S4: **REMD folding simulations of HP35.** (A,B) REMD simulation of HP35 folding performed using ff99cufix. (A) The number of microscopic states observed in the 297 K replica (heat map) having the specified RMSD from the native structure is shown as a function of the simulation time. As the simulation progresses, the folded conformations (RMSD $<$ 2 Å) become dominant. The heatmap was made by splitting the 20 ps sampled trajectory of the 297 K replica into 30 ns intervals each containing 1,500 microscopic configurations and binning the configurations according to their RMSD values with a 0.05 interval. The RMSD value was computed using coordinates of the C_α atoms excluding the two residues at each terminal of the protein. (B) The free energy landscapes of the HP35 *versus* RMSD from the native structure at five representative temperatures. The free energy landscape was obtained by binning 25,000 microscopic configurations observed during the last 0.5 μs of the REMD simulation according to their RMSD values with a 0.02 interval. (C,D) Same as in panels A and B, respectively, but for the REMD simulation of HP35 folding performed using ff99.

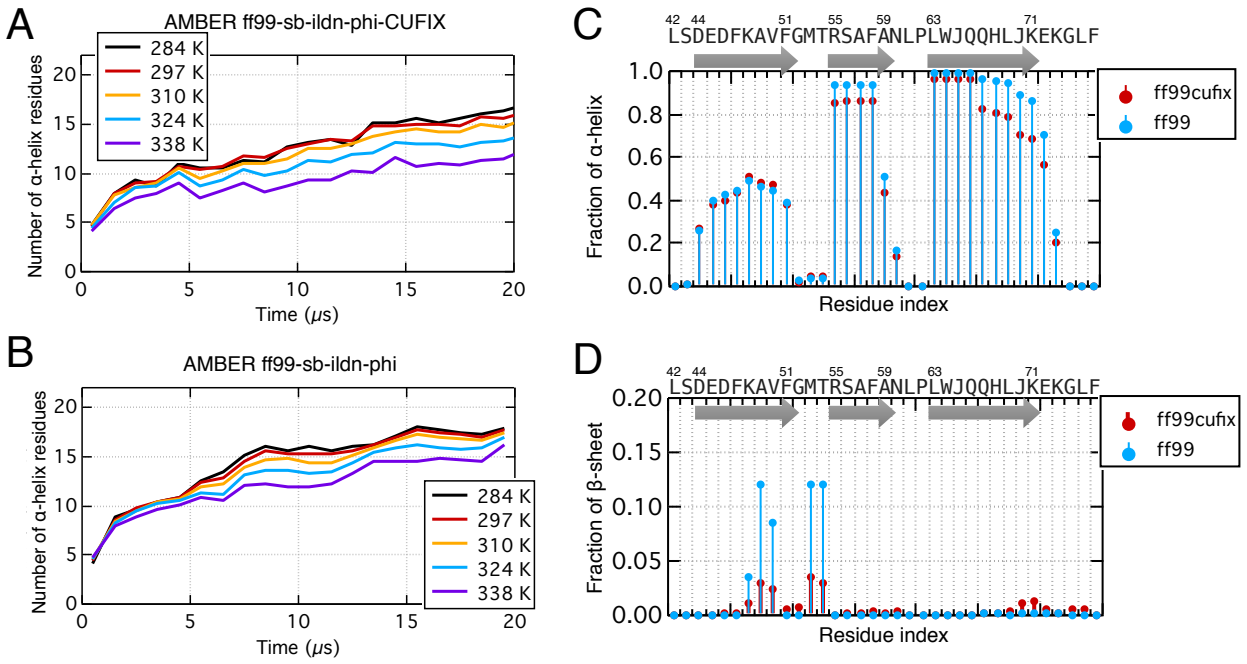

Figure S5: **Secondary structure analysis of HP35 folding.** (A,B) The number of residues forming an α-helix fold *versus* simulation time at five representative temperatures of the HP35 system obtained using ff99cufix (A) and ff99 (B). The secondary structure type of each residue was determined using the database of secondary structure assignments (DSSP) program.[37] The plots show 2 ns block averages of 20 ps sampled trajectories. Note that the number of residues forming an α-helix fold in the experimental structure[24] is 22. (C,D) The fraction of α-helix (C) and β-sheet (D) of each residue in the 297 K replica averaged over the 1.5 to 2.0 μ fragment of the REMD trajectory. Arrows indicate the residues that form α-helices in the native protein structure. The amino acid sequence and the residue index of the protein are shown at the top of the plots.

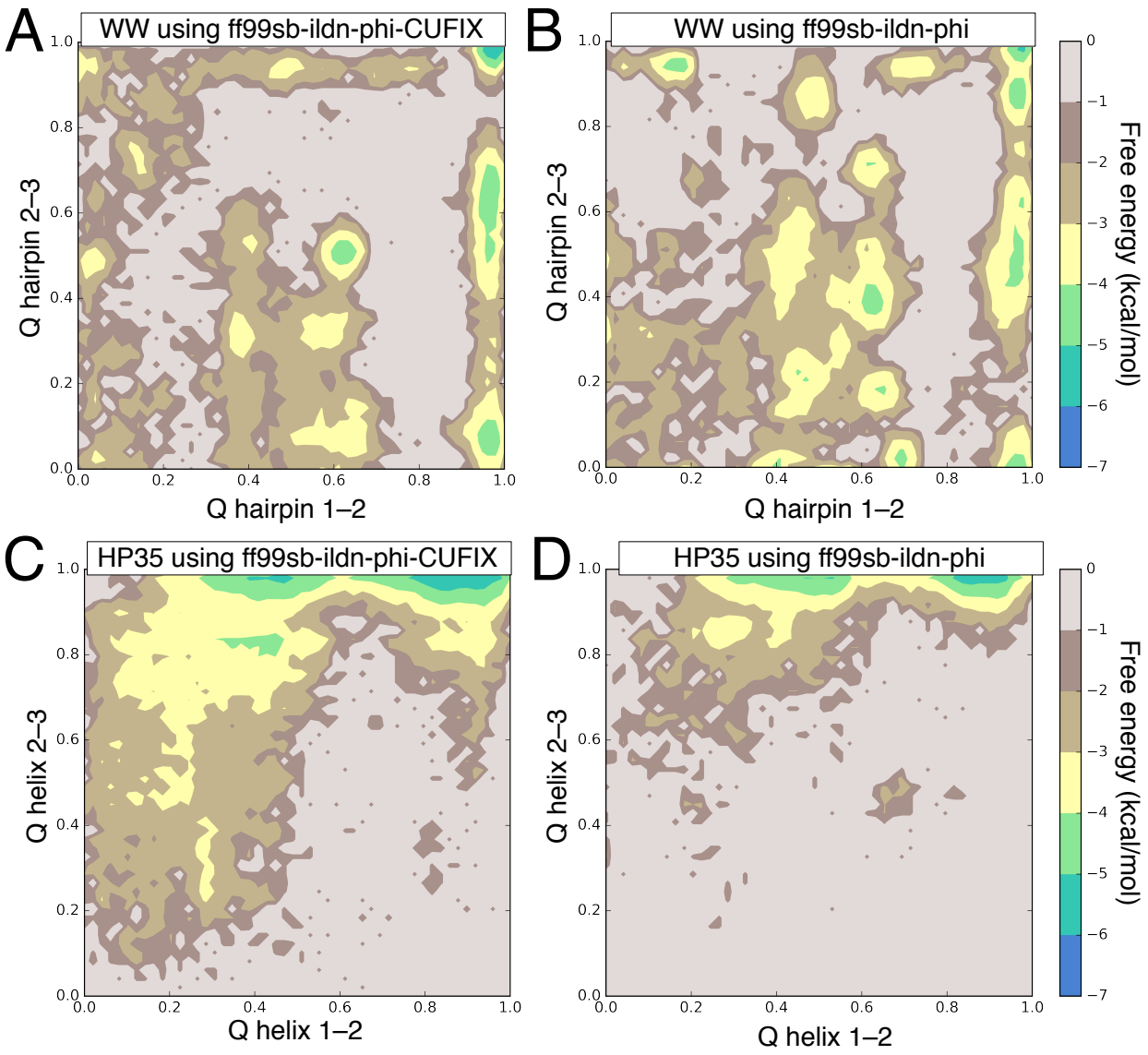

Figure S6: **The effect of CUFIX corrections on the 2D free energy landscapes.** (A,B) The free energy of the WW domain at 297 K as a function of the fraction of native contacts within the first and second hairpins (Q^{β}_{1-2} and Q^{β}_{2-3}, respectively) in REMD simulations performed using ff99cufix (A) and ff99 (B). Q^{β}_{1-2} and Q^{β}_{2-3} were computed using C_{α} coordinates of residues $11-26$ and $22-34$, respectively. Each free energy landscape was computed by counting the populations of microscopic states in the 297 K replica having the prescribed values of Q^{β}_{1-2} and Q^{β}_{2-3} using the last 0.5 μ of the respective REMD simulation; the bin size was 0.02 for both variables. (C,D) Same as in panels A and B but for the REMD simulations of HP35 using the fraction of native contacts within and betwen helix 1 and 2 (Q^{α}_{1-2}) or helix 2 and 3 (Q^{α}_{2-3}). Q^{α}_{1-2} and Q^{α}_{2-3} were computed using C_{α} coordinates of residues $44-59$ and $55-71$, respectively.

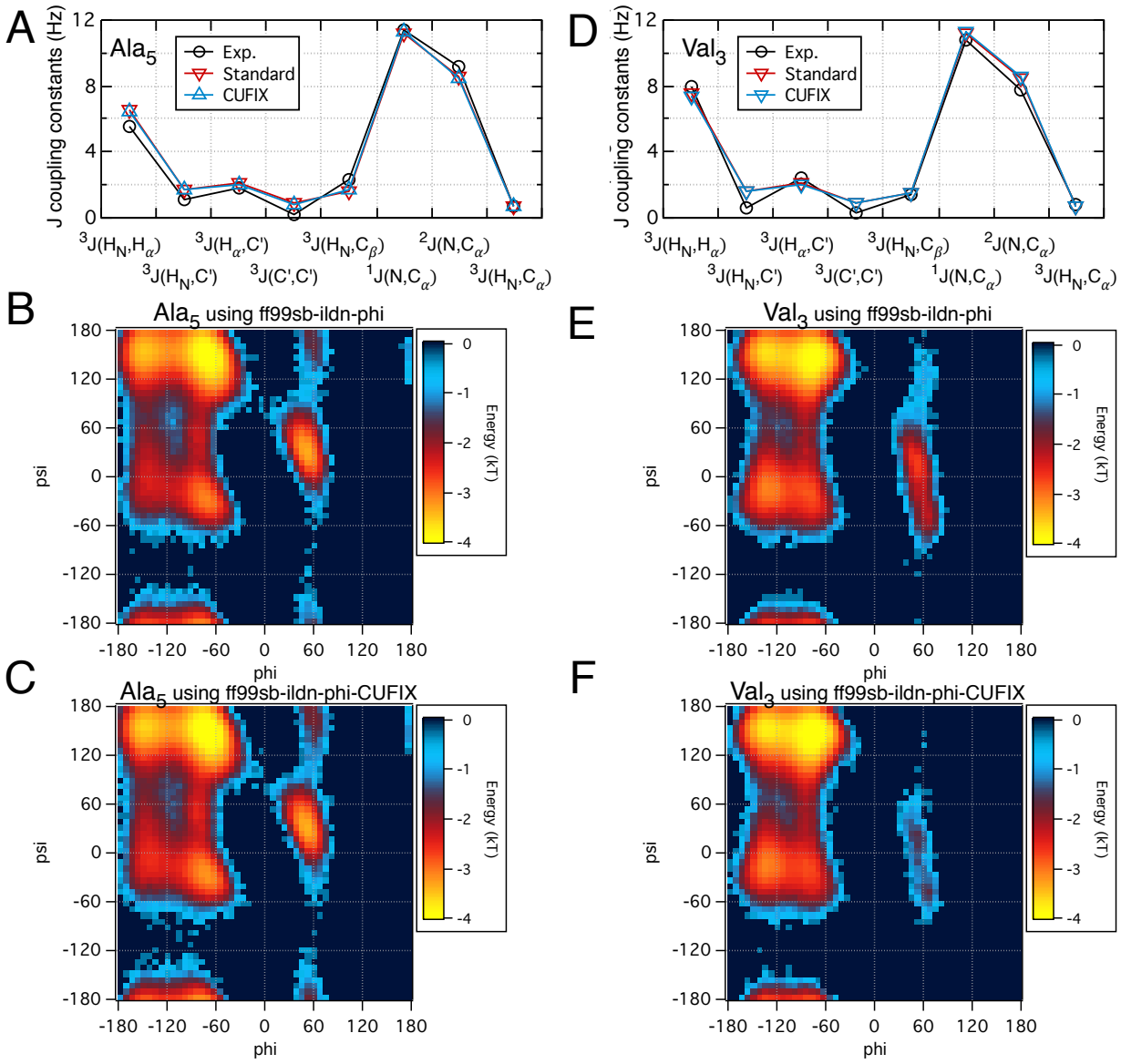

Figure S7: **The effect of CUFIX corrections on backbone conformations of Ala_5 and Val_3 peptides.** (A) Comparison of the simulated J coupling constants for the second residue of Ala_5 with experiment.[33] (B,C) Ramachandran plots for the second residue of Ala_5 computed using ff99 (B) and ff99cufix (C). (D) Comparison of the simulated J coupling constants for the second residue of Val_3 with experiment.[33] (E,F) Ramachandran plots for the second residue of Val_3 computed using ff99 (E) and ff99cufix (F). The error bars (standard error) of the simulated J coupling values in panels A and D are smaller than the symbols.

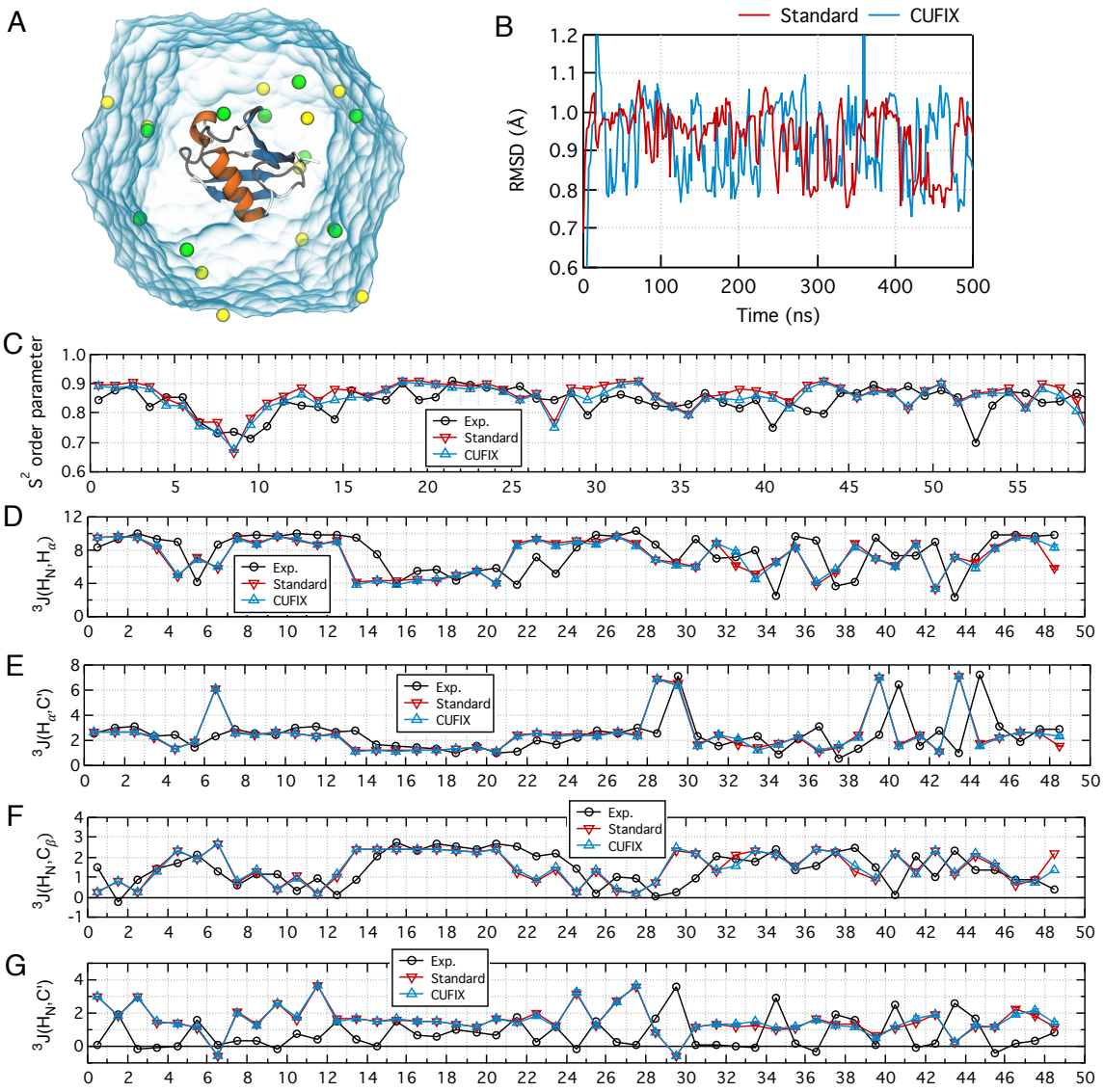

Figure S8: **The effect of CUFIX corrections on MD simulations of ubiquitin.** (A) A unit periodic cell of the simulation system containing a single copy of a 76-residue ubiquitin protein (PDB: 1UBQ[35]) solvated in a rhombic dodecahedron volume of 150 mM aqueous solution of NaCl (semi-transparent surface). Ubiquitin is shown using a cartoon representation; Na^+ and Cl^- ions are shown as yellow and green spheres, respectively. (B) Root-mean-square deviation (RMSD) of the ubiquitin's coordinates from the crystal structure during the equilibration simulations performed with and without our CUFIX corrections to the ff99sb-ildn-phi force fields. Because the C-terminal domain of ubiquitin is unstructured,[21] the RMSD calculation were performed for the non-hydrogen backbone atoms of residues 1–71. The RMSD curves were computed using MD trajectories sampled every 2 ps; 2-ns block average of the data is shown. (C) Experimentally measured[38] and simulated Lipari-Szabo order parameter (S^2) of ubiquitin at 300 K. The experimental isotropic values of the order parameter were taken from the NMR study. (D–G) Experimentally measured[39] and simulated J-couplings of ubiquitin at 300 K: $^3J(H_N,H_\alpha)$, D; $^3J(H_\alpha,C')$, E; $^3J(H_N,C_\beta)$, F; $^3J(H_N,C')$, G.

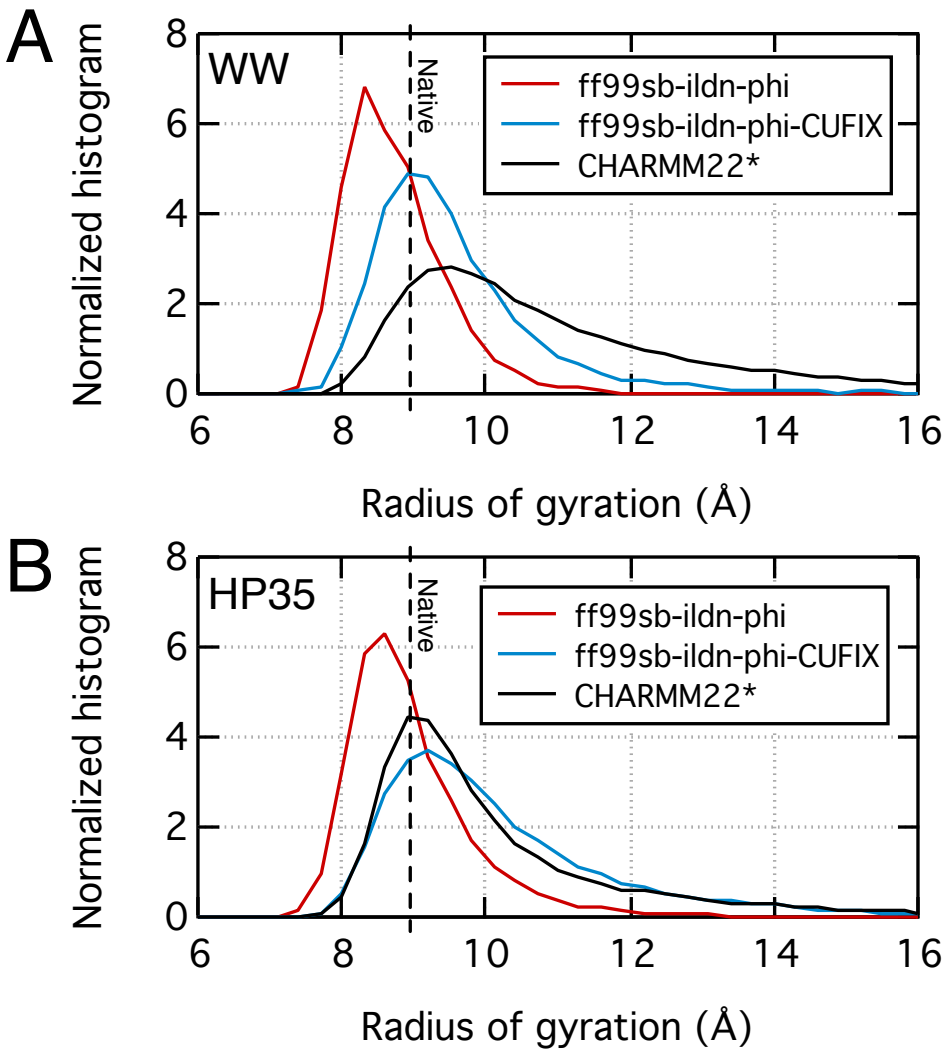

Figure S9: **The effect of CUFIX corrections on radius of gyration (R_g) of denatured conformations.** (A,B) Distribution of R_g of denatured conformations ($Q < 0.5$) at 360 K for WW domain (A) and HP35 (B). Trajectories using the CHARMM22* were taken from Ref. 40. Note that the first two amino acids (Gly_1 and Ser_2) of the GTT variant of WW domain in Ref. 40 differ from ours and the experimental structure (Glu_1 and Glu_2).

References

(1) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(2) Nose, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055–76.

(3) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(4) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–90.

(5) Darden, T. A.; 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.

(6) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Molecules. *J. Comput. Chem.* **1992**, *13*, 952–962.

(7) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–72.

(8) 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., Func., Bioinf.* **2006**, *65*, 712–25.

(9) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Func., Bioinf.* **2010**, *78*, 1950–8.

(10) Nerenberg, P. S.; Head-Gordon, T. Optimizing ProteinSolvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J. Chem. Theory Comput.* **2011**, *7*, 1220–1230.

(11) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; 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*, 1999–2012.

(12) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.

(13) Yoo, J.; Aksimentiev, A. Improved Parametrization of Li^+, Na^+, K^+, and Mg^{2+} Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *J. Phys. Chem. Lett.* **2012**, *3*, 45–50.

(14) Yoo, J.; Aksimentiev, A. Improved Parameterization of Amine–Carboxylate and Amine–Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *J. Chem. Theory Comput.* **2016**, *12*, 430–443.

(15) Yoo, J.; Wilson, J.; Aksimentiev, A. Improved Model of Hydrated Calcium Ion for Molecular Dynamics Simulations Using Classical Biomolecular Force Fields. *Biopolymers* **2016**, *105*, 752–763.

(16) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(17) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

(18) Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(19) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47–9.

(20) Johnson, M. E.; Malardier-Jugroot, C.; Murarka, R. K.; Head-Gordon, T. Hydration Water Dynamics Near Biological Interfaces. *J. Phys. Chem. B* **2009**, *113*, 4082–4092.

(21) Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T. Optimizing Solute-Water van der Waals Interactions to Reproduce Solvation Free Energies. *J. Phys. Chem. B* **2012**, *116*, 4524–4534.

(22) Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. Computational Design and Experimental Testing of the Fastest-Folding β-Sheet Protein. *J. Mol. Biol.* **2011**, *405*, 43–8.

(23) Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. Structure-Function-Folding Relationship in a WW Domain. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10648–10653.

(24) Kubelka, J.; Chiu, T. K.; Davies, D. R.; Eaton, W. A.; Hofrichter, J. Sub-Microsecond Protein Folding. *J. Mol. Biol.* **2006**, *359*, 546–53.

(25) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(26) van der Spoel, D.; Seibert, M. M. Protein Folding Kinetics and Thermodynamics from Atomistic Simulations. *Phys. Rev. Lett.* **2006**, *96*, 238102.

(27) Patriksson, A.; van der Spoel, D. A Temperature Predictor for Parallel Tempering Simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073–2077.

(28) Eastwood, M.; Hardin, C.; Luthey-Schulten, Z.; Wolynes, P. Evaluating Protein Structure-Prediction Scheme Using Energy Landscape Theory. *IBM J. Res., Dev.* **2001**, *45*, 475–497.

(29) Duan, Y.; Kollman, P. A. Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution. *Science* **1998**, *282*, 740–744.

(30) Ermak, D. L.; McCammon, J. A. Brownian Dynamics with Hydrodynamic Interactions. *J. Chem. Phys.* **1978**, *69*, 1352.

(31) Elcock, A. H.; Sept, D.; ; McCammon, J. A. Computer Simulation of Protein-Protein Interactions. *J. Phys. Chem. B* **2001**, *105*, 1504–1518.

(32) Yang, S.; Onuchic, J. N.; García, A. E.; Levine, H. Folding Time Predictions from All-Atom Replica Exchange Simulations. *J. Mol. Biol.* **2007**, *372*, 756–63.

(33) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.

(34) Karplus, M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.* **1959**, *30*, 11–15.

(35) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8Å Resolution. *J. Mol. Biol.* **1987**, *194*, 531–544.

(36) Hura, G.; Sorenson, J. M.; Glaeser, R. M.; Head-Gordon, T. Solution X-ray Scattering as a Probe of Hydration-Dependent Structuring of Aqueous Solutions. *Perspect. Drug Discov.* **1999**, *17*, 97–118.

(37) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577–2637.

(38) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. Rotational Diffusion Anisotropy of Human Ubiquitin from 15N NMR Relaxation. *J. Am. Chem. Soc.* **1995**, *117*, 12562–12566.

(39) Wang, A. C.; ; Bax, A. Determination of the Backbone Dihedral Angles ϕ in Human Ubiquitin from Reparametrized Empirical Karplus Equations. *J. Am. Chem. Soc.* **1996**, *118*, 2483–2494.

(40) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.