Further optimization of a hybrid united-atom and coarse-grained force field for folding simulations: Improved backbone hydration and interactions between charged side chains

Wei Han,[†,‡] and Klaus Schulten[*,†,‡]

[†] Beckman Institute and [‡] Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, USA

[*]Please address all correspondence to kschulte@ks.uiuc.edu

Supporting Information

Calculation of hydration free energy and potential mean force Hydration free energy, ΔG_{HFE}, is defined as the free energy difference between the state where a solute is immersed in water and the state where the solute is isolated from water. The thermodynamic integration (TI) method [1] calculates hydration free energy by introducing a coupling parameter (λ) to interaction potentials between a solute and water. Through choosing a series of windows with varying λs, ΔG_{HFE} can be calculated as

$$\Delta G_{\mathrm{HFE}} = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \,, \tag{S1}$$

where $H(\lambda)$ is parameterized Hamiltonian, with $\lambda = 0$ for fully present interactions between the solute and water, and with $\lambda = 1$ for complete decoupling between the solute and water. To avoid the singularity problem when λ is close to unity or zero, a soft-core LJ potential was applied [2]. As we are only interested in HFEs of single peptide conformation, starting structures were constrained by applying a harmonic potential ($K = 300$ kJ·mol^{-1}rad^{-2}) on all its ϕ and ψ backbone dihedral angles. For TI calculations with all-atom force fields, the particle mesh Ewald summation was used to account for lone-range electrostatic interactions [3]. We carried out a three-stage procedure by turning off separately electrostatic and vdW interactions between a solute and water, as suggested by Shirts and coworkers [4], for better convergence. We first turned off electrostatic interactions by discharging a solute while keeping vdW interactions on. The following λ_{ele} values were used: 0.0, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.15, 0.25, 0.5, 0.75 and 1.0. Then we decoupled vdW interactions through λ_{vdW} windows of 0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95 and 1.0. Finally, we recharged the solute through five λ windows. For PACE, as there are no electrostatic terms for interactions between solutes and CG water, we can turn off the interactions in one TI step, through λ windows of 0.0, 0.05, 0.1, 0.2 ,0.3 ,0.4 ,0.5 , 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95 and 1.0. Examples of plots of $\langle dH/d\lambda \rangle$ against λ are shown in Figure S1. For each HFE, we performed three independent TI simulations to estimate average values and errors. The systems studied with TI calculations are summarized in Table S1.

Potential mean forces (PMF) were calculated in the same way as previous studies [5]. The calculations were performed in a rectangular box with size of 5.0×3.0×3.0 nm. A pair of solute molecules were placed in the center of the box solvated by about 500 CG water particles. To obtain 1D PMFs of solute pairs, solutes were constrained to move along a straight line. The positions of a solute pair are restrained so that the pair is either in a close contact or separated far away (> 1.0 nm). The positions can be made a function of a coupling parameter λ. As λ slowly varies from zero to unity through 50 ns simulations, the solutes are moving away from each other. The free energy difference between the pair in any intermediate distance and the close contact is determined as the accumulated $\langle dH/d\lambda \rangle d\lambda$. The free energy in the longest distance is set to zero as a reference point. In each PMF simulation, 100 evenly distributed intermediate distances were chosen to plot PMF profiles and, for each pair, five PMF simulations were performed to generate the average result.

GROMACS 3.3 was used in both HFE and PMF calculations. All simulation parameters are the same as described in the main text, except that in GROMACS, constant temperature and pressure was maintained by a Nosé-Hoover thermostat [6] with $\tau_{\mathrm{T}} = 1.0$ ps^{-1}, and a Parrinello-Rahman barostat [7] with $\tau_{\mathrm{P}} = 2.0$ ps^{-1}, respectively.

Modification of backbone dihedral potentials To illustrate how change of hydration parameters for backbones affects backbone conformations, we carried out REMD simulations to generate equilibrium (ϕ, ψ) distributions of dialanine (Ace-Ala-NMe), and then derive its free energy map at 300 K. The map was compared with those from a coil library (Figure S2a) [8], and from simulations using the original PACE (Figure S2b). As shown in the figure, the re-optimized hydration parameters do not significantly alter (ϕ, ψ) free energy surface, except that the free energy difference between α and polyproline II (PPII) conformations (Figure S2c) becomes 0.1 ± 0.1 kJ·mol^{-1}, smaller compared to 1.4 ± 0.2 kJ·mol^{-1} for the original PACE. The α-PPII free energy difference was estimated to be ~ 2.0 kJ·mol^{-1} through analysis of the coil library, or ~ 1.0 kJ·mol^{-1} through high-level quantum mechanics calculation in aqueous medium ($\epsilon_{\mathrm{r}} = 80$) [9]. As α and PPII differ from each other mainly in ψ (Figure S2c), we refined the dihedral terms related to ψ, *i.e.*, force constant K_{dih_n} with

$n = 1$, as described in Eq 3. The re-optimized dihedral parameters, as shown in Table S3, raised the α-PPII free energy difference to 1.3 ± 0.2 kJ·mol^{-1} (Figure S2d).

For most of the other amino acids, the modification of backbone hydration parameters generally causes a shift in (ϕ, ψ) distributions in favor of α conformations. The shift was eliminated by refinement of the dihedral parameters relevant to ϕ, ψ and χ_1 (Table S3). The resulting parameters yield a good match between side-chain rotamer distributions from the coil library and the simulated distributions with new PACE. The fitting between two two distributions leads to $R^2 = 0.903$ and the slope of the fitted line as 1.0006 (Figure S3a). To assess how well the new PACE could reproduce the (ϕ, ψ) dependence on χ_1 rotamers, we calculated the similarity between the (ϕ, ψ) distributions, upon the three rotamers, from the coil library and the simulations, using a cosine similarity score S that is expressed as

$$S(\mathcal{X}\{x_i\}, \mathcal{Y}\{y_i\}) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}}, \tag{S2}$$

where $\mathcal{X}\{x_i\}$ and $\mathcal{Y}\{y_i\}$ are sets of data for two (ϕ, ψ) distributions. Each data point in the sets stands for probability of backbone dihedral angles being in a certain $10^o \times 10^o$ grid on the (ϕ, ψ) map. S should range between 0-1 with 1 for a perfect match. The calculation of S has been done in previous studies to compare (ϕ, ψ) distributions which showed that a S value of 0.8 is a robust indicator of a good match between two distributions [8]. As shown in Figure S3b, S values between the distributions from the coil library and simulations are, for most of the cases, above 0.8. The average S value is 0.85 ± 0.08, slightly better than 0.83 ± 0.10 for original PACE.

Parametrization of hydrogen bond interactions for backbones In the original PACE, backbone HB terms, as shown in Eq 5, were parameterized through reproducing secondary-structure contents of specific α-helical and β-sheet peptides [5]. To improve the transferability of PACE, we used PMF of amide-amide interactions, simulated with all-atom force fields, as a more general reference for fitting. Following Masunov and Lazaridis [10], we generated a 1D PMF of amide-amide interaction in a coplanar and head-to-head direction

(Figure S4), through the TI method.

The backbone HB parameters were empirically adjusted through repeated simulations so that the difference between the PMFs generated by all-atom and PACE force fields was minimized. In the PMF generated by the re-optimized HB parameters (Figure S4), the first free energy minimum, namely contact HB pair, which is -8.0 ± 0.8 kJ·mol^{-1} deep, is located at $r_{O-N} = 0.28$ nm, consistent with both position ($r_{\mathrm{O-N}} = 0.28$ nm) and depth (-9.0 ± 1.0 kJ·mol^{-1}) of the first minimum in the all-atom PMF. The height of the first barrier is $\sim$ 10.5 kJ·mol^{-1}, lower than $\sim$ 12.1 kJ·mol^{-1} in the all-atom PMF. The second minimum at $r_{\mathrm{O-N}} = 0.52$ nm in the all-atom PMF, known as water-bridged HB pair [10], is not captured by re-optimized PACE, presumably due to the lack of explicit HB interactions between CG water and backbone amide groups.

The new backbone HB parameters, as listed in Table S2, are similar to those in the original PACE, except that $\epsilon_{\mathrm{attr,O-N}}$, responsible for HB attraction (Eq 5), was reduced from originally 20.7 kJ·mol^{-1} to 17.7 kJ·mol^{-1}. The reduction of $\epsilon_{\mathrm{attr,O-N}}$ is expected since with reduced hydration of backbone amide groups, the cost of dehydration for HB formation is also decreased. For the same reason, other HB parameters involved with backbone amides may also be affected. As such, we examined and, when necessary, modified them by fitting respective all-atom PMFs. All the modified parameters are summarized in Table S2.

LJ parameters of charged-pair interactions As electrostatic interactions between charged-pairs have been modeled by Coulomb potentials in the new PACE, all LJ parameters $\epsilon_{\mathrm{A-B}}$ and $\sigma_{\mathrm{A-B}}$ for interactions between atoms A and B in charged pairs were generated by the combination role, $\epsilon_{\mathrm{A-B}} = \sqrt{\epsilon_{\mathrm{A-A}}\epsilon_{\mathrm{B-B}}}$ and $\sigma_{\mathrm{A-B}} = (\sigma_{\mathrm{A-A}} + \sigma_{\mathrm{B-B}})/2$, using values of $\epsilon_{\mathrm{A-A}}$ and $\sigma_{\mathrm{A-A}}$ in the original PACE [5]. For Arg^+...Asp^-/Glu^- pair, the contact pair minimum is modeled, in addition to Coulomb potentials, by effective potentials, as shown in Figure S5 and the equation below:

$$E_{\mathrm{R...D/E}} = \sum 4\epsilon_{\mathrm{attr,O_{D/E}-N_R}} \left(\frac{\sigma^{12}_{\mathrm{O_{D/E}-N_R}}}{r^{12}_{\mathrm{O_{D/E}-N_R}}} - \frac{\sigma^{6}_{\mathrm{O_{D/E}-N_R}}}{r^{6}_{\mathrm{O_{D/E}-N_R}}} \right) + \sum \frac{C_{\mathrm{rep,C_{D/E}-C_R}}}{r^{6}_{\mathrm{C_{D/E}-C_R}}}, \quad \text{(S3)}$$

where R, D and E are one-letter names of amino acids. The optimized parameters are: $\epsilon_{\mathrm{attr,O_{D/E}-N_R}} = 9.1$ kJ·mol^{-1}, $\sigma_{\mathrm{O_{D/E}-N_R}} = 0.243$ nm and $C_{\mathrm{rep,C_{D/E}-C_R}} = 0.05$ kJ·mol^{-1}nm^6.

References

[1] van Gunsteren, W. F.; Berendsen, H. J. C. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 171-176.

[2] Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Greber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529-539.

[3] Bogusz, S.; Cheatham, T. E.; Brooks, B. R. *J. Chem. Phys.* **1998**, *108*, 7070-7084.

[4] Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740-5761.

[5] Han, W.; Wan, C.-K.; Jiang, F.; Wu, Y.-D. *J. Chem. Theory Comput.* **2010**, *6*, 3373-3389.

[6] Nose, S. *J. Chem. Phys.* **1984**, *81*, 511519.

[7] Parinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, emph52, 7182.

[8] Jiang, F.; Han, W.; Wu, Y.-D. *J. Phys. Chem. B* **2010**, *114*, 5840-5850.

[9] Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. *J. Comput. Chem.* **2006**, *27*, 781-790.

[10] Masunov, A.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 1722-1730.

[11] Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B. *Biopolymers* **2003**, *68*, 91-109.

Table S1: Summary of hydration free energies (HFE) calculations, including peptides, force fields used in simulations, the number of water molecules, length of each λ window and averaged HFEs.

System	Force Field[a]	n_{water}	Time per window (ns)	HFE ($kJ \cdot mol^{-1}$)
Ala_1	OPLS	815	5	-48.2 ± 0.7
	PACEold	507	5	-78.8 ± 0.6
	PACEnew	507	5	-41.5 ± 0.8
Ala_2	OPLS	1007	5	-69.0 ± 1.0
	PACEold	541	5	-121.0 ± 1.0
	PACEnew	541	5	-64.3 ± 0.8
Ala_3	OPLS	1236	5	-88.4 ± 1.2
	PACEold	875	5	-162.0 ± 1.1
	PACEnew	875	5	-86.7 ± 1.1
Ala_4	OPLS	1701	5	-107.6 ± 1.9
	PACEold	987	5	-203.0 ± 1.5
	PACEnew	987	5	-110.4 ± 1.3

[a]OPLS, PACEold and PACEnew denote OPLS-AA, the original PACE and the new PACE force fields, respectively.

Table S2: Summary of the modified parameters[a] for polar interactions involving backbone.

A...B[b]	$\epsilon_{\mathrm{attr,A-B}}$[c] ($\mathrm{kJ{\cdot}mol^{-1}}$)	$\sigma_{\mathrm{attr,A-B}}$[c] (nm)
bb...bb		
$\mathrm{O_{bb}...N_{bb}}$	17.7	0.24
$\mathrm{O_{bb,}}_{i}\mathrm{...N_{bb,}}_{i+3/i+4}$	14.7	0.24
$\mathrm{C_{bb}...C}\alpha_{\mathrm{bb}}$	2.6	0.33
$\mathrm{C_{bb}...C_{bb}}$	2.6	0.33
sc amide...bb		
$\mathrm{O_{sc}...H_{bb}}$	27.0	0.16
$\mathrm{H_{sc}...O_{bb}}$	20.0	0.16
-$\mathrm{COO^-}$...bb		
$\mathrm{O_{sc}...H_{bb}}$	25.0	0.16
-$\mathrm{NH_3^+}$...bb		
$\mathrm{N_{sc}...O_{bb}}$	6.0	0.26
-NH-$\mathrm{C^+}$-$\mathrm{(NH_2)_2}$...bb		
$\mathrm{N_{sc}...O_{bb}}$	4.5	0.26
-OH...bb		
$\mathrm{O_{sc}...O_{bb}}$	0.8	0.28
CGW...bb		
$\mathrm{CGW...O_{bb}}$	4.0	0.36
$\mathrm{CGW...H_{bb}}$	3.0	0.36

[a]Only the parameters that are different from those in the original PACE are listed in the table. [b]"bb" and "sc" denote backbone and side chain, respectively. [c]The parameters are for a LJ potential: $4\epsilon\left(\frac{\sigma^{12}}{r^{12}}-\frac{\sigma^{6}}{r^{6}}\right)$.

Table S3: Summary of refined parameters[a] of dihedral angle terms K_{dih_n} (kJ·mol^{-1}), n and ζ_{0_n} (deg) in Eq 3.

	Dihedral	K_{dih_n}	n	ζ_0
Ala	ψ	1.7	1	0.0
Gly	ϕ	2.0	3	0.0
		3.0	1	-180.0
Lys[b]	ψ	1.0	1	0.0
Asn[c]	ψ	1.0	1	0.0
Asp[c]	ψ	1.5	1	0.0
Cys	ψ	2.0	1	0.0
Ser	ψ	1.5	1	0.0
	χ_1	2.0	1	-120.0
Pro	ψ	3.0	1	0.0
His	ψ	2.0	1	0.0

[a]Only the parameters different from the original PACE are shown in the table. [b]Arg, Gln, Glu, Met and Leu have the same dihedral parameters as Lys. [c]The parameters ($\epsilon_{\mathrm{short},ij}$ in Eq 3) for short-range pairs between $O_{\delta 1/\delta 2}$ in Asp or $O_{\delta 1}/H_{\delta 21}$ in Asn and the adjacent backbone amides were set to 19.0 kJ·mol^{-1}.

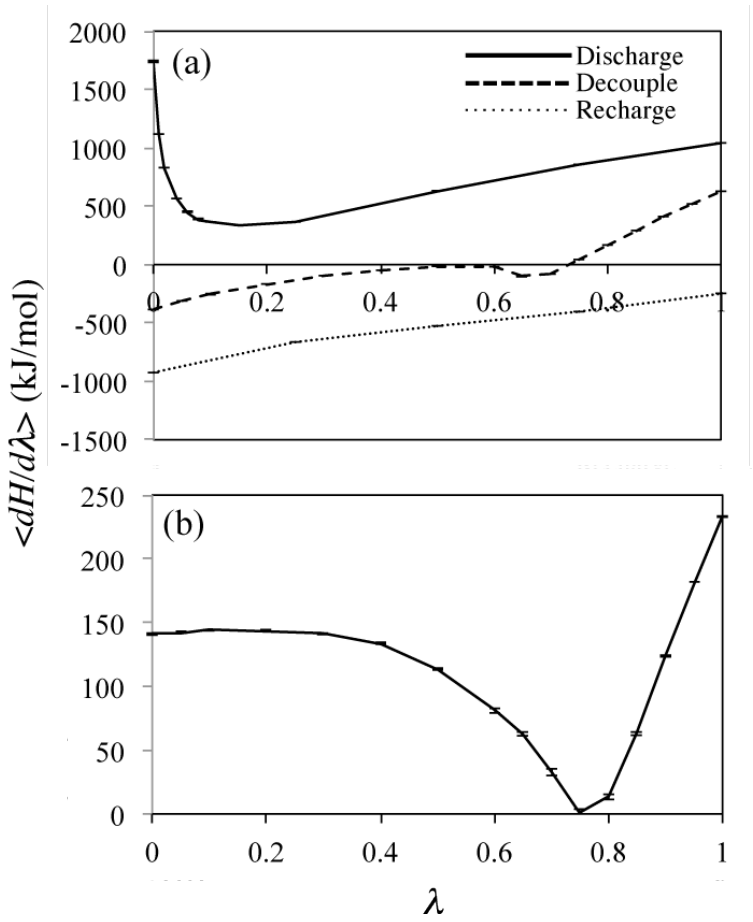

Figure S1: Plots of $\langle dH/d\lambda \rangle$ against λ for Ala_4 with OPLS-AA in TIP3P water (a) and Ala_4 with the new PACE (b). The error bars in each λ window were estimated through a block average with a bin size of 1 ns.

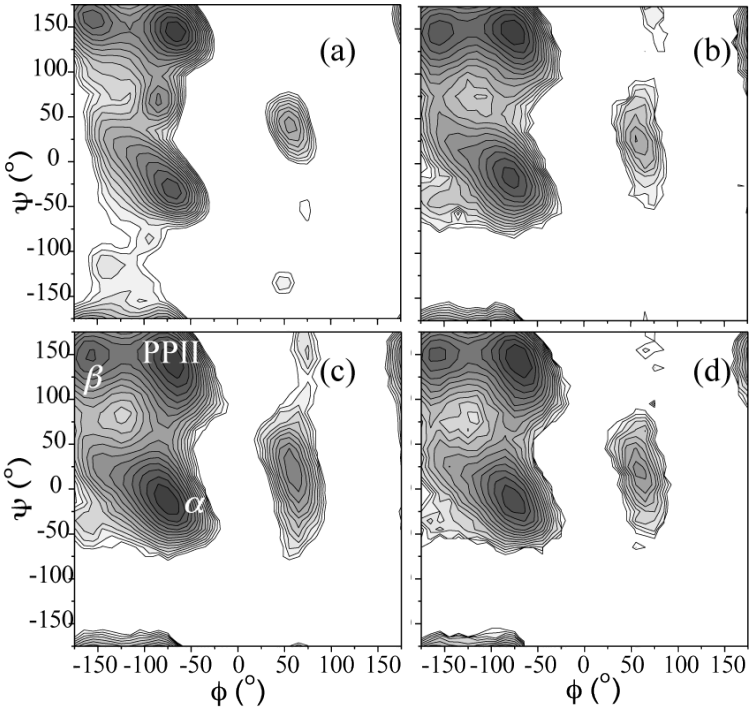

Figure S2: $(\phi\,,\psi)$ free energies maps of dialanine obtained from the coil library (a) and from REMD simulations using original PACE (b), using re-optimized hydration parameters only (c) and using both re-optimized hydration and backbone parameters (d). The gap between neighboring contour lines denotes 1.0 kJ·mol^{-1} free energy difference at 300 K.

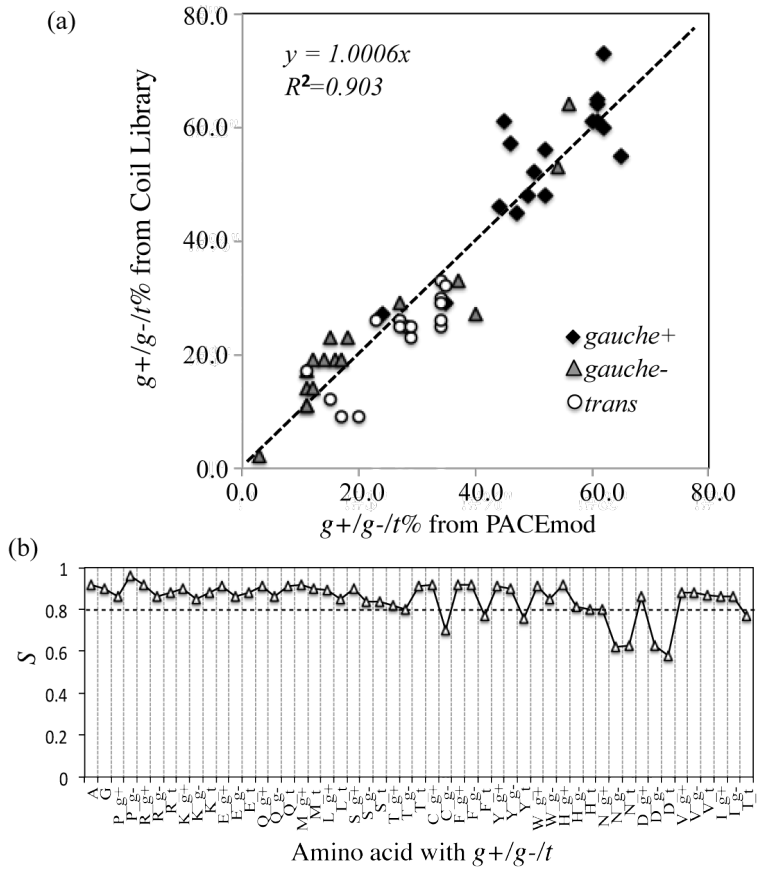

Figure S3: (a) Plot of χ_1 rotamer distributions from the coil library against those from simulation using PACE with the refined backbone hydration and dihedral parameters. (b) Scores (S) of similarity between (ϕ, ψ) distributions from the coil library and those simulated using PACE with the refined parameters.

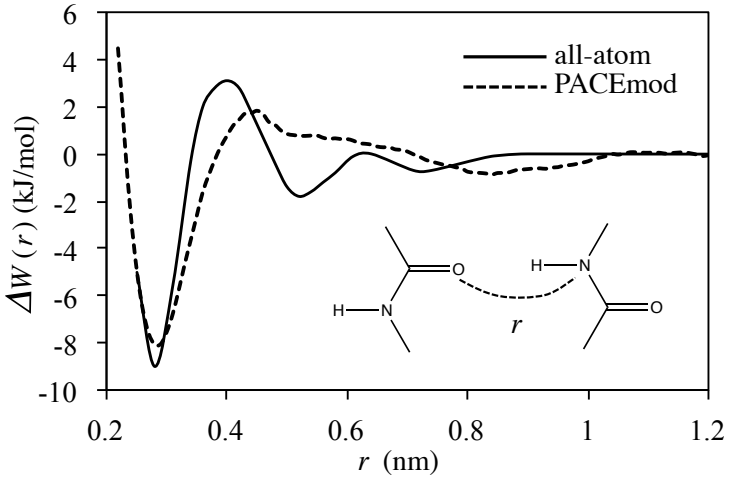

Figure S4: Potential mean forces of amide-amide interactions by OPLS-AA and TIP3P water (solid line) and new PACE (dashed line).

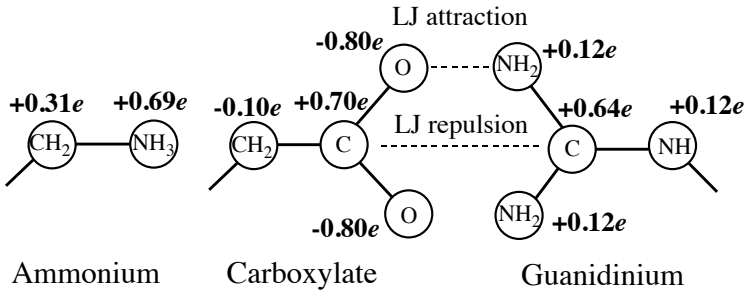

Figure S5: Schematic representation of ammonium, carboxylate and guanidinium groups in PACE and their assigned partial charges.

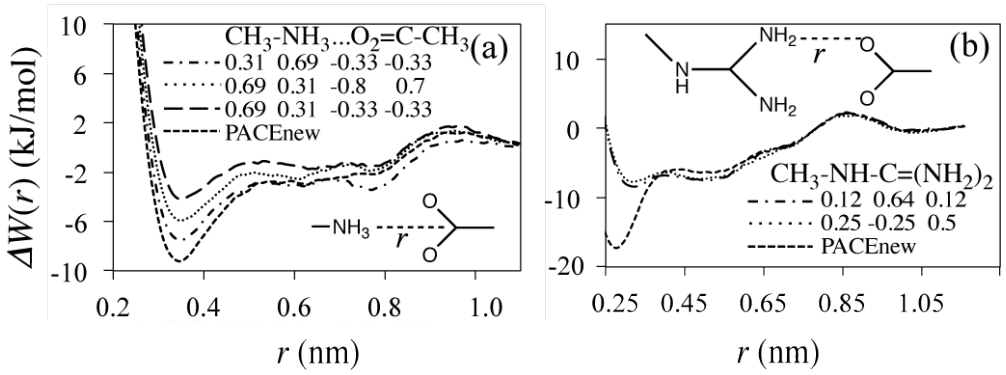

Figure S6: The different schemes of partial charges and the corresponding PMFs for Lys^+...Asp^-/Glu^- (a) and Arg^+...Asp^-/Glu^- (b).

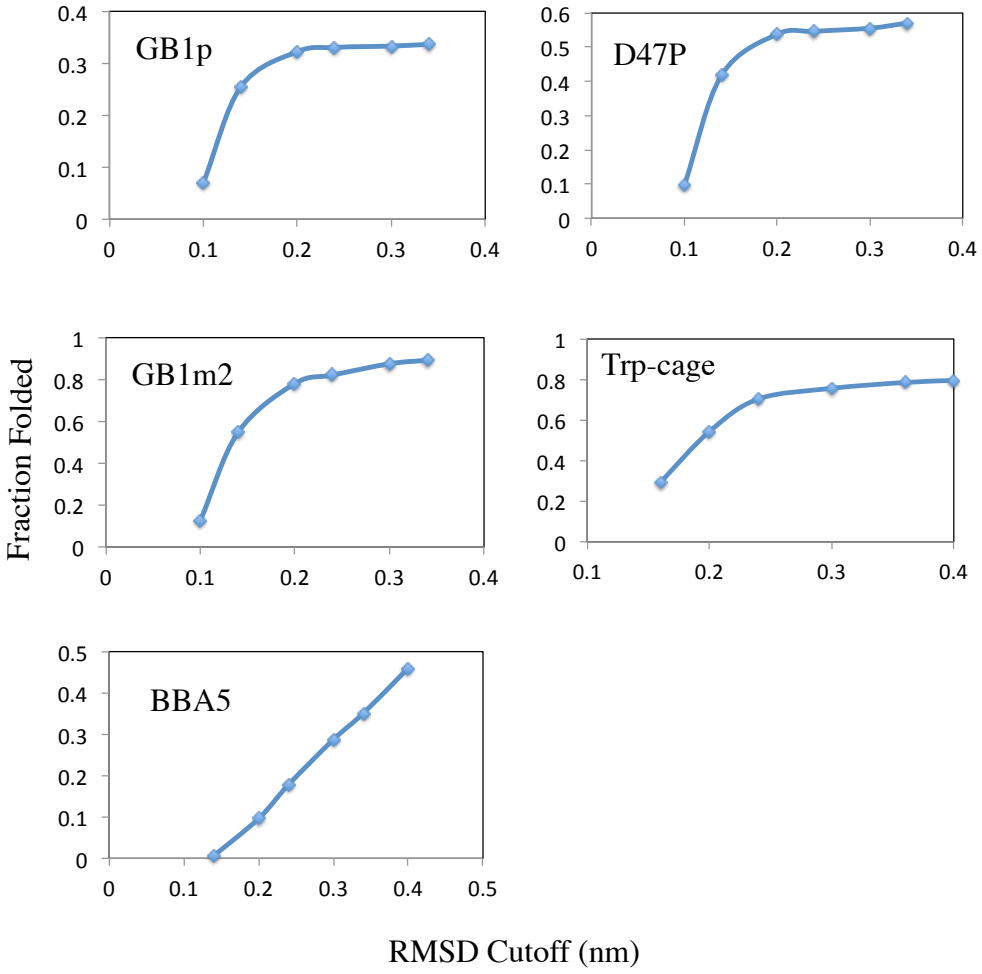

Figure S7: Fractions of folded structures determined with different RMSD cutoff.

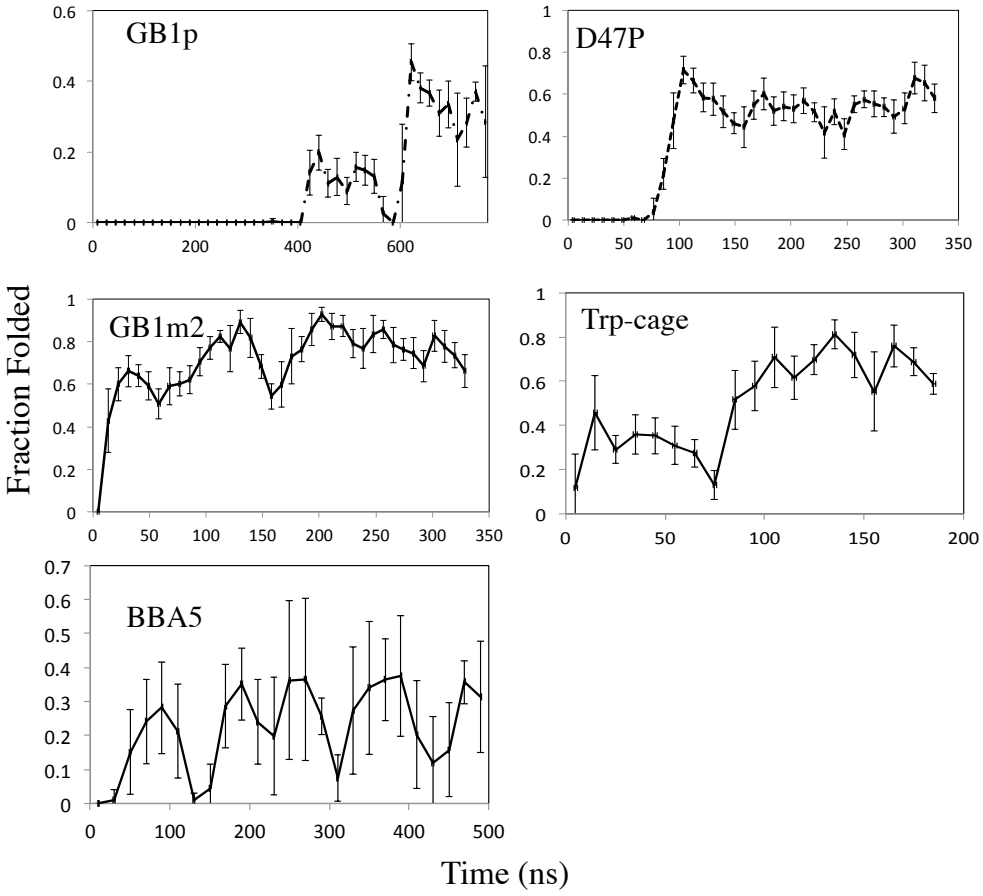

Figure S8: Fractions of folded structures against simulation time during REMD simulations. Each point in the plots denotes an average value over a block of 10 (D47P, GB1m2 and Trp-cage) or 20 ns (GB1p and BBA5). The error bars for the points were estimated with block average with a bin size of 1 ns.

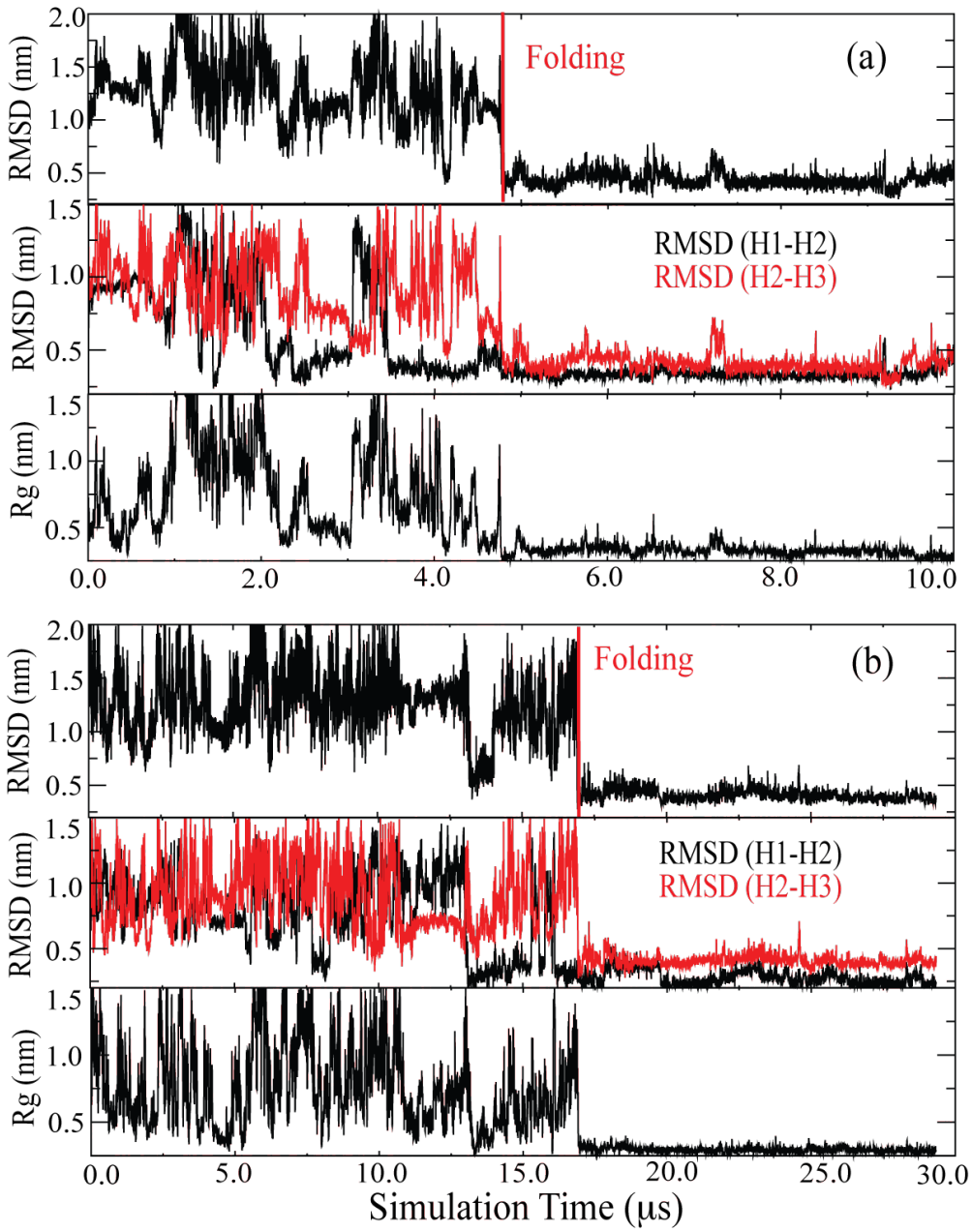

Figure S9: RMSD and Rg of two α3D simulations (a) and (b), with top panels for all C_α RMSD, middle panels for RMSDs of H1-H2 (black) and H2-H3 (red) and bottom panels for Rg.Folded states are reached when C_α RMSD is below 0.35 nm.

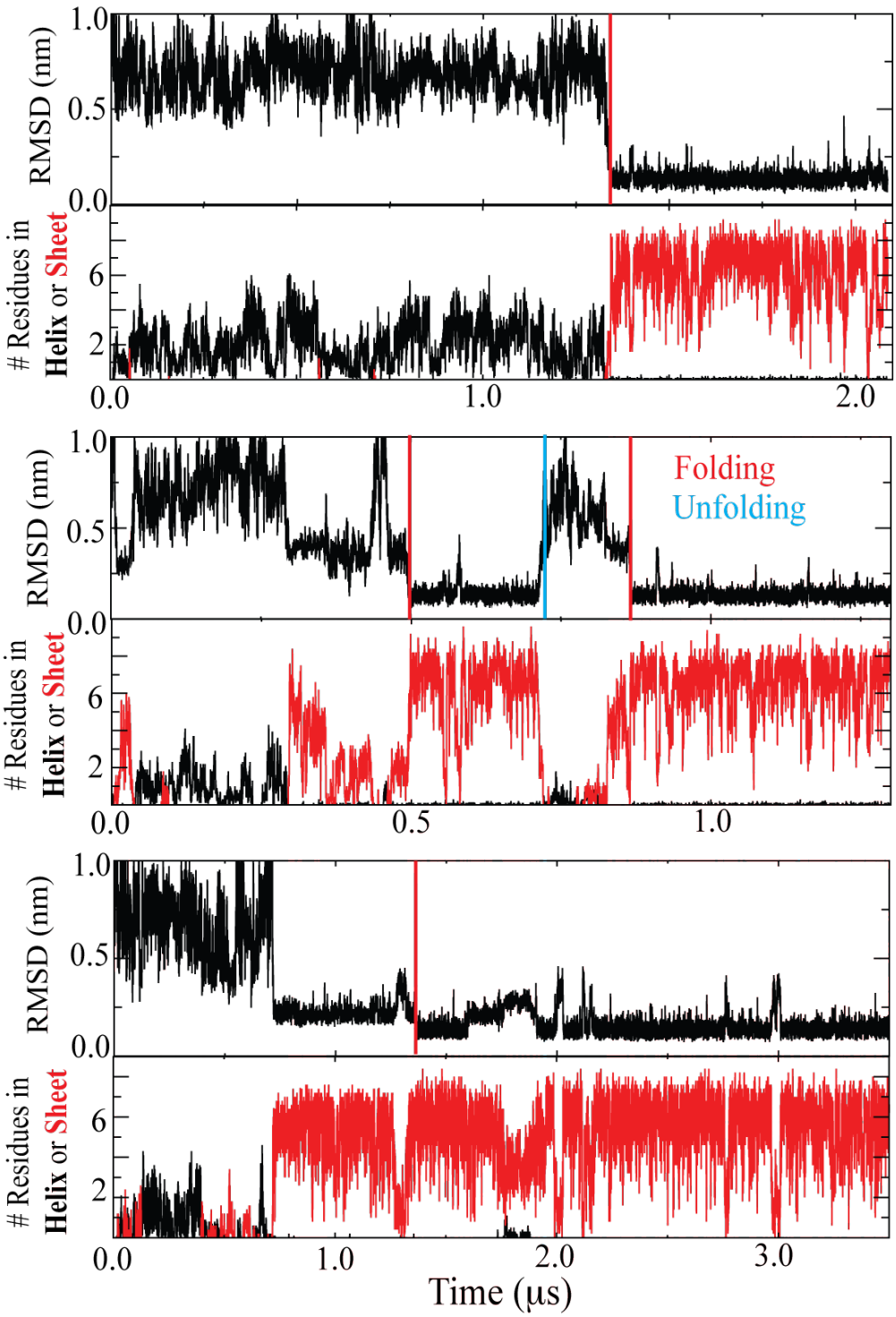

Figure S10: Backbone RMSD (top panels) and the number of residues (bottom panels) with secondary structures against time in standard MD simulations of D47P. The peptides fold when the RMSD drops below 0.15 nm, starting from unfolded states. The peptides unfold when the RMSD is beyond 0.6 nm, from folded states.

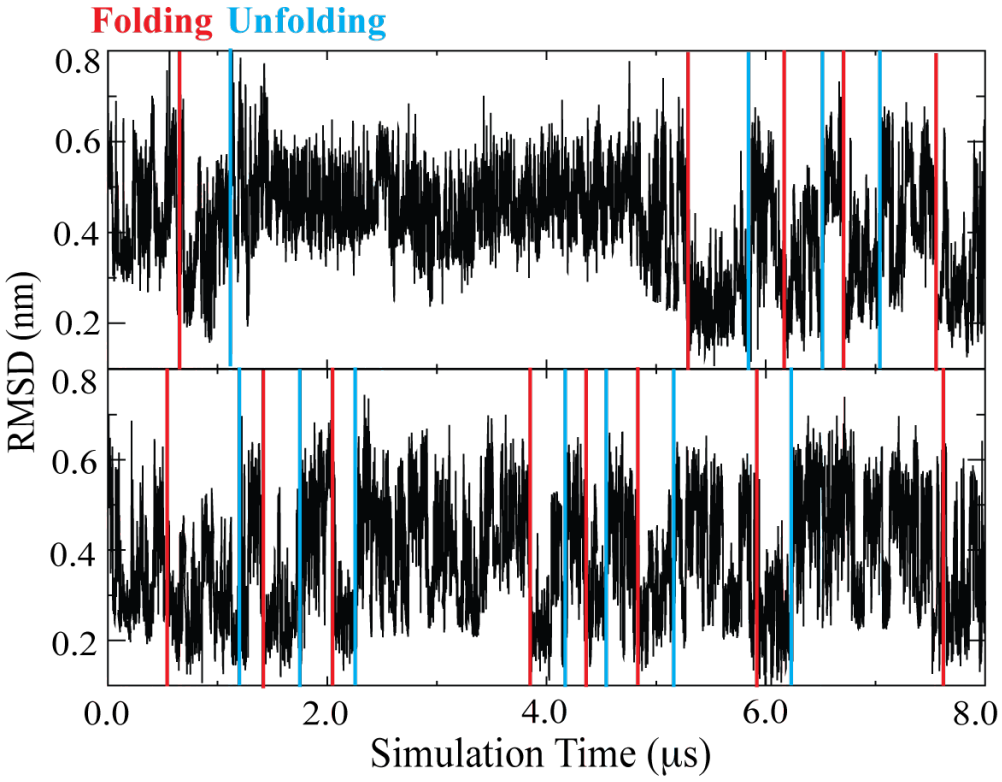

Figure S11: Backbone RMSD change during standard MD simulations of BBA5. The peptides fold when the RMSD drops below 0.2 nm, from unfolded states. The peptides unfold when the RMSD is beyond 0.6 nm, from folded states.

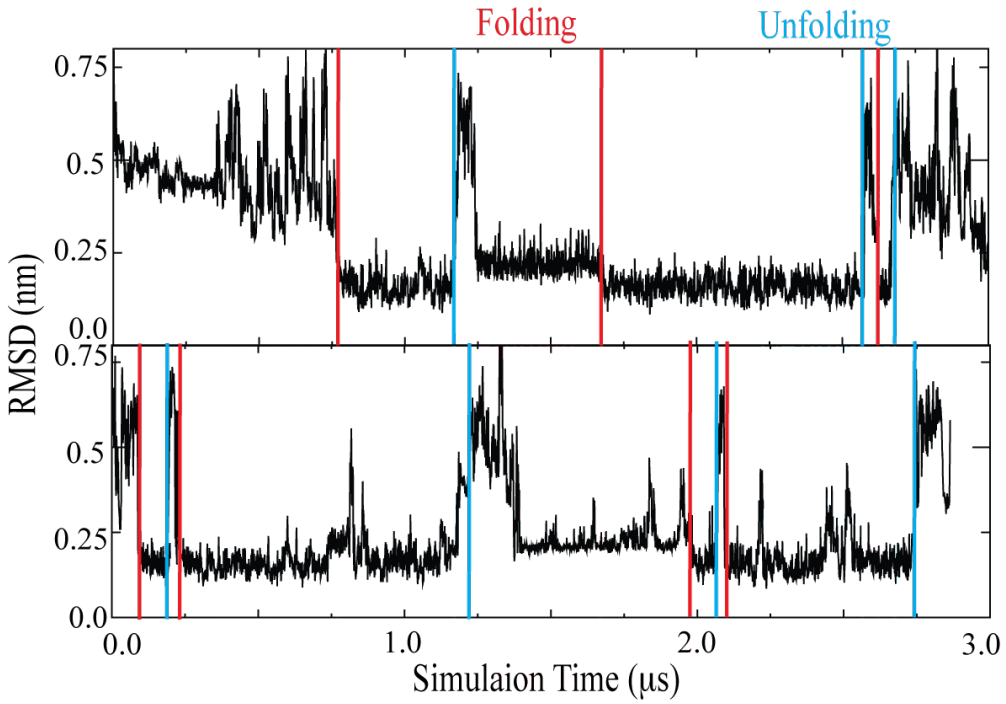

Figure S12: Backbone (3-19) RMSD change during standard MD simulations of Trp-cage. The peptides fold when the RMSD drops below 0.15 nm, from unfolded states. The peptides unfold when the RMSD is beyond 0.7 nm, from folded states.

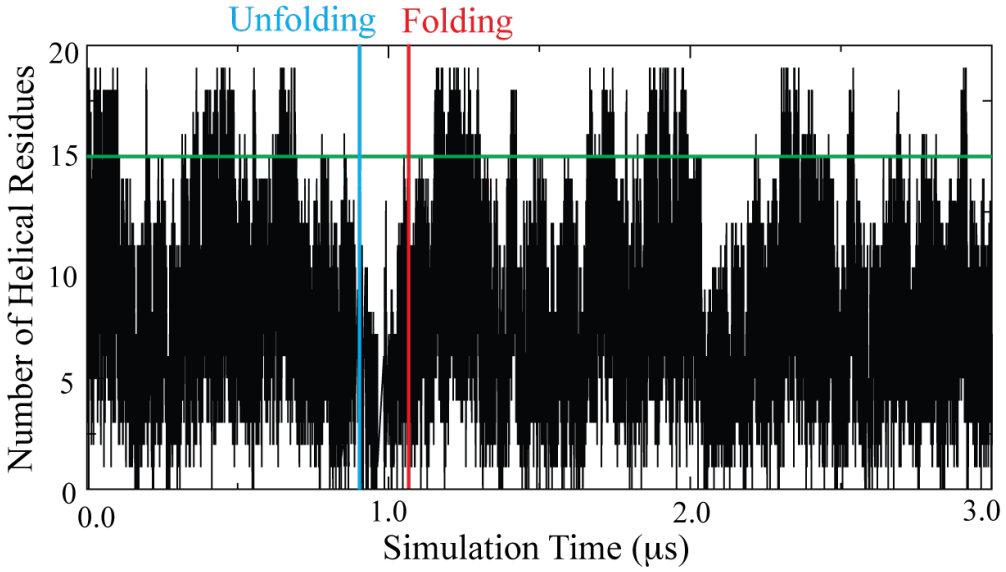

Figure S13: Number of helical residues N_h of Fs simulations. The peptides fold when N_h reaches 15 (green line), from unfolded states, as suggested by previous work [11]. The peptides unfold when N_h drops to zero, from folded states.