Supporting Information for

CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins

Jing Huang,[1] Sarah Rauscher,[2] Grzegorz Nawrocki,[3] Ting Ran,[1] Michael Feig,[3] Bert L. de Groot,[2] Helmut Grubmüller,[2] Alexander D. MacKerell Jr.[1 *]

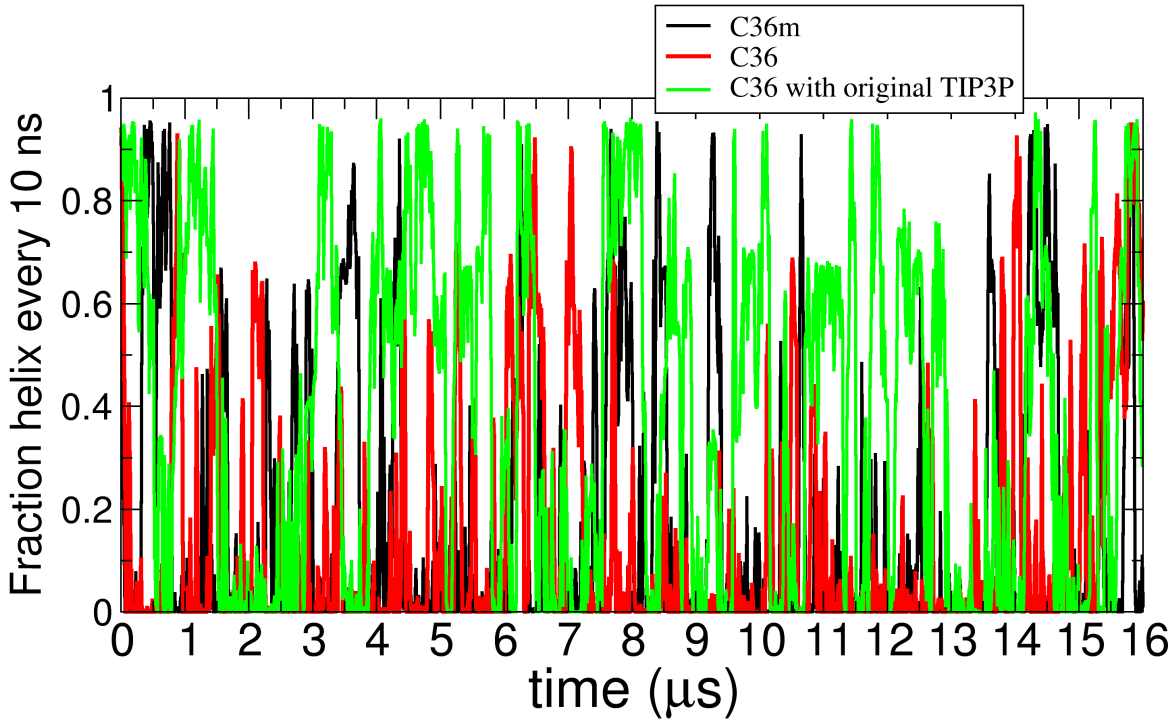
[1.] Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, 20 Penn St. Baltimore, MD 21201, USA

[2.] Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen 37077, Germany

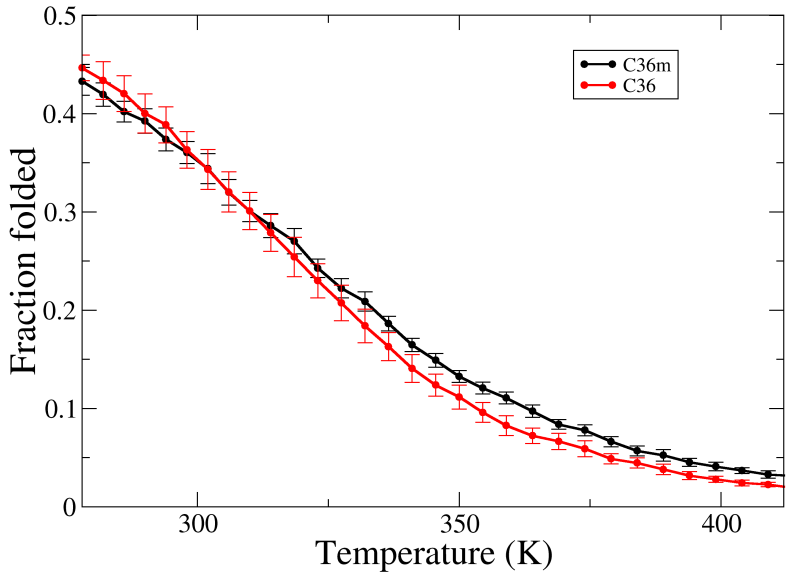
[3.] Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

[*] amackere@rx.umaryland.edu

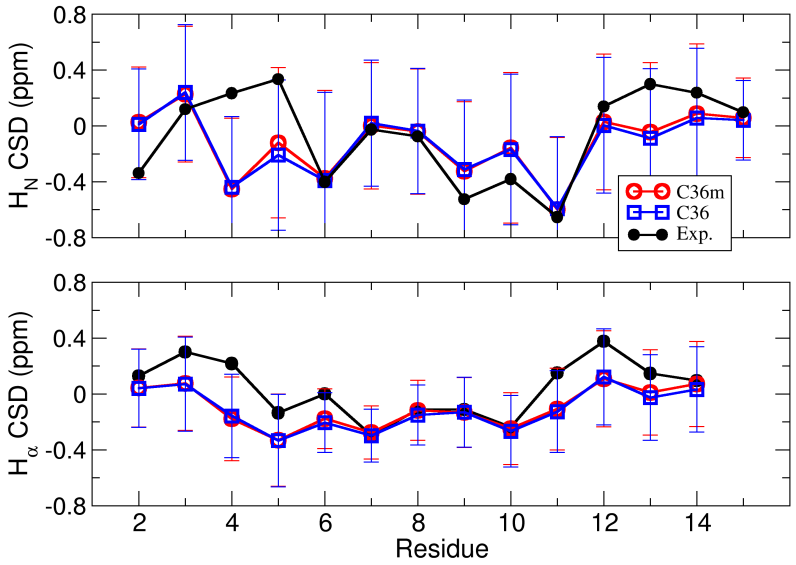
Supplementary Figure 1. Fraction helix of $(AAQAA)_3$ computed over 10 ns blocks from 16 μs MD trajectories generated by the C36m FF, the C36 FF, and the C36 FF with the original TIP3P water model. Tens of helix-coil transitions were observed during, indicating that converged thermodynamic properties were obtained.



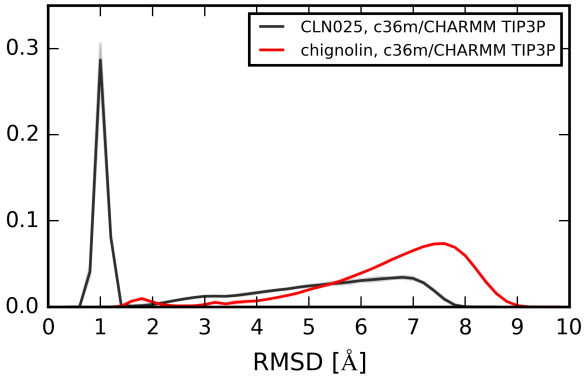
Supplementary Figure 2. The melting curves for the GB1 hairpin determined from T-REX simulations with the C36m (black) and the C36 (red) FFs. The folded states were identified as those with a Cα RMSD less than 2.0 Å compared to the pdb structure 1PGA.



Supplementary Figure 3. H_N (top panel) and H_α (bottom panel) chemical shift deviations for the GB1 hairpin at 278K from experimental measurement[1] (black) and T-REX simulations with the C36m (red) and C36 (blue) FFs. Error bars on calculated data are the typical RMSD between experiment and chemical shifts predicted by SPARTA+.

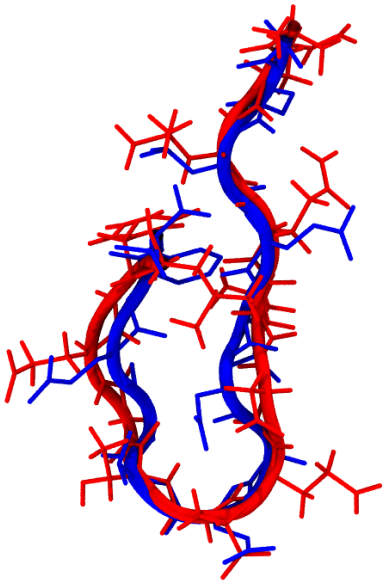


Supplementary Figure 4. RMSD histograms for Cln025 and Chignolin Hairpin Ensembles (backbone RMSD compared to the NMR structures - PDB 2RVD for cln025 and PDB 1UAO for chignolin). For CLN025, the folded state has a population of 41 %, and the chignolin hairpin has a population of 2.6 %. There is also a crystal structure available for cln025 (PDB 5AWL); comparing to this structure similarly gives a population of 41 % for the folded state. The folded state is defined using the low RMSD peak for each hairpin (cutoff of < 2.2 Å for chignolin and < 1.8 Å for cln025).

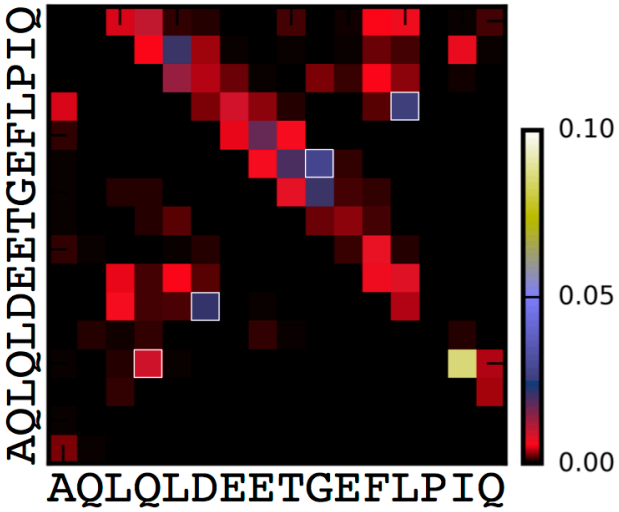


Supplementary Figure 5. (A) The Nrf2 hairpin conformation from the ensemble with the lowest RMSD (0.9 Å) is shown, with the crystal structure 2FLU in blue and the same 13 residues from the Nrf2 peptide shown in red. The folded hairpin is fully formed in 3 % of the MD ensemble generated with the C36m FF. There is no experimental data available for the population of folded states for the Nrf2 hairpin.[2] (B) The backbone hydrogen bonding contact map for the Nrf2 ensemble. The backbone hydrogen bonds formed in the folded hairpin (indicated by white outlined squares) are some of the most populated hydrogen bonds in the ensemble.

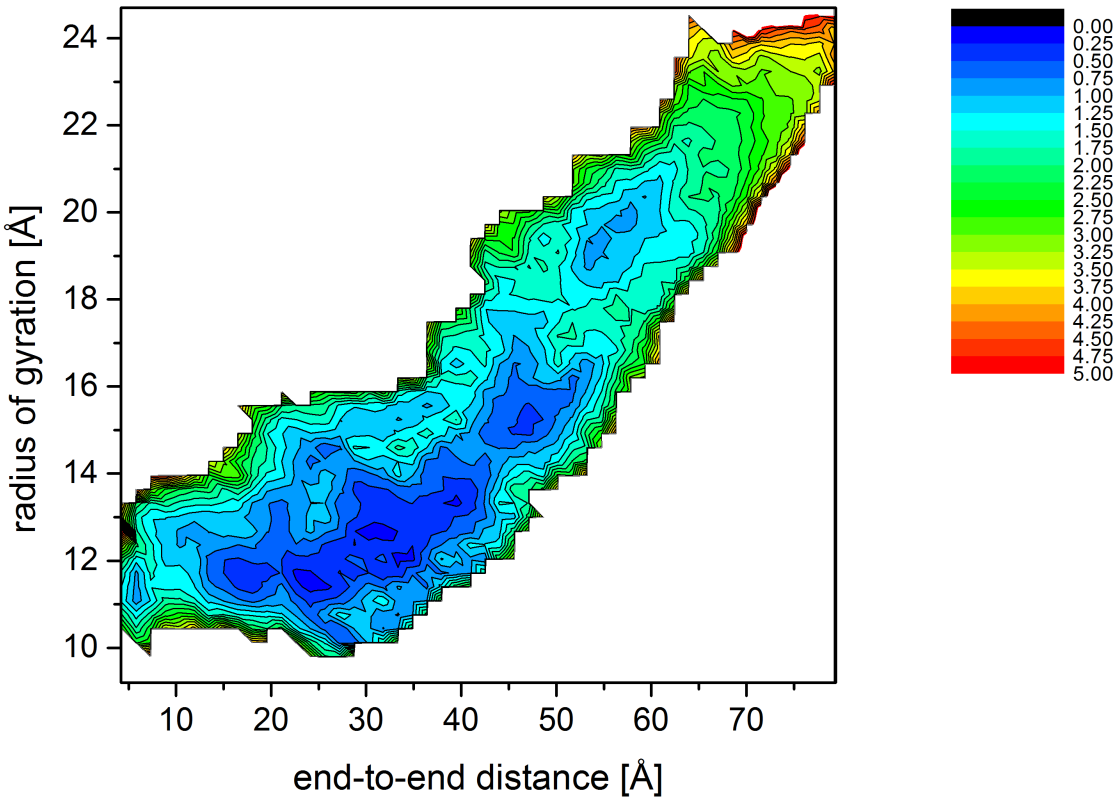
A



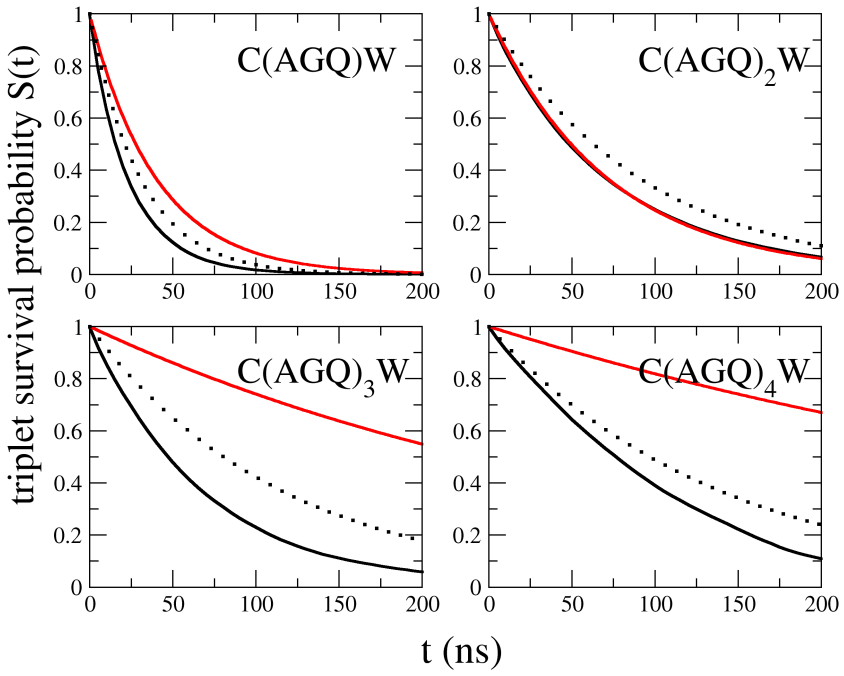
B



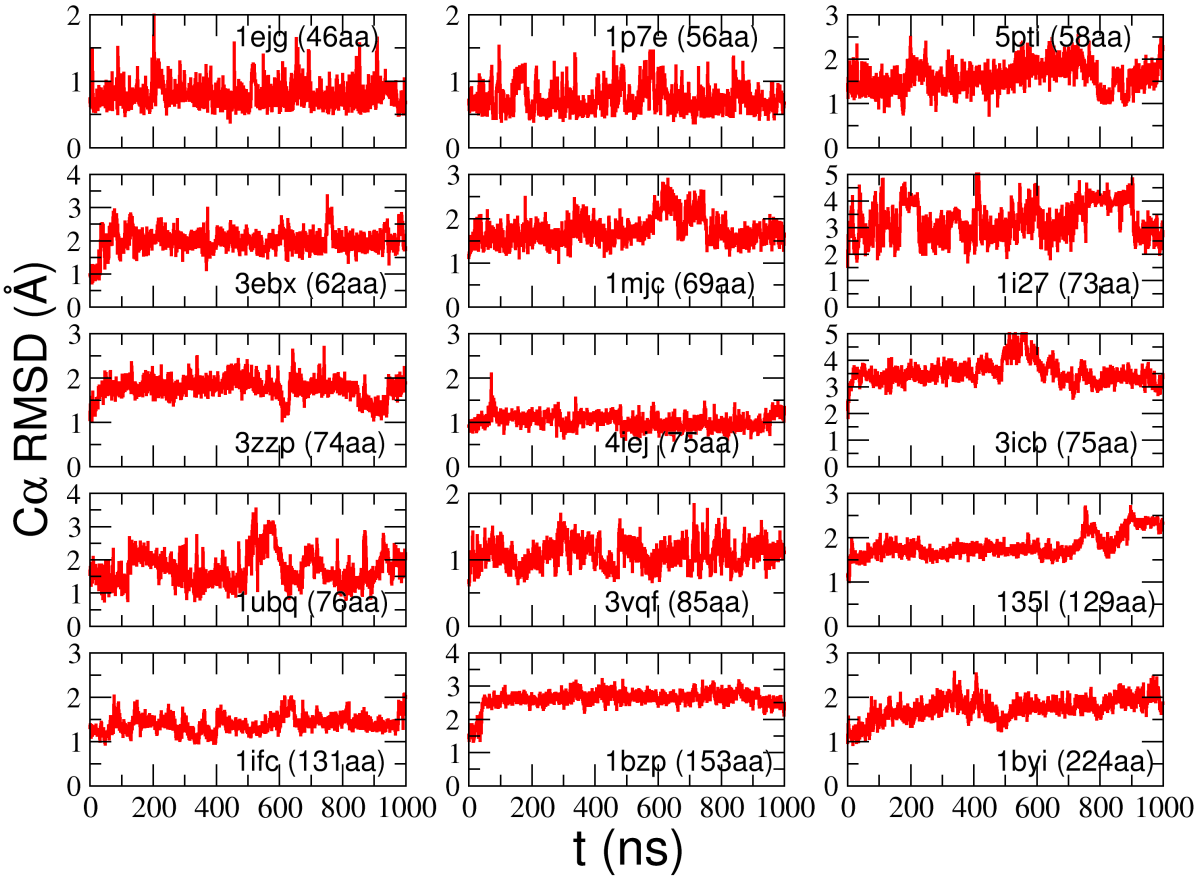
Supplementary Figure 6. The overall conformational sampling of the polyglutamine Q_{30} characterized by the potential of mean force (PMF) as a function of end-to-end distance and radius of gyration obtained via two-dimensional WHAM analysis from Hamiltonian replica exchange simulations using the C36m FF. The color bar indicates the energy in units of kcal/mol. Experimental data based on FRET analysis for peptides with inserted Q stretches reports end-to-end distances up to 24 glutamine residues.[3] Extrapolation of that data to Q_{30} suggests an end-to-end distance of 25 Å but since FRET analysis is not able to distinguish between states with very long end-to-end separations, this value should be more of a lower bound for the end-to-end distance of highly disordered peptides such as Q_{30}. Our simulation results with the C36m FF, where the first major minimum is at 25 Å, are, therefore, in excellent agreement with the experimental data, although we note that, as discussed in the analysis by Fluitt et al.[4], C36 generates ensembles that are broader than most other force fields. Integrating over the PMF with a Forster radius of 21.0 Å[3], the mean FRET efficiency <E> was calculated to be 0.20 ± 0.01 for Q_{30}, which compares favorably to the value of 0.22 estimated from linear extrapolation of experimental FRET efficiencies.[3]



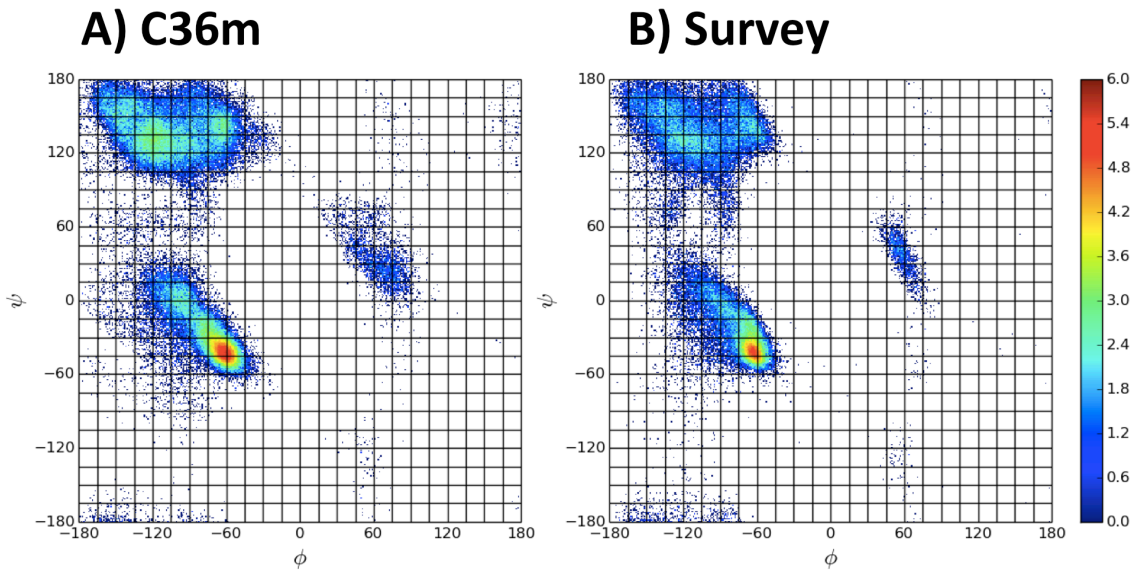
Supplementary Figure 7. Decays of tryptophan triplet state computed from MD simulations (black solid line) of the $C(AGQ)_nW$ peptides for n = 1 - 4, and from optical measurements (red solid line). The tryptophan triplet survival probability *S(t)* is computed as $S(t) = \langle exp\left[-\int_{t_0}^{t_0+t} q_c H(r_c - r(t'))dt'\right]\rangle_{t_0}$, where *r(t')* is the distance between the excited triplet and its quencher computed as the minimal distance between the cysteine sulfur atom and the tryptophan indole ring at time *t'*, with r_c = 4.0 Å, q_c =0.8 ns^{-1}, and *H* is the Heaviside step function.[5] The ensemble average is taken for every possible starting time t_0. Also shown is TIP3P viscosity corrected decay curves (black dotted lines) where the diffusion-limited part of quenching rate was scaled down by 2.87, the ratio of TIP3P water viscosity to experimental water viscosity.[5]



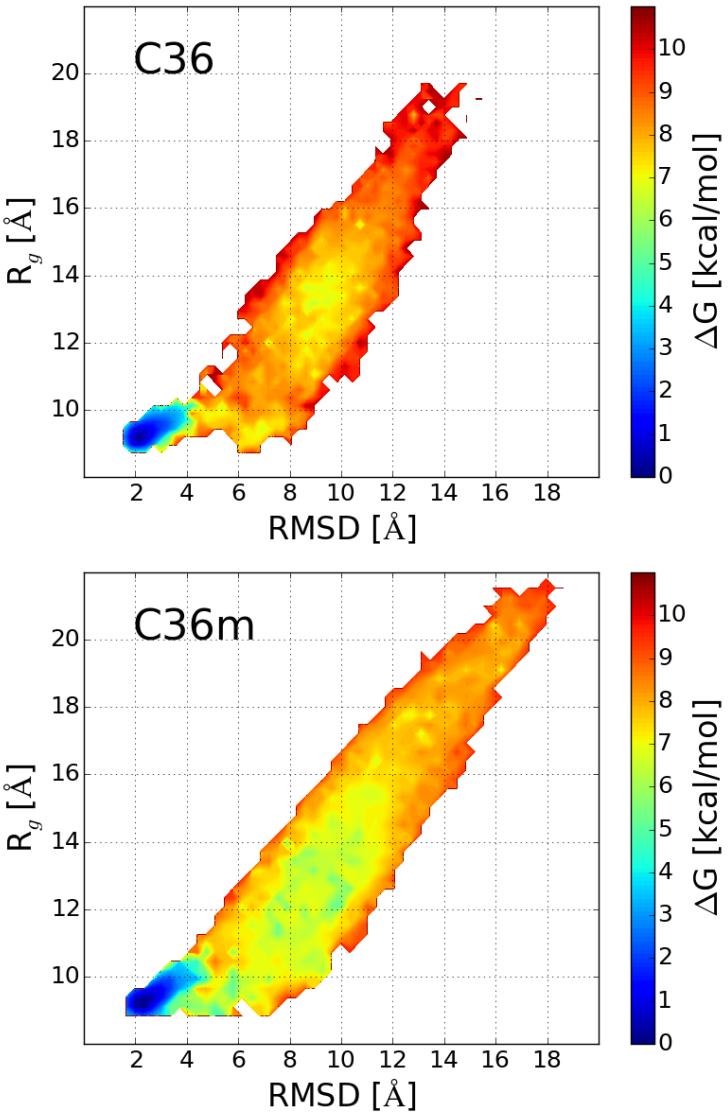
Supplementary Figure 8. Cα RMS Difference (RMSD) plots along 1 μs MD simulations for 15 proteins. The pdb id and number of amino acids for each protein are listed. All proteins are solvated in 150mM NaCl solution. Co-ions (Zn^{2+} for 1i27 and Ca^{2+} for 3icb) were not included in the simulation system, which might be related to the relatively large RMSD of 1i27 and 3icb. The Cα atoms of all residues in the proteins were included in the RMSD calculations.



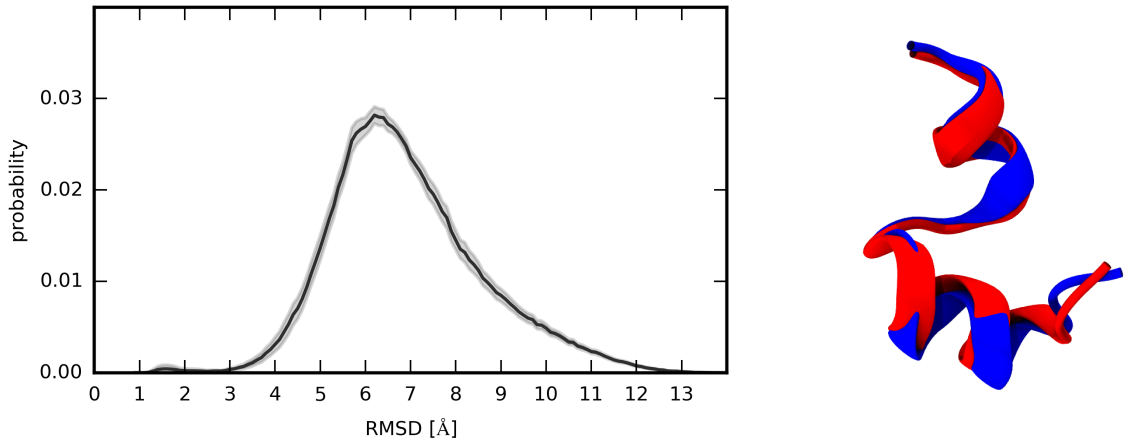
Supplementary Figure 9. ϕ, ψ distributions of non-Gly, non-Pro and *i*-1 non-Pro residues from MD simulations with the C36m FF (A) and from the "top500" pdb structures (B). The natural logarithm of the probability is plotted.



Supplementary Figure 10. The overall conformational sampling of the villin headpiece subdomain characterized by the 2D potential of mean force (PMF) as a function of RMSD and radius of gyration with the C36 and the C36m FF. The folded states were centered at RMSD=2.0 Å, R_g=9.5 Å, and the unfolded states at RMSD=8.5 Å, R_g=13 Å.

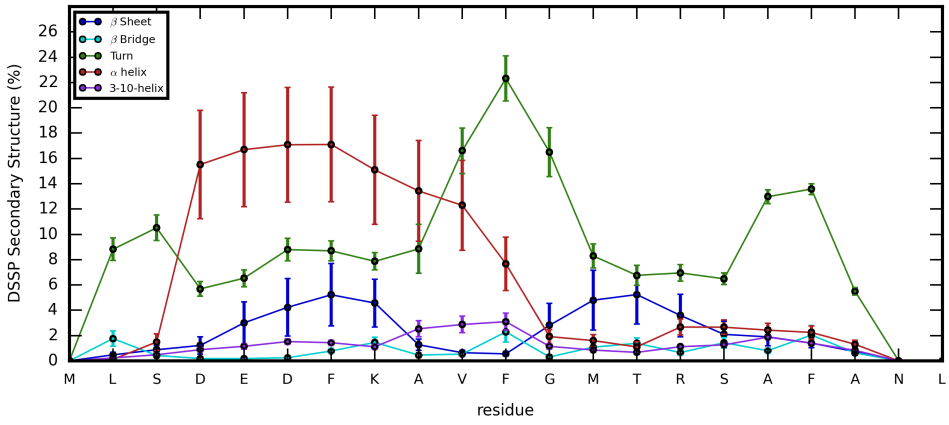


Supplementary Figure 11. The conformational sampling of HP21, the N-terminal fragment of villin headpiece subdomain. NMR studies have indicated the existence of α-helical structure in the ensemble of HP21[6,7] and a preference for native-like structure.[7] The ensemble obtained with the C36m FF contains a small population of the folded state as defined based on the backbone RMSD compared to the same fragment in the NMR structure of HP36 (pdb id: 1vii). The structure with the lowest RMSD (1 Å) is also shown (red cartoon) along with the NMR structure (blue cartoon). Shading indicates standard error of the mean.

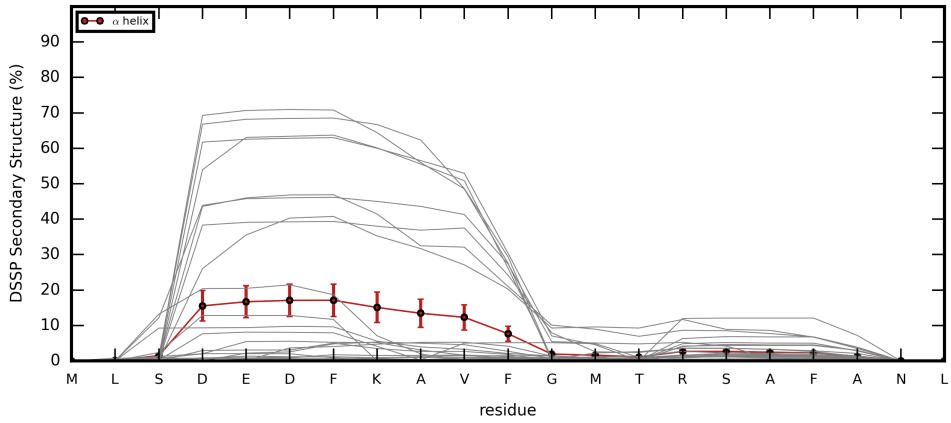


Supplementary Figure 12. The most populated type of secondary structure of HP21 is α-helix. Consistent with a recent simulation study,[8] the N-terminal helix is found to have a significantly higher population than the C-terminal helix. (A) The population of different types of secondary structure determined with DSSP[9] for each residue. (B) The population of helix for each of the 30 MD trajectories is shown in gray. Eight of them have a high population of helix.

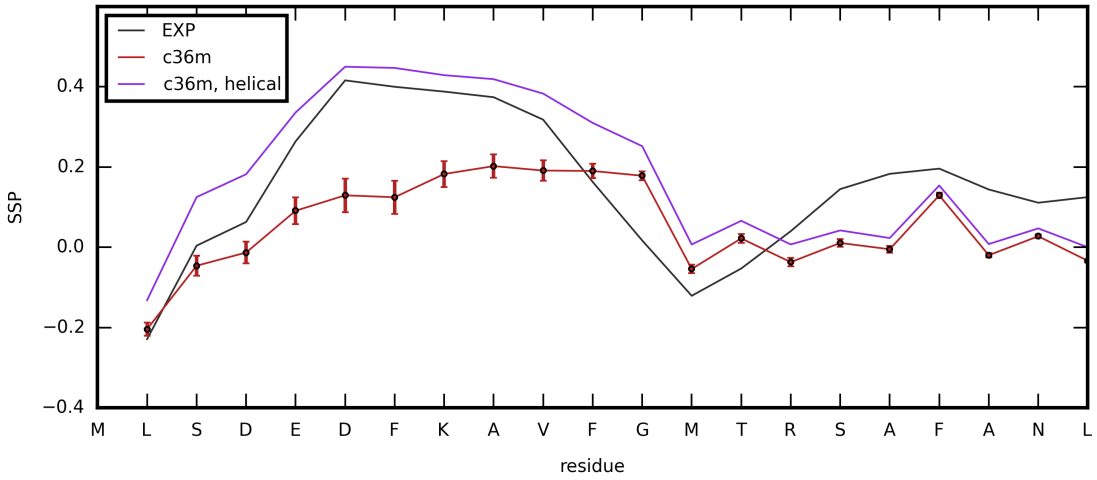
(A)



(B)

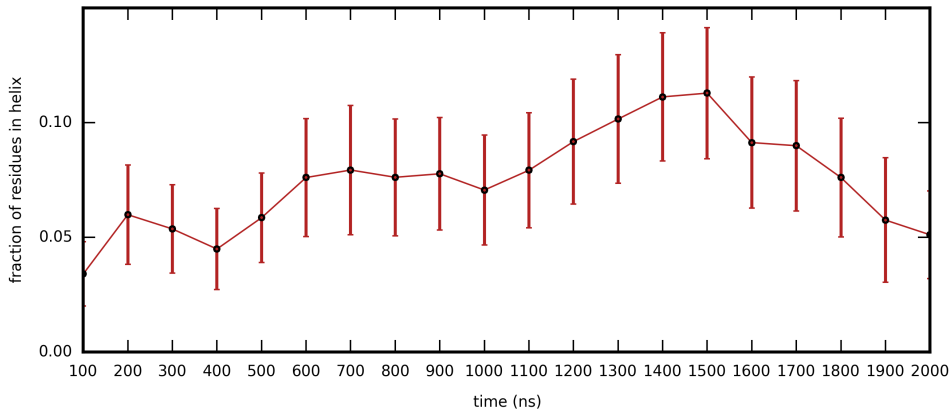


Supplementary Figure 13. Secondary structure propensity of HP21. For each residue, the secondary structure propensity (SSP)[10] is derived using the experimental chemical shifts (gray) and SHIFTX2[11] computed shifts for the C36m ensemble (red). To investigate the discrepancy between the experiment and simulation (Supplementary Table 14 and 15), SSP is computed for the eight MD trajectories that have a high population of helix (purple), and better agreement with experimentally derived SSP is observed. The population of helix in this sub-ensemble is 21.96%, compared to 7.47% in the complete ensemble. Since this helical sub-ensemble is in reasonable agreement with the experimental measurement, we estimate that the helical state is destabilized in the C36m ensemble by 1.08 *kT*.

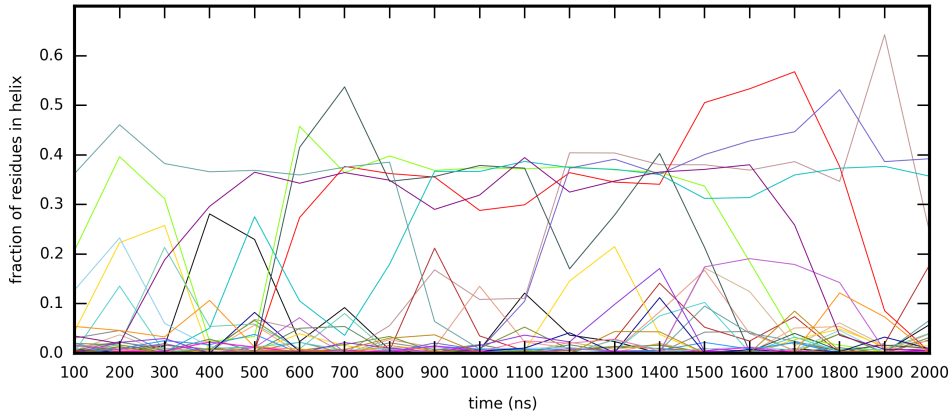


Supplementary Figure 14. Fraction of residues in α-helix for HP21. (A) The fraction of residues in α-helix is shown for twenty 100 ns windows of the trajectory. The average of all thirty replicas is plotted in red, and error bars indicated standard error of the mean. (B) The fraction of residues in α-helix is plotted for each of the thirty replicas individually, each with a unique color. Many transitions are observed between helical/non-helical states, indicating sufficient sampling of conformational spaces.

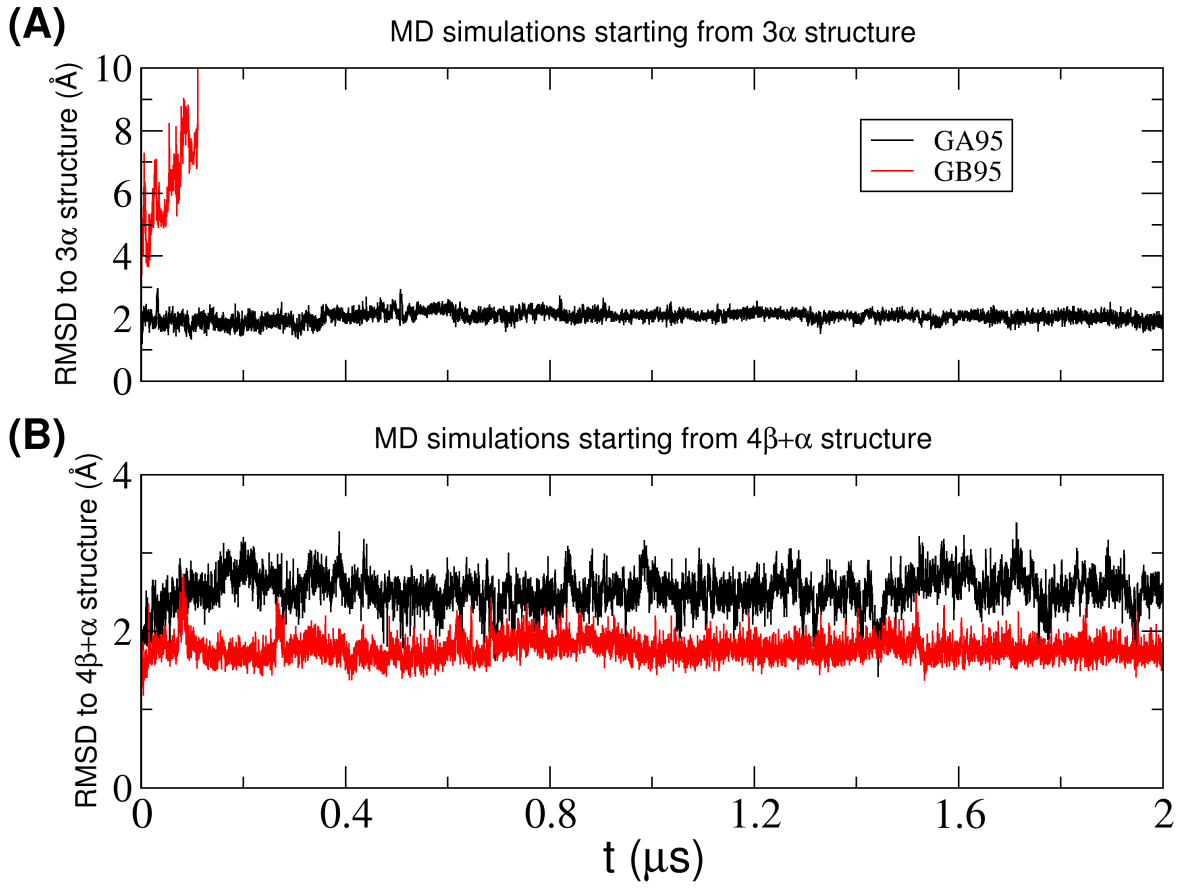
(A)



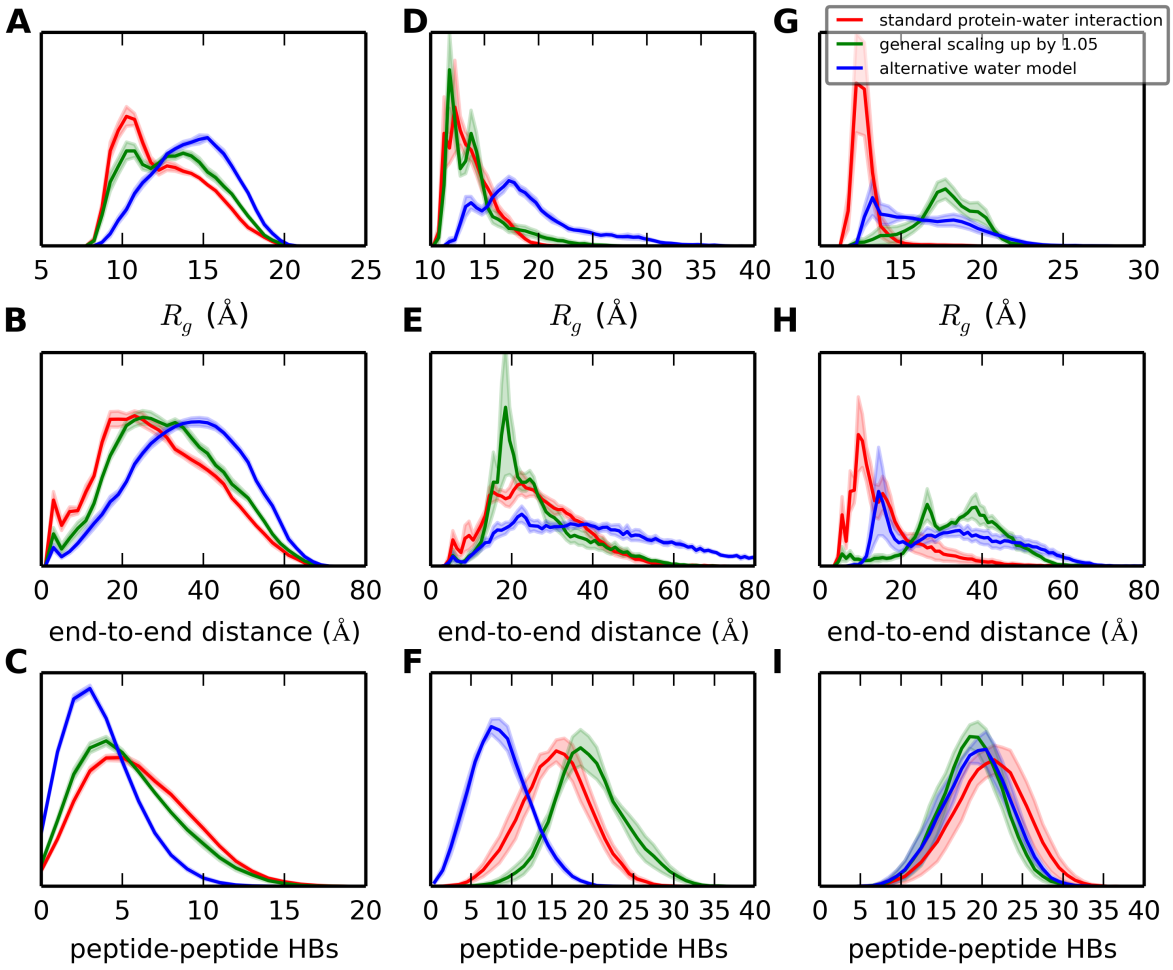
(B)



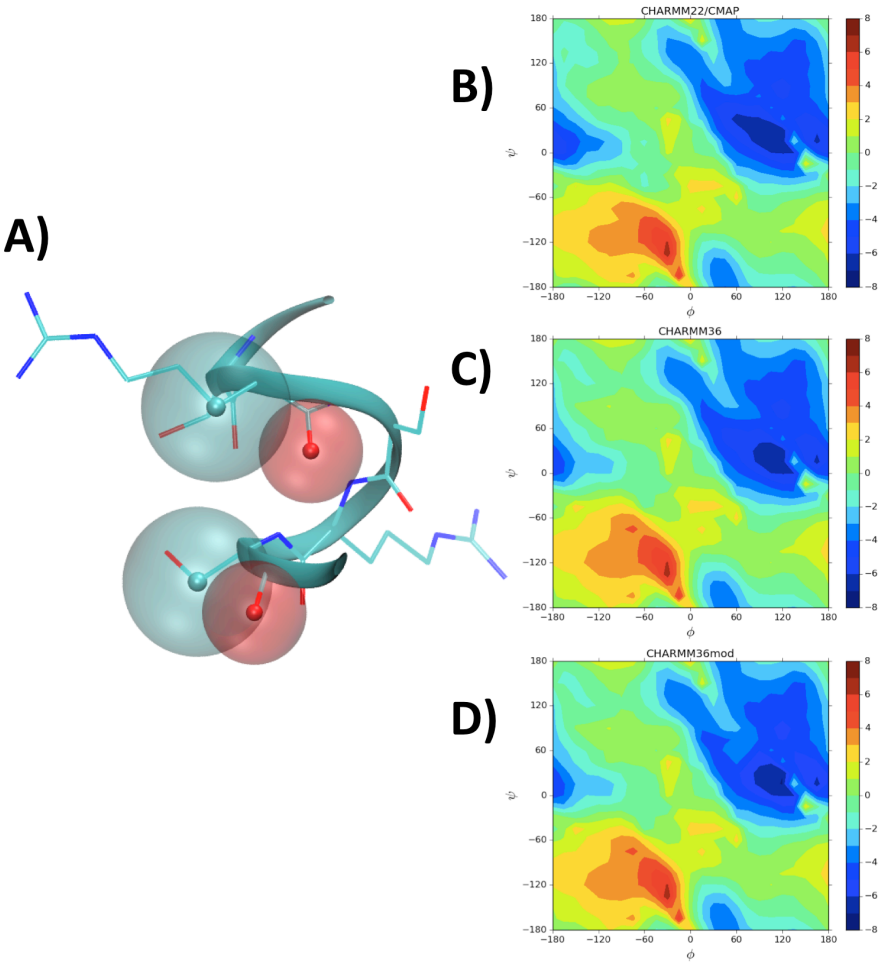
Supplementary Figure 15. RMS differences along the MD trajectories of designed proteins GA95 and GB95 with the C36m FF. GA95 and GB95 differ by only three residues in their sequence (L20, I30 and L45 for GA95, versus A20, F30 and Y45 for GB95) but have completely different folds (3-α for GA95, and 4β+α for GB95).[12] (A) MD simulations of GA95 and GB95 sequences starting from the 3-α initial structure (pdb id: 2kdl) show that GA95 is stable during a 2 μs simulation while the GB95 quickly unfolds within 100 ns. (B) Simulations of GA95 and GB95 starting from the 4β+α initial structure (pdb id: 2kdm). Both sequences are stable in the 2 μs MD timescale, while GB95 has lower RMSDs compared to GA95. In summary, the C36m FF is able to qualitatively differentiate this pair of proteins with 95% sequence identity but alternative folds.



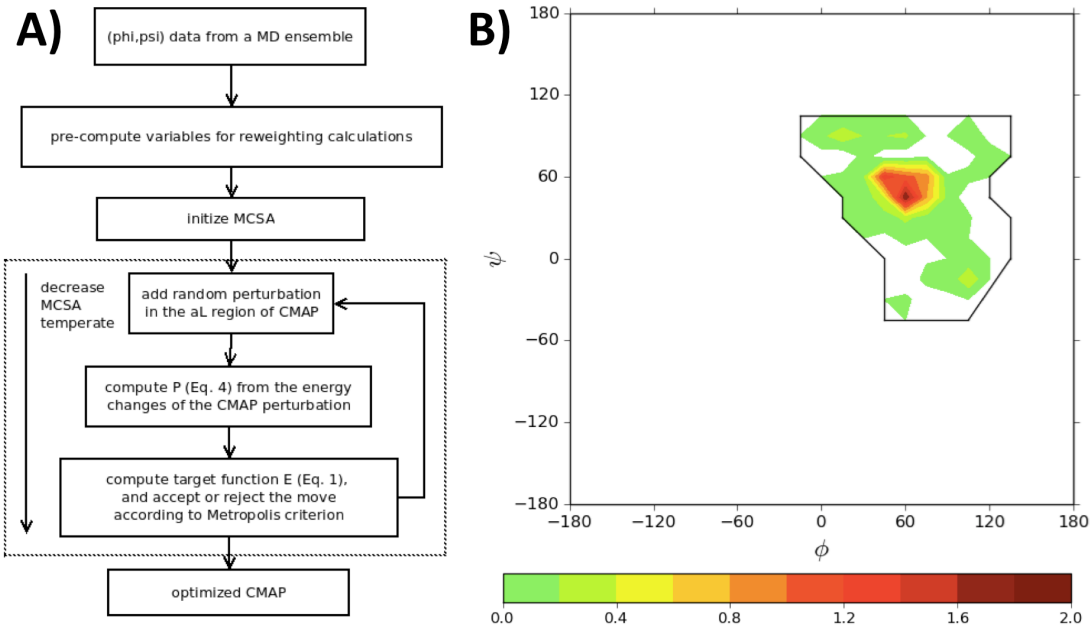
Supplementary Figure 16. Probability distribution of radius of gyration, end-to-end distance and peptide-peptide hydrogen bond number for the RS peptide (A, B and C), the IN protein (D, E and F) and the CspTm protein (G, H and I) from MD simulations using the C36m FF with different water models. Results obtained with the standard protein-water interactions based on the CHARMM TIP3P water model (red), with a general scaling of protein-water VdW interactions by 1.05 (green), and with an alternative water model with ε_H= -0.1 kcal/mol (blue), are plotted.



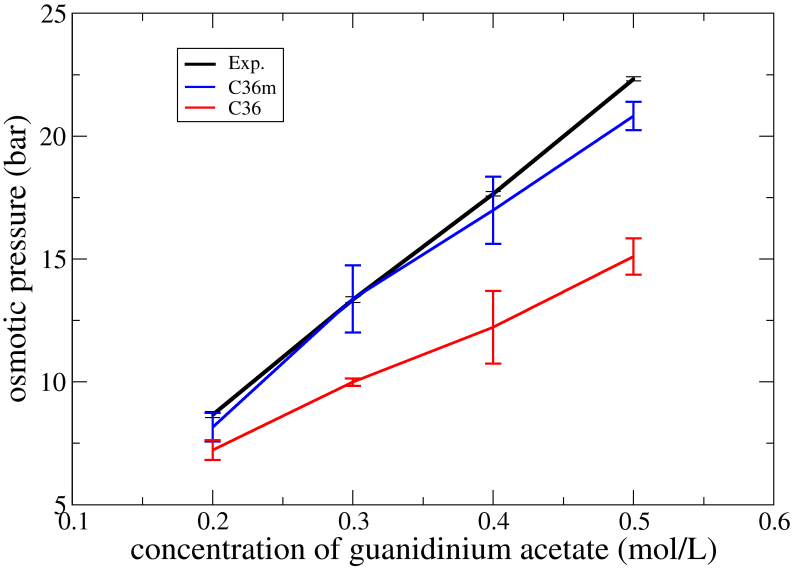
Supplementary Figure 17. A turn of left α-helix in the RS peptide is shown with spheres around the Cβ atom and O atom plotted using their VdW radius in the C36 FF (A). Also plotted are the 2D backbone CMAP potentials employed in the C22/CMAP (B), C36 (C) and C36m FFs. (D) The unit of the energy scale is kcal/mol.



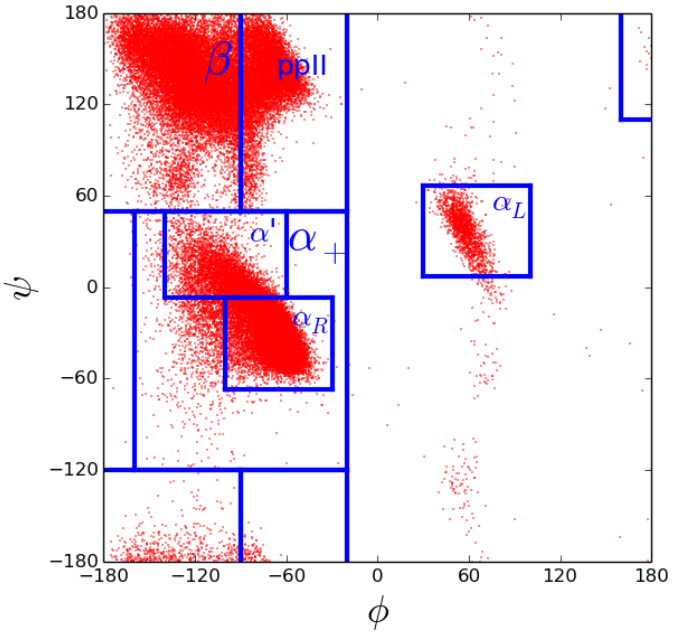
Supplementary Figure 18. A) Flow diagram of the CMAP potential optimization. Starting from the ϕ, ψ values, intermediate variables for reweighting were calculated including the observable h_i in Eq. 4 and CMAP lookup indexes and residuals. These pre-computed variables allow efficient evaluation of energies and target functions in the Monte Carlo simulated annealing. B) The difference of the optimized and C36 CMAP potentials. Units are kcal/mol and the region subjected to MC optimization is indicated by the black line.



Supplementary Figure 19. Osmotic pressure of guanidinium acetate solution with different concentrations obtained from MD simulations with the C36 (no NBFIX term) and the C36m (a NBFIX term between with R_{min}=3.637 Å) FF. The experimental data are from a personal communication with Shen and Roux.



Supplementary Figure 20. Definition of different regions in a Ramachandran plot. α_L is defined as 30°<ϕ<100° and 7°<ψ< 67°, α_R as -100°<ϕ<-30° and -67°<ψ< -7°, α' as -140°<ϕ<-60° and -7°<ψ< 50°. A broadly defined α^+ covers -160°<ϕ<-20° and -120°<ψ< 50°, while the ppII region covers -90°<ϕ<-20° and 50°<ψ< 180° or -180°<ψ<-120°, and the β region is defined as -180°<ϕ<-90° and 50°<ψ< 180° plus -180°<ϕ<-90° and -180°<ψ< -120°, or 160°<ϕ<180° and 110°<ψ< 180°. The ϕ, ψ data of the "top 500" pdb structures compiled by Lovell *et al* are plotted as scattering points.



Supplementary Table 1. Peptides and proteins used for optimization and validation of the C36m force field. Systems were solvated with 150mM NaCl, except for $(AAQAA)_3$ and ubiquitin where no ions were added, and for $(Ala)_5$ and HP36 one and two Cl^- ions were added, respectively, to neutralize the system. All simulations listed were carried out with the CHARMM TIP3P water model.

Protein/Peptide PDB ID/Sequence	Box size (Å)	ions	MD engine	Simulation details
1) FG peptide GAFGSNTNNTNNSNTS	59.5	NaCl	Gromacs	1 μs * 20 with C36m FF
2) RS peptide GAMGPSYG$(RS)_8$	84.1	NaCl	Gromacs	T-REX with C36m FF (0.63 μs * 34 replica)
3) HEWL19 KVFGM*QLAAAMKRHGLDN	62.3	NaCl	OpenMM	5.1 μs MD with C36 FF 5.1 μs MD with C36m FF
4) apo N-terminal zinc-binding domain of HIV-1 integrase (IN) 1wjb, 55 aa	70.0	NaCl	OpenMM	5.1 μs MD with C36 FF 5.1 μs MD with C36m FF
5) Ac-$(AAQAA)_3$-NH_2	38.9	--	OpenMM	16 μs MD with C36 FF 16 μs MD with C36m FF
6) GB1 hairpin GEWTYDDATKTFTVTE	42.1	NaCl	Gromacs	T-REX with C36 FF T-REX with C36m FF (0.8 μs * 32 replica)
7) Nrf2 hairpin AQLQLDEETGEFLPIQ	67.5	NaCl	Gromacs	T-REX with C36m FF (1 μs * 28 replicas)
8) Chignolin GYDPETGTWG	51.8	NaCl	Gromacs	T-REX with C36m FF (6 μs * 29 replicas)
9) CLN025 YYDPETGTWY	51.8	NaCl	Gromacs	T-REX with C36m FF (6 μs * 29 replicas)
10) $polyQ_{30}$ $(Q)_{30}$	82.7	--	CHARMM	H-REX with C36m FF (20 ns * 15 replicas)
11) Ac-$(AGQ)_n$-NH_2 peptides			OpenMM	
CAGQW	35.0	NaCl		10.1 μs MD with C36m FF
$C(AGQ)_2W$	37.0	NaCl		10.1 μs MD with C36m FF
$C(AGQ)_3W$	48.0	NaCl		10.1 μs MD with C36m FF
$C(AGQ)_4W$	56.0	NaCl		10.1 μs MD with C36m FF
12) 15 folded proteins[a]	see Supplementary Note		OpenMM	1 μs each with C36m FF
13) Ubiquitin 1ubq, 76 aa	58.4	--	NAMD	1.2 μs MD with C36m FF
14) villin headpiece subdomain HP36 1vii, 36 aa	60.0	--	CHARMM	H-REX with C36 FF H-REX with C36m FF (100 ns * 20 replicas)
15) HP36 N-terminal fragment HP21 MLSDEDFKAVFGMTRSAFANL	77.0	NaCl	Gromacs	2 μs * 30 with C36m FF
16) GA95 and GB95 2kdl and 2kdm, 56 aa	59.0	--	OpenMM	2 μs each with C36m FF
17) Cold-shock protein from Thermotoga maritima (CspTm) 1g6p, 66 aa	69.0	NaCl	OpenMM	5 μs MD with C36 FF 5 μs MD with C36m FF

a) folded proteins include crambin (pdb id:1ejg), protein G B3 domain (1p7e), bovine pancreatic trypsin inhibitor (5pti), erabutoxin B (3ebx), cold shock protein A (1mjc), C-terminal domain of human transcription factor IIF (1i27), ribosomal protein S6 (3zzp), DNA methyltransferase 1 associated protein 1 (4iej), bovine intestinal calcium-binding protein (3icb), ubiquitin (1ubq), PDZ domain (3vqf), lysozyme (135l), intestinal fatty acid-binding protein (1ifc), deoxy-myglobin (1bzp), and dethiobiotin synthetas (1byi).

Supplementary Table 2. J-coupling constants for the central alanine residues in the HEWL19 peptide from experiments and 5 μs MD simulations with the C36 and C36m FFs using the Karplus equation with parameters from Vögeli and Bax.[13] Uncertainties are given in brackets, and all the units are in Hz. Weighted χ^2 for each residue[14] is also listed.

residue	J-coupling type	C36	C36m	Exp.	C36 χ^2	C36m χ^2
A9	$^3J(H_N, H\alpha)$	5.87 (0.27)	5.19 (0.22)	5.18 (0.04)	0.42	0.30
	$^3J(H_N, C')$	1.47 (0.12)	1.38 (0.02)	1.39 (0.02)		
	$^3J(H\alpha, C')$	2.10 (0.11)	1.66 (0.09)	2.06 (0.19)		
	$^3J(H_N, C\beta)$	1.94 (0.27)	2.40 (0.11)	2.26 (0.04)		
	$^1J(N, C\alpha)$	10.94 (0.18)	10.78 (0.17)	10.54 (0.05)		
	$^2J(N, C\alpha)$	7.69 (0.19)	7.56 (0.20)	7.24 (0.06)		
A10	$^3J(H_N, H\alpha)$	5.92 (0.32)	5.11 (0.12)	5.10 (0.02)	0.59	0.13
	$^3J(H_N, C')$	1.16 (0.14)	1.43 (0.05)	1.33 (0.05)		
	$^3J(H\alpha, C')$	2.36 (0.10)	1.85 (0.16)	1.72 (0.04)		
	$^3J(H_N, C\beta)$	2.26 (0.05)	2.41 (0.06)	2.19 (0.08)		
	$^1J(N, C\alpha)$	10.55 (0.07)	10.63 (0.10)	10.58 (0.04)		
	$^2J(N, C\alpha)$	7.12 (0.18)	7.35 (0.13)	7.02 (0.06)		
	$^3J(N, C\alpha)$	0.51 (0.03)	0.46 (0.04)	0.46 (0.03)		
A11	$^3J(H_N, H\alpha)$	5.53 (0.16)	5.48(0.25)	5.67 (0.03)	0.17	0.09
	$^3J(H_N, C')$	1.21 (0.05)	1.17 (0.04)	1.09 (0.04)		
	$^3J(H\alpha, C')$	2.00 (0.15)	2.20 (0.32)	2.20 (0.03)		
	$^3J(H_N, C\beta)$	2.41 (0.06)	2.47 (0.10)	2.21 (0.06)		
	$^1J(N, C\alpha)$	10.76 (0.16)	10.45 (0.10)	10.57 (0.05)		
	$^2J(N, C\alpha)$	7.50 (0.25)	6.99 (0.14)	7.17 (0.03)		
	$^3J(N, C\alpha)$	0.44 (0.03)	0.43 (0.04)	0.43 (0.00)		
total					0.39	0.16

Supplementary Table 3. Chain dimensions of the RS peptide computed from the C36 and the C36m ensembles. The C36m FF leads to larger chain dimensions, consistent with experimental measurements from both SAXS and PFG-NMR.

	radius of gyration (Å)	hydrodynamic radius (Å)
C36	11.04 ± 0.11	11.05 ± 0.01
C36m	13.11 ± 0.08	12.01 ± 0.03
Experiment	12.62 ± 0.07	11.9 ± 0.1

Supplementary Table 4. The correlation between computed and experimental NMR scalar couplings and chemical shifts of the RS peptide from the C36 and the C36m ensembles. Average unsigned error (AUE) for each type of experimental measurements and total weighted error (χ^2) are listed. Chemical shifts were calculated using SHIFTX2.[11]

		C36	C36m
AUE for each type of scalar coupling (Hz)	$^3J(H_N, H\alpha)$	0.59	0.71
	$^1J(C\alpha, C\beta)$	1.04	1.18
	$^1J(C\alpha, H\alpha)$	2.10	1.35
	$^3J(N, C\gamma)$	0.28	0.25
	$^3J(C, C\gamma)$	0.76	0.59
χ^2 for all scalar couplings		2.00	1.58
AUE for each type of chemical shift (ppm)	$C\alpha$	0.39	0.28
	C’	0.54	0.52
χ^2 for all chemical shifts		1.46	1.18

Supplementary Table 5. Conformational sampling of left-handed and right-handed α helices of $(AAQAA)_3$ obtained from MD simulations using different FF parameters. The α_R fraction column is highlighted as it can be directly compared with experimental estimates of ~19% and ~22% at 300K.[15,16] Results of the C36 FF with original TIP3P water model were listed as this is used in previous studies.[14,17] MD simulations using the C36 FF with the original TIP3P water model[18] leads to significantly larger fraction of right-handed α helix, consistent with a recent replica exchange study of $(AAQAA)_3$ using both water models.[19]

protein FF	water model	α_L probability	α_L fraction	α_L propensity	α_R probability	α_R fraction	α_R propensity
C36m	CHARMM TIP3P	0.2% ± 0.1%	0.04% ± 0.01%	2.6% ± 0.1%	32.5% ± 3.0%	**17.0% ± 2.2%**	26.0% ± 1.9%
C36	CHARMM TIP3P	5.3% ± 1.5%	1.2% ± 0.4%	6.7% ± 0.5%	27.6% ± 5.1%	**12.7% ± 3.0%**	21.8% ± 2.7%
C36	original TIP3P	3.9% ± 0.5%	0.9% ± 0.1%	5.7% ± 0.4%	61.3% ± 5.4%	**37.3% ± 3.6%**	44.3% ± 3.3%

Supplementary Table 6. Population of folded states of β-hairpins at 300 K obtained with temperature replica exchange simulations using the C36m FF.

β hairpins	Simulation	Experiment
Gb1 hairpin	34% ± 2%	~30%[1]
Cln025	41% ± 2%	~90%[20]
Chigolin	2.6% ± 0.2%	~60%[21]
Nrf2 hairpin	3% ± 1%	no experimental data

Supplementary Table 7: Loop closure kinetics of the C(AGQ)$_n$W peptides. The contact formation rates, as well as their diffusion-limited parts k_{D+} and a reaction-limited parts k_R, are listed. Contact formation rates were calculated by integrating the triplet survival probability *S(t)* via $k^{-1} = \int_0^\infty S(t)\mathrm{d}t$. The diffusion-limited rate k_{D+} can be calculated using the survival probability $S(t) = \langle H(t_c(t_0) - t - t_0)\rangle_{t_0}$, where t_c is the first contact time after t_0 that r becomes less than r_c.[5] The reaction-limited rate k_R can either be deduced from $k_R^{-1} = k_{calc}^{-1} - k_{D+}^{-1}$, or computed using the distance probability *P(r)* via $k_R = q_c \int_0^{r_c} P(r)\mathrm{d}r$, and both approaches lead to similar results as listed. The uncertainties of computational rates are all less than 0.002 ns^{-1} based on block analysis.

	Simulation			Experiment		
peptide	k_{calc} (ns^{-1})	k_{D+} (ns^{-1})	k_R (ns^{-1})	k_{obs} (ns^{-1})	k_{D+} (ns^{-1})	k_R (ns^{-1})
C(AGQ)$_1$W	0.046	0.213	0.058 / 0.061	0.025	0.136	0.031
C(AGQ)$_2$W	0.015	0.077	0.018 / 0.020	0.014	0.074	0.017
C(AGQ)$_3$W	0.012	0.057	0.016 / 0.018	0.0030	0.0041	0.011
C(AGQ)$_4$W	0.010	0.047	0.013 / 0.015	0.0020	0.0027	0.008

Supplementary Table 8. Scalar couplings across backbone hydrogen bonds in ubiquitin from NMR experiments and 1.2 μs MD simulations with the C36 and C36m FFs. The RMS difference between computed and experimental scalar couplings across hydrogen bonds equals 0.10 Hz for both C36 and C36m simulations, and the correlation coefficient between computed couplings from these two models was 0.99. The residues for hydrogen bond donor and acceptor are listed, and all units are in Hz.

Residues	C36	C36m	Exp.
3-15	-0.496 ± 0.006	-0.514 ± 0.008	-0.445
4-65	-0.579 ± 0.031	-0.574 ± 0.035	-0.605
6-67	-0.748 ± 0.010	-0.758 ± 0.018	-0.594
7-11	-0.466 ± 0.027	-0.408 ± 0.036	-0.579
13-5	-0.738 ± 0.010	-0.736 ± 0.015	-0.725
15-3	-0.616 ± 0.022	-0.625 ± 0.022	-0.616
23-54	-0.509 ± 0.040	-0.576 ± 0.017	-0.553
26-22	-0.282 ± 0.005	-0.297 ± 0.005	-0.289
27-23	-0.465 ± 0.017	-0.487 ± 0.009	-0.515
28-24	-0.331 ± 0.011	-0.345 ± 0.006	-0.263
29-25	-0.348 ± 0.010	-0.372 ± 0.021	-0.268
30-26	-0.336 ± 0.011	-0.332 ± 0.011	-0.376
31-27	-0.411 ± 0.004	-0.406 ± 0.005	-0.405
32-28	-0.321 ± 0.008	-0.340 ± 0.023	-0.309
33-29	-0.392 ± 0.016	-0.431 ± 0.019	-0.211
34-30	-0.751 ± 0.015	-0.726 ± 0.016	-0.681
35-31	-0.117 ± 0.004	-0.118 ± 0.002	-0.208
42-70	-0.617 ± 0.017	-0.571 ± 0.012	-0.513
44-68	-0.714 ± 0.010	-0.711 ± 0.009	-0.605
45-48	-0.520 ± 0.018	-0.493 ± 0.013	-0.515
50-43	-0.636 ± 0.012	-0.612 ± 0.015	-0.626
56-21	-0.354 ± 0.019	-0.326 ± 0.014	-0.536
57-19	-0.448 ± 0.008	-0.435 ± 0.006	-0.412
61-56	-0.479 ± 0.014	-0.434 ± 0.004	-0.203
64-2	-0.767 ± 0.019	-0.794 ± 0.016	-0.835
67-4	-0.844 ± 0.006	-0.838 ± 0.004	-0.687
68-44	-0.583 ± 0.005	-0.591 ± 0.011	-0.645
69-6	-0.607 ± 0.006	-0.633 ± 0.012	-0.534
70-42	-0.657 ± 0.008	-0.641 ± 0.021	-0.603

Supplementary Table 9. Q factors for backbone N-H RDCs in ubiquitin with different alignment media from 1.2 μs simulations with the C36 and the C36m FF. The medium number is the same as in Ref. [22] and [23], a denotes those involved in D23M, b denotes those included in D10, c denotes the media used in the side-chain methyl RDC calculation, and d denotes those included in D36M. The average Q factors increase from 0.24 ± 0.01 with the C36 FF to 0.26 ± 0.01 with C36m, which is mainly due to the deviation of RDCs in residues Lys11 and Ile61. Excluding these two residues from the RDC calculations results in an average Q factor of 0.24 for both MD ensembles generated from the C36 and C36m FF.

medium	Q factor		medium	Q factor	
	C36	C36m		C36	C36m
1[cd]	0.221	0.243	23[ad]	0.236	0.244
2[acd]	0.181	0.209	24[ad]	0.240	0.246
3[cd]	0.263	0.276	25[ad]	0.259	0.265
4[acd]	0.187	0.224	26[ad]	0.232	0.241
5[d]	0.214	0.229	27[ad]	0.182	0.223
6[d]	0.225	0.239	28[ad]	0.218	0.239
7[cd]	0.173	0.200	29[ad]	0.167	0.211
8[cd]	0.259	0.267	30[d]	0.256	0.294
9[acd]	0.666	0.662	31[d]	0.308	0.342
10[acd]	0.292	0.300	32[ad]	0.265	0.298
11[d]	0.225	0.241	33[ad]	0.255	0.284
12[cd]	0.205	0.223	34[ad]	0.241	0.270
13[ad]	0.182	0.209	35[ad]	0.201	0.228
14[ad]	0.212	0.232	36[ad]	0.203	0.220
15[ad]	0.404	0.414	37[b]	0.289	0.299
16[ad]	0.199	0.232	38[b]	0.432	0.451
17[d]	0.216	0.221	39[b]	0.303	0.315
18[d]	0.209	0.224	40[b]	0.237	0.275
19[abd]	0.166	0.176	41[b]	0.239	0.295
20[abd]	0.175	0.206	42[b]	0.217	0.266
21[d]	0.223	0.243	43[b]	0.193	0.246
22[ad]	0.227	0.238	44[b]	0.200	0.231

Supplementary Table 10. Through-bond and through-space J couplings for lysine side chains in ubiquitin. Experimental values were taken from Ref. [24] and calculated results were from 1.2 μs simulations with the C36 and the C36m FFs, respectively. For hydrogen bond scalar couplings, the residues containing the acceptor C=O group are given in parentheses.

		J_{exp} (Hz)	J_{calc} (Hz)	
			C36m	C36
$^{3}J_{C\gamma N\zeta}$	Lys6	1.78 ± 0.25	1.65 ± 0.08	1.63 ± 0.07
	Lys11	1.89 ± 0.03	1.95 ± 0.08	1.91 ± 0.09
	Lys27	2.45 ± 0.03	2.34 ± 0.01	2.35 ± 0.00
	Lys29	1.26 ± 0.03	1.42 ± 0.04	1.36 ± 0.07
	Lys33	1.60 ± 0.01	1.73 ± 0.10	1.74 ± 0.06
	Lys48	1.49 ± 0.01	1.52 ± 0.02	1.49 ± 0.08
	Lys63	1.71 ± 0.01	1.60 ± 0.03	1.73 ± 0.06
$^{h3}J_{N\zeta C'}$	Lys29 (Glu16)	-0.23 ± 0.03	-0.27 ± 0.01	-0.27 ± 0.04
	Lys33 (Thr14)	-0.17 ± 0.02	-0.06 ± 0.01	-0.06 ± 0.02

Supplementary Table 11. Relaxation order parameter S^2 for the lysine NH_3^+ groups in ubiquitin. The experimental values were taken from Ref. [25] and the calculated results were obtained from 1.2 μs simulations with the C36 and the C36m FFs.

S^2	exp	C36m	C36
Lys6	-	0.266 ± 0.038	0.222 ± 0.016
Lys11	0.415 ± 0.039	0.417 ± 0.034	0.444 ± 0.022
Lys27	0.709 ± 0.021	0.794 ± 0.011	0.816 ± 0.017
Lys29	0.378 ± 0.017	0.193 ± 0.009	0.192 ± 0.013
Lys33	0.248 ± 0.005	0.301 ± 0.009	0.301 ± 0.006
Lys48	0.192 ± 0.005	0.123 ± 0.006	0.111 ± 0.009
Lys63	0.267 ± 0.006	0.186 ± 0.005	0.186 ± 0.009

Supplementary Table 12. Side-chain methyl order parameter S^2 in ubiquitin from experiments and 1.2 μs simulations with the C36 and the C36m FFs.

Residues	C36	C36m	Exp.	Residues	C36	C36m	Exp.
I3γ	0.25 ± 0.13	0.27 ± 0.14	0.98	I30δ	0.41 ± 0.03	0.32 ± 0.01	0.77
I3δ	0.15 ± 0.10	0.16 ± 0.10	0.75	I36γ	0.86 ± 0.01	0.84 ± 0.02	0.83
V5γ1	0.83 ± 0.01	0.85 ± 0.01	0.91	I36δ	0.36 ± 0.02	0.37 ± 0.03	0.58
V5γ2	0.84 ± 0.01	0.86 ± 0.01	0.88	L43δ1	0.40 ± 0.03	0.37 ± 0.03	0.55
T7γ	0.26 ± 0.06	0.06 ± 0.09	0.75	L43δ2	0.36 ± 0.03	0.34 ± 0.03	0.61
L8δ1	0.16 ± 0.00	0.14 ± 0.01	0.27	I44γ	0.79 ± 0.03	0.72 ± 0.03	0.71
L8δ2	0.11 ± 0.01	0.16 ± 0.04	0.21	I44δ	0.11 ± 0.01	0.09 ± 0.01	0.31
T9γ	0.31 ± 0.03	0.08 ± 0.07	0.64	A46β	0.75 ± 0.01	0.73 ± 0.00	0.95
T12γ	0.77 ± 0.02	0.73 ± 0.03	0.93	L50δ1	0.84 ± 0.00	0.84 ± 0.00	0.89
I13γ	0.64 ± 0.04	0.67 ± 0.03	0.56	L50δ2	0.81 ± 0.00	0.81 ± 0.01	0.86
I13δ	0.21 ± 0.03	0.19 ± 0.03	0.55	T55γ	0.91 ± 0.00	0.93 ± 0.00	0.93
T14γ	0.65 ± 0.02	0.70 ± 0.06	0.78	L56δ1	0.56 ± 0.02	0.54 ± 0.02	0.60
L15δ1	0.55 ± 0.05	0.56 ± 0.05	0.58	L56δ2	0.55 ± 0.02	0.52 ± 0.02	0.62
L15δ2	0.52 ± 0.05	0.52 ± 0.04	0.62	I61γ	0.88 ± 0.00	0.87 ± 0.00	0.95
V17γ1	0.72 ± 0.06	0.68 ± 0.06	0.89	I61δ	0.27 ± 0.01	0.32 ± 0.02	0.56
V17γ2	0.73 ± 0.06	0.69 ± 0.05	0.89	L67δ1	0.26 ± 0.07	0.29 ± 0.07	0.30
T22γ	0.89 ± 0.00	0.88 ± 0.00	0.95	L67δ2	0.26 ± 0.06	0.29 ± 0.05	0.29
I23γ	0.87 ± 0.00	0.88 ± 0.00	0.95	L69δ2	0.69 ± 0.02	0.69 ± 0.04	0.55
I23δ	0.75 ± 0.00	0.75 ± 0.01	0.51	V70γ2	0.44 ± 0.06	0.32 ± 0.06	0.35
V26γ1	0.89 ± 0.00	0.88 ± 0.00	0.86	L71δ1	0.41 ± 0.04	0.41 ± 0.06	0.29
V26γ2	0.88 ± 0.00	0.87 ± 0.00	0.99	L73δ1	0.05 ± 0.01	0.07 ± 0.01	0.19
I30γ	0.93 ± 0.00	0.92 ± 0.00	0.93	L73δ2	0.06 ± 0.00	0.06 ± 0.02	0.17

Supplementary Table 13. The folding free energies, ΔG_f, of villin headpiece subdomain HP36 derived from the C36 and the C36m simulations and from experiments. The second experimental value refers to HP35, i.e. villin headpiece without the N-terminal Met residue. The C36 FF leads to a less favorable unfolded state as compared to the experimentally derived ΔG_f, and the C36m FF decreases ΔG_f leading to better agreement with the experimental data.

	ΔG_f [kcal/mol]
C36 simulation	-4.940
C36m simulation	-4.161
Experiment	-2.4[26], -3.1[27]

Supplementary Table 14. $^{3}J_{HNH\alpha}$ scalar couplings for the HP21 peptide. Experimental values were taken from Ref. [6] and calculated results were from 30 x 2 μs simulations with the C36m FFs. The Vögeli Karplus parameters were used.[13] Errors provided are standard error of the mean.

	Residue	J_{exp} (Hz)	J_{calc} (Hz)
			C36m
$^{3}J_{HNH\alpha}$	2	6.6	6.29 ± 0.02
	5	5.4	5.35 ± 0.09
	6	6.7	5.05 ± 0.10
	8	5.0	6.65 ± 0.20
	9	4.9	5.56 ± 0.08
	10	7.3	6.00 ± 0.08
	13	4.5	5.88 ± 0.06
	14	6.3	5.94 ± 0.06
	15	4.4	7.08 ± 0.07
	16	5.5	5.77 ± 0.04
	17	5.7	6.17 ± 0.03
	19	4.9	6.08 ± 0.02
	20	8.7	6.70 ± 0.01
	21	7.0	8.12 ± 0.01
Average unsigned error			1.08 Hz

Supplementary Table 15. Comparison to chemical shifts for the HP21 peptide. Experimental values were taken from Ref. [7] and calculated results were from 30 x 2 μs simulations with the C36m FFs using SHIFTX2.[11] Errors provided are standard error of the mean.

Chemical Shift	RMS error (ppm)
Cα	1.13 ± 0.04
Cβ	0.69 ± 0.01
C	0.94 ± 0.01
Hα	0.10 ± 0.002
HN	0.26 ± 0.004
N	1.92 ± 0.04

Supplementary Table 16. Radius of gyration of RS, IN and CspTM from MD simulations using the C36m FF with different water models. Results with the standard CHARMM TIP3P water model, as well as a general scaling of protein-water VdW interaction by 1.05, and an alternate water model with ε_H=-0.1 kcal/mol, are listed.

$\langle R_g \rangle$ (Å)	RS	IN	CspTm
Standard protein-water interaction	13.11 ± 0.08	13.8 ± 0.2	12.8 ± 0.2
General scaling up by 1.05	13.66 ± 0.09	14.3 ± 0.4	17.8 ± 0.4
Alternate water model with ε_H=-0.1 kcal/mol	14.67 ± 0.06	20.2 ± 0.3	16.4 ± 0.5
Exp.	12.62 ± 0.07[28]	24[29]	15[29], 16[30]

Supplementary Table 17. Mean FRET efficiency <E> of IN and CspTM from MD simulations using the C36m FF with different water models. Results with the standard CHARMM TIP3P water model, as well as a general scaling of protein-water VdW interaction by 1.05, and an alternate water model with ε_H=-0.1 kcal/mol, are listed.

<E>	IN	CspTm
Standard protein-water interaction	0.95 ± 0.02	0.99 ± 0.06
General scaling up by 1.05	0.95 ± 0.05	0.88 ± 0.03
Alternate water model with ε_H=-0.1 kcal/mol	0.74 ± 0.01	0.85 ± 0.05
Exp.[29]	0.53	0.85

Supplementary Table 18. MCSA optimization results with different weighting factors *w*. The predicted α_L probability (*P*) and RMS difference with the original C36 CMAP (RMS_{CMAP}) are listed. All runs are carried out using an initial MCSA temperature of 10 K with the temperature decreasing exponentially. Other initial temperatures were also tested and the results are not sensitive to them.

w (in *kT*)	*P*	RMS_{CMAP} (kcal/mol)
100	0.3880	0.0019
50	0.1525	0.0187
20	0.0148	0.0815
10	0.0123	0.0950
5	0.0115	0.1044
2 (C36m)	0.0113	0.1196
1	0.0114	0.1380
0.1	0.0114	0.1599
0.01	0.0115	0.1676

Supplementary Note. Detailed information for each simulation system.

1. Temperature replica exchange simulations of RS peptide.

The simulation system consisted of the RS peptide in a rhombic dodecahedral box with 14240 water molecules, 39 Na^+ and 47 Cl^- ions. GROMACS version 4.6.31 was used.[31,32] The RS peptide was built in a fully extended conformation with protonation states to match experimental conditions (arginine residues as well as the N- and C-termini were simulated in their charged states). Prior to the production runs, energy minimization with the steepest descent algorithm was performed. The lengths of bonds with hydrogen atoms were constrained using the LINCS algorithm.[33] An integration time step of 2 fs was used. A cutoff of 9.5 Å was used for the Lennard-Jones interactions and short-range electrostatic interactions. Long-range electrostatic interactions were calculated by particle-mesh Ewald summation with a grid spacing of 1.2 Å and a fourth order interpolation.[34] The velocity rescaling thermostat was used.[35] Equilibration was performed at 298 K for 1 ns using Berendsen pressure coupling[36] followed by 5 ns of simulation in the NPT ensemble using the Parrinello−Rahman algorithm.[37] Temperature replica exchange (T-REX)[38] simulations were performed in the canonical ensemble with a total of 34 temperatures between 298 and 350 K. Temperature exchanges were attempted every 2 ps for a total of 21.42 μs (0.63 μs of simulation per temperature). The first 30 ns of simulation were discarded as equilibration (based on radius of gyration and hydrogen bonding). Coordinates were stored before each temperature exchange; a total of 300000 conformations were collected per temperature.

2. Simulations of the FG peptide.

The simulation system consisted of the FG peptide in a rhombic dodecahedral box with 4897 water molecules, 14 Na^+ and 14 Cl^- ions. The simulations were performed with simulation parameters the same as described above for the RS peptide. Twenty independent MD simulations of 1 μs each in length were performed (2000000 conformations were obtained).

3. MD simulations of HEWL19.

The HEWL19 peptide was solvated in a cubic water box of ~ 62.3 Å per side containing 5677 water molecules, 16 Na^+ and 21 Cl^- ions. The starting structure of HEWL19 was taken from the previous run with the C36 FF.[28] MD simulations were carried out using OpenMM[39] in the NPT ensemble at 300 K and 1 atm. Temperature control is performed using the Andersen thermostat with a collision frequency of 1 ps^{-1}, and pressure control performed based on a Monte Carlo barostat. Periodic boundary conditions were applied and Lennard-Jones interactions were truncated at 12 Å with a potential switching function from 10 to 12 Å. Electrostatic interactions were calculated using the particle mesh Ewald[40] method with a real space cutoff of 12 A on an approximately 1 Å grid with a 4th-order spline. Covalent bonds to hydrogen atoms were constrained using the constant constraint matrix approximation.[41] The integration time step equals 2 fs and coordinates were saved every 100 ps.

4. MD simulations of IN.

Apo N-terminal zinc-binding domain of HIV-1 integrase (IN) was solvated in a cubic water box of ~70.0 Å per side containing 10303 water molecules, 32 Na^+ and 28 Cl^- ions. The starting protein structure taken from protein data bank[42] (pdb id: 1wjb) was unfolded with a 1 ns gas phase Langevin dynamics simulation at 800K using CHARMM,[43] and then the simulation system was built with the CHARMM-GUI.[44] MD simulations were carried out using OpenMM in the NPT ensemble at 300 K and 1 atm, with the same setup as HEWL19. Simulations with standard CHARMM modified TIP3P water model as well as alternative water models were each carried out for 5.1 μs with the first 0.1 μs discarded as the equilibrium phase. MD trajectories with the length of several μs were shown to lead to converged $\langle R_g \rangle$ for IDPs.[45] Although a relatively large water box was used, there exists the possibility that when the protein is highly extended it forms VdW contacts with its periodic image. A larger solvation box could reduce such a possibility but will significantly increase the already large

computational cost, so we simply discard from the analysis all conformations in which any atom of the protein is within 8 Å of any atom in any of the protein's periodic images.

5. MD simulations of $(AAQAA)_3$

Ac-$(AAQAA)_3$-NH_2 was solvated in a cubic water box of ~38.9 Å per side containing 1943 water molecules, and the starting structure was taken from previous simulations.[46] MD simulations were carried out using OpenMM in the NPT ensemble at 300 K and 1 atm, with the same setup as HEWL19.

6. Temperature replica exchange simulations of GB1 hairpin.

The simulation system consisted of the GB1 hairpin in a rhombic dodecahedral box with 1816 water molecules, 6 Na^+ and 3 Cl^- ions. The simulations were performed with simulation parameters the same as described above for the RS peptide. Temperature replica exchange simulations were performed with 32 temperatures between 278 K and 419 K for a total of 25.6 μs (0.8 μs of simulation per temperature).

7. Temperature replica exchange simulations of Nrf2 hairpin.

The simulation system consisted of the Nrf2 peptide in a rhombic dodecahedral box with 7203 water molecules, 24 Na^+ and 20 Cl^- ions. The simulations were performed with simulation parameters the same as described above for the RS peptide. Temperature replica exchange simulations were performed with 28 temperatures between 300 K and 367 K for a total of 28 μs (1 μs of simulation per temperature).

8. Temperature replica exchange simulations of the chignolin and CLN025 peptides.

The simulation system consisted of the CLN025 peptide in a rhombic dodecahedral box with 3240 water molecules and 2 Na^+ ions. The simulations were performed with simulation parameters the same as described above for the RS peptide. Temperature replica exchange simulations were performed with

29 temperatures between 278 K and 368 K. The T-REX simulation was run for a total of 174 μs (6 μs of simulation per temperature). The population of the folded hairpin did not converge until 4 μs of simulation per temperature, and therefore only the final 2 μs of simulation were used in the analysis. For chignolin, the simulation system consisted of the chignolin peptide in a rhombic dodecahedral box with 3255 water molecules and 2 Na^+ ions. The same simulation protocol used for the CLN025 peptide was followed. The population of the folded hairpin remained low, and the entire 6 μs of simulation were used in the analysis.

9. Hamiltonian replica exchange of polyQ

Partially extended Q_{30}, terminated with zwitterionic termini, was placed in a cubic box with a length of 82.74 Å in each dimension and solvated with 18,967 TIP3P water molecules. Periodic boundaries were applied and particle-mesh Ewald summation[40] was used to calculate electrostatic interactions. The direct sum and Lennard-Jones interactions were truncated a 9 Å cutoff (with a switching function effective between 8 and 9 Å). The peptide was sampled via Hamiltonian replica exchange molecular dynamics simulations. The end-to-end distance between Cα atoms of residues 1 and 30 was used as the biasing reaction coordinate. Harmonic umbrella potentials with a force constant of 0.2 kcal/mol/$Å^2$ were applied to target end-to-end distance ranging from 5 to 75 Å spaced at 5 Å intervals to require 15 replicas. Each replica was simulated at 300K in the NVT ensemble using a Langevin thermostat with a friction constant of 0.5 ps^{-1}. SHAKE[47] was applied and an integration time step of 2 fs was used. Replica exchanges were attempted every 2 ps and a total of 10,000 cycles (20 ns) were completed for each replica. The simulations were carried out using CHARMM[43], version c41a1, in combination with OpenMM[39] to take advantage of GPU acceleration. The replica exchange functionality was provided by the MMTSB tool set.[48] The overall conformational sampling was characterized by the potential of mean force as a function of radius of gyration and the end-to-end distance obtained via two-dimensional WHAM analysis using the wham-2d program.[49]

10. MD simulations of Ac-$(AGQ)_n$-NH_2 peptides.

The Ac-$(AGQ)_n$-NH_2 peptides with n=1-4 were simulated using OpenMM in the NPT ensemble at 293 K and 1 atm. CAGQW was solvated in a cubic water box of ~35 Å per side containing 1285 water molecules, 2 Na^+ and 3 Cl^- ions. $C(AGQ)_2W$ was solvated in a cubic water box of ~37 Å per side containing 1518 water molecules, 3 Na^+ and 4 Cl^- ions. $C(AGQ)_3W$ was solvated in a cubic water box of ~48 Å per side containing 3376 water molecules, 6 Na^+ and 7 Cl^- ions. $C(AGQ)_4W$ was solvated in a cubic water box of ~56 Å per side containing 5413 water molecules, 10 Na^+ and 11 Cl^- ions. The simulation setup is the same as HEWL19.

11. MD simulations of 15 folded proteins.

1000 ns MD simulations were carried out using OpenMM for 15 folded proteins, including crambin (pdb id:1ejg), protein G B3 domain (1p7e), bovine pancreatic trypsin inhibitor (5pti), erabutoxin B (3ebx), cold shock protein A (1mjc), C-terminal domain of human transcription factor IIF (1i27), ribosomal protein S6 (3zzp), DNA methyltransferase 1 associated protein 1 (4iej), bovine intestinal calcium-binding protein (3icb), ubiquitin (1ubq), PDZ domain (3vqf), lysozyme (135l), intestinal fatty acid-binding protein (1ifc), deoxy-myglobin (1bzp), and dethiobiotin synthetas (1byi). Protein structures taken from pdb files[42] were solvated in pre-equilibrated cubic TIP3P water boxes with at least 10 Å larger at each direction in the x, y, z dimensions. Counter-ions of Na^+ or Cl^- were added to keep systems neutral. Simulations were carried out in the NPT ensemble at 300 K and 1 atm, with the same simulation setup as HEWL19, except coordinates were saved every 1 ns.

12. MD simulations of ubiquitin.

The 1.2 μs MD simulation of ubiquitin with the C36m FF was carried out using NAMD[50] with the same setup as the C36 trajectory reported in Ref. [23] Briefly, ubiquitin was solvated in a cubic water

box of ~58.4 Å per side containing 6353 water molecules and no ions. Periodic boundary conditions were applied and Lennard-Jones (LJ) interactions were truncated at 12 Å with a force switch smoothing function from 10 Å to 12 Å. The non-bonded interaction lists were generated with a distance cutoff of 16 Å and updated heuristically. Electrostatic interactions were calculated using the particle mesh Ewald method[40] with a real space cutoff of 12 Å on an approximately 1 Å grid with 6th order spline. Covalent bonds to hydrogen atoms were constrained by SHAKE.[47] Langevin thermostat with a damping factor of 5 ps^{-1} and the Nosé-Hoover Langevin piston method with a barostat oscillation time scale of 200 fs were used for the NPT simulation at 300 K and 1 atm. The time step equals 2 fs and coordinates were stored every 100 ps.

13. Hamiltonian replica exchange simulations of HP36

All-atom Hamiltonian Replica Exchange Molecular Dynamic (H-REMD) simulations of chicken villin headpiece subdomain (HP36, pdb id: 1vii) were performed to determine the folding free energy ΔG_f with the C36 and the C36m FFs. The protein was placed in a cubic box with a length of 60 Å in each dimension and solvated with 6685 TIP3P water molecules and 2 Cl^- ions. First, the energy was minimized using steepest descent algorithm. Second, an NPT equilibration was performed with increasing temperature and constrained positions of the protein's heavy atoms. Third, extended protein conformations were generated by gradually increasing the distance between C_α of residue 15 (the second α-helix) and C_α of residue 29 (the third α-helix). 20 replicas, each with a different distance restraint, were selected to start Hamiltonian replica exchange simulations using this distance as the biasing reaction coordinate. The smallest distance was equal to 10.5 Å, which is slightly less than the native 10.86 Å. The subsequent distances increased every 0.6 Å up to 21.9 Å. The distances were restrained with a harmonic potential with a force constant k=4.5 kcal/mol/Å. Each replica was simulated at 298K in the NVT ensemble using a Langevin thermostat with a friction constant of 0.01 ps^{-1}. Periodic boundaries were applied and particle-mesh Ewald summation[40] was used to calculate

electrostatic interactions. The direct sum and Lennard-Jones interactions were truncated a 9 Å cutoff (with a switching function effective between 8 and 9 Å). SHAKE[47] was applied and an integration time step of 2 fs was used. Replica exchanges were attempted every 10 ps and a total of 100 ns were completed for each replica with the first 10 ns omitted from the analysis as the equilibration phase. The simulations were carried out using CHARMM[43], version c41a1, in combination with OpenMM[39] to take advantage of GPU acceleration. The replica exchange functionality was provided by the MMTSB tool set.[48]

2D WHAM analysis[49] was used for potential of mean force (PMF) calculations, as a function of both the biased distance (*d*) and root-mean-square deviation (RMSD). *d* was than integrated to obtain the unbiased 1D PMF on RMSD. Errors were calculated based on 10 ns non-overlapping intervals. The folding free energy ΔG_f was calculated based on the probabilities of finding the protein in the folded (*f*) and unfolded (*uf*) state:

$$\Delta G_f = -k_B \mathrm{T} \ln \frac{\sum_{i,f} p(RMSD_i)}{\sum_{i,uf} p(RMSD_i)}$$

The folded state was defined as RMSD being less than 4 Å and the unfolded as larger than 6.0 Å and 6.25 Å for C36 and C36m, respectively. PMFs were also calculated as a function of RMSD and R_g. Multistate Bennett Acceptance Ratio (MBAR) method[51] was used to unbias all the three coordinates (*d*, RMSD, and R_g), and then the 2D PMF was obtained by integrating *d*.

14. MD Simulations of HP21.

The simulations of HP21 were performed with simulation parameters the same as described above for the RS peptide. Thirty independent MD simulations of 2 μs each in length were performed with GROMACS.

15. MD Simulations of GA95 and GB95.

The designed GA95 and GB95 proteins were solvated in cubic water boxes of ~59.0 Å per side containing 6138 water molecules for both systems. The starting structure were taken from the NMR structures for either the 3α fold (pdb id: 2kdl) and the 4β+α fold (pdb id:2kdm). The systems were built using CHARMM, and subjected to 1 ns NVT simulations with CHARMM as equilibrium before the 2 μs production NPT simulations with OpenMM. Simulation parameters are the same as described above for the HEWL19 peptide. RMSDs were calculated by aligning protein structures with the starting NMR structure using the Cα atom positions of residues 9 to 52.

16. MD simulations of CspTm.

Cold-shock protein from Thermotoga maritima (CspTm) was solvated in a cubic water box of ~ 69.0 Å per side containing 9743 water molecules, 28 Na^+ and 28 Cl^- ions. The starting protein structure taken from protein data bank (pdb id: 1g6p) was unfolded with 1 ns gas phase Langevin dynamics simulation at 800K using CHARMM,[43] and then the simulation system was built with CHARMM-GUI.[44] MD simulations were carried out using OpenMM in the NPT ensemble at 300 K and 1 atm, with the same setup as HEWL19. Simulations with standard CHARMM modified TIP3P water model as well as alternative water models were each carried out for 5.1 μs with the first 0.1 μs discarded as the equilibrium phase. All conformations in which the protein forms VdW contact with its periodic images based on an 8 Å distance between non-hydrogen atoms were discarded from the analysis.

Reference

1. R. M. Fesinmeyer, F. M. Hudson, N. H. Andersen, *Journal of the American Chemical Society* 2004, 126, 7238-7243.
2. E. A. Cino, W.-Y. Choy, M. Karttunen, *Journal of Chemical Theory and Computation* 2012, 8, 2725-2740.
3. R. H. Walters, R. M. Murphy, *Journal of Molecular Biology* 2009, 393, 978-992.
4. Aaron M. Fluitt, Juan J. de Pablo, *Biophysical Journal* 2015, 109, 1009-1018.
5. I.-C. Yeh, G. Hummer, *Journal of the American Chemical Society* 2002, 124, 6563-6568.
6. Y. Tang, M. J. Goger, D. P. Raleigh, *Biochemistry* 2006, 45, 6940-6946.
7. W. Meng, B. Shan, Y. Tang, D. P. Raleigh, *Protein Science* 2009, 18, 1692-1701.
8. A. S. Baltzis, N. M. Glykos, *Protein Science* 2016, 25, 587-596.
9. W. Kabsch, C. Sander, *Biopolymers* 1983, 22, 2577-2637.
10. J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, *Protein Science* 2006, 15, 2795-2804.
11. B. Han, Y. Liu, S. W. Ginzinger, D. S. Wishart, *Journal of Biomolecular NMR* 2011, 50, 43-57.
12. P. A. Alexander, Y. He, Y. Chen, J. Orban, P. N. Bryan, *Proceedings of the National Academy of Sciences* 2009, 106, 21149-21154.
13. B. Vögeli, J. Ying, A. Grishaev, A. Bax, *Journal of the American Chemical Society* 2007, 129, 9377-9385.
14. R. B. Best, X. Zhu, J. Shim, P. Lopes, J. Mittal, M. Feig, A. D. M. Jr., *J. Chem. Theory Comput.* 2012.
15. W. Shalongo, L. Dugad, E. Stellwagen, *Journal of the American Chemical Society* 1994, 116, 8288-8293.
16. C. A. Rohl, R. L. Baldwin, *Biochemistry* 1997, 36, 8435-8442.
17. R. B. Best, J. Mittal, M. Feig, A. D. MacKerell, *Biophysical journal* 2012, 103, 1045-1051.
18. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* 1983, 79, 926-926.
19. S. Boonstra, P. R. Onck, E. van der Giessen, *The Journal of Physical Chemistry B* 2016.
20. S. Honda, T. Akiba, Y. S. Kato, Y. Sawada, M. Sekijima, M. Ishimura, A. Ooishi, H. Watanabe, T. Odahara, K. Harata, *Journal of the American Chemical Society* 2008, 130, 15327-15331.
21. S. Honda, K. Yamasaki, Y. Sawada, H. Morii, *Structure* 2004, 12, 1507-1518.
22. N. A. Lakomek, K. F. Walter, C. Fares, O. F. Lange, B. L. de Groot, H. Grubmuller, R. Bruschweiler, A. Munk, S. Becker, J. Meiler, C. Griesinger, *Journal of biomolecular NMR* 2008, 41, 139-155.
23. J. Huang, A. D. MacKerell, *Journal of Computational Chemistry* 2013, 34, 2135-2145.
24. L. Zandarashvili, D. W. Li, T. Wang, R. Bruschweiler, J. Iwahara, *Journal of the American Chemical Society* 2011, 133, 9192-9195.
25. A. Esadze, D. W. Li, T. Wang, R. Bruschweiler, J. Iwahara, *Journal of the American Chemical Society* 2011, 133, 909-919.
26. M. Wang, Y. Tang, S. Sato, L. Vugmeyster, C. J. McKnight, D. P. Raleigh, *Journal of the American Chemical Society* 2003, 125, 6032-6033.
27. J. Kubelka, W. A. Eaton, J. Hofrichter, *Journal of Molecular Biology* 2003, 329, 625-630.

28. S. Rauscher, V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. de Groot, H. Grubmüller, *Journal of Chemical Theory and Computation* 2015, 11, 5513-5524.
29. S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, L. Reymond, D. Nettels, B. Schuler, *Proceedings of the National Academy of Sciences* 2010, 107, 14609-14614.
30. R. Wuttke, H. Hofmann, D. Nettels, M. B. Borgia, J. Mittal, R. B. Best, B. Schuler, *Proceedings of the National Academy of Sciences* 2014, 111, 5213-5218.
31. B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *Journal of Chemical Theory and Computation* 2008, 4, 435-447.
32. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, *Bioinformatics* 2013, 29, 845-854.
33. B. Hess, *Journal of Chemical Theory and Computation* 2008, 4, 116-122.
34. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *The Journal of Chemical Physics* 1995, 103, 8577-8593.
35. G. Bussi, D. Donadio, M. Parrinello, *The Journal of Chemical Physics* 2007, 126, 014101.
36. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *The Journal of Chemical Physics* 1984, 81, 3684-3690.
37. M. Parrinello, A. Rahman, *Journal of Applied Physics* 1981, 52, 7182-7190.
38. Y. Sugita, Y. Okamoto, *Chemical Physics Letters* 1999, 314, 141-151.
39. P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts, V. S. Pande, *Journal of Chemical Theory and Computation* 2013, 9, 461-469.
40. T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* 1993, 98, 10089-10092.
41. P. Eastman, V. S. Pande, *Journal of Chemical Theory and Computation* 2010, 6, 434-437.
42. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Research* 2000, 28, 235-242.
43. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* 2009, 30, 1545-1614.
44. S. Jo, T. Kim, V. G. Iyer, W. Im, *Journal of Computational Chemistry* 2008, 29, 1859-1865.
45. S. Piana, A. G. Donchev, P. Robustelli, D. E. Shaw, *The Journal of Physical Chemistry B* 2015, 119, 5113-5123.
46. J. Huang, Alexander D. MacKerell Jr, *Biophysical Journal* 2014, 107, 991-997.
47. J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* 1977, 23, 327-341.
48. M. Feig, J. Karanicolas, C. L. Brooks III, *J. Mol. Graph. Model.* 2004, 22, 377-395.
49. A. Grossfield. University of Rochester, 2008.
50. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *Journal of Computational Chemistry* 2005, 26, 1781-1802.
51. M. R. Shirts, J. D. Chodera, *The Journal of Chemical Physics* 2008, 129, 124105.