Supporting Information for "Protein Loop Conformational Free Energy Changes via an Alchemical Path without Reaction Coordinates"

Shima Arasteh, Bin W. Zhang,* and Ronald M. Levy*

Center for Biophysics and Computational Biology, and Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, United States

E-mail: bin.w.zhang@temple.edu; ronlevy@temple.edu

Simulation Details

We ran simulations for a solution that contains a protein Ubiquitin, 5884 TIP3P[1] water molecules, 18 Na^+ and 18 Cl^- (0.15 M) ions in a rhombic dodecahedron simulation box by using the GROMACS version 2018/8.[2] The edge length of the rhombic dodecahedron is 6.49971 nm. The system was modeled by the AMBER99SB-ILDN force field[3]. The PDB files for the protein Ubiquitin 3NHE (chain B) and 1UBQ were downloaded from the protein data bank to construct the initial structures for simulations. Before the NVT production simulations, the system was equilibrated by energy minimization and a 0.1 ns long NPT simulation. For the NPT simulations during the equilibration period, the pressure was kept at 1 bar by the Parrinello-Rahman pressure barostat with a pressure relaxation time of 2.0 ps;[4] and the temperature was kept constant at 300 K by the velocity-rescaling with a time constant $\tau_t = 0.1$ ps. For the NVT simulations during the production period, the temperature was maintained at 300 K by the leapfrog stochastic dynamics integrator (Langevin dynamics) with a time constant $\tau_t = 2.0$ ps.[5] The time step for the MD simulations is 2 fs with the LINCS algorithm used to constrain bond lengths involving hydrogen atoms.[6] The water molecules were kept rigid with SETTLE.[7] The long-range electrostatic interaction was treated using the smooth particle-mesh Ewald approach with a real-space cutoff of 1.0 nm and a spline order of 4.[8]

Umbrella Sampling

Periodic Ramachandran Map

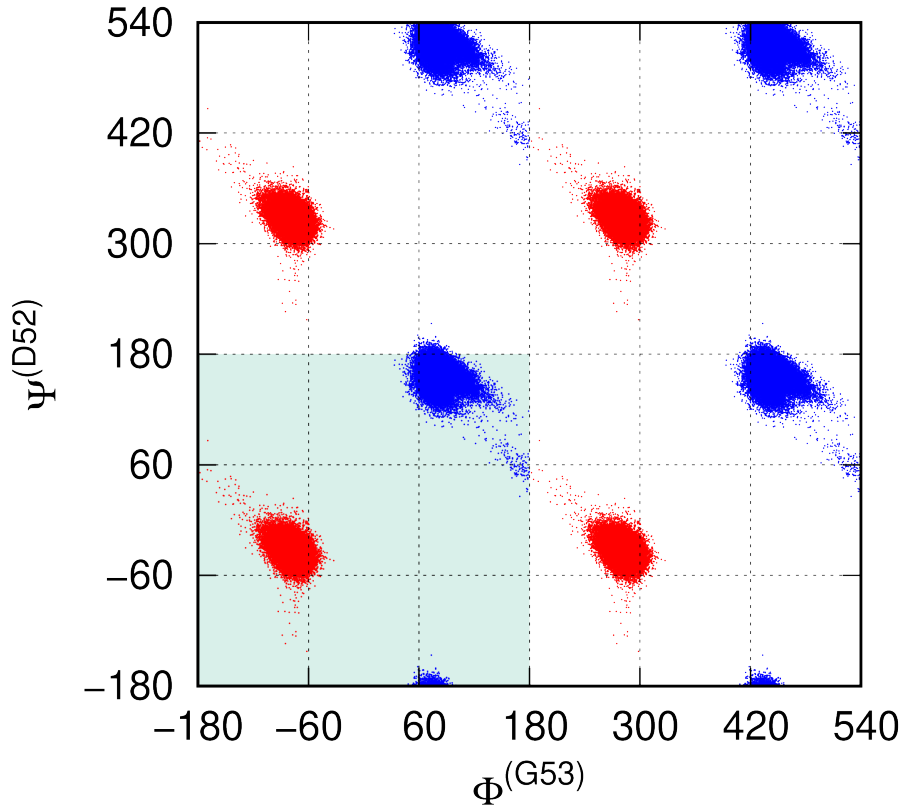

Figure S1: Periodic Ramachandran Map. The β-I states are marked with red ink; while the β-II states are marked with blue ink. This picture includes the central cell that is filled with light cyan, and three image cells.

Fig.S1 shows the periodic Ramachandran map. As can be seen, four nearest neighboring β-I states exist around each β-II state, and vice versa. Therefore, there are multiple choices to add umbrella sampling windows that leads to the β-I state when starting from one chosen β-II state. In the main text, we only show the β-I states (the final states) in the central cell and in the first image cell on the right side of the central cell when the β-II state in the central cell is chosen as the initial state.

Setup and Analyses

To construct the initial state (window) for umbrella sampling, we applied two parabolic restraints that are located at $\phi_0^{(G53)} = 90^\circ$ and $\psi_0^{(D52)} = 150^\circ$ to restrain the Ubiquitin molecule in the β-II state. The free energy change of applying these restraints on the free Ubiquitin molecule in the β-II state was estimated by free energy perturbation (FEP) with replica exchange. The force constant of the restraints were increased gradually from 0 to 200 kJ/mol/rad^2 by using 10 λ-states. The chosen FEP control parameter λ_i values are 0, 0.01, 0.03, 0.06, 0.1, 0.2, 0.3, 0.5, 0.75, and 1.0. The MD simulation at each λ-state lasted totally 10 ns. After every 1 ps MD simulation, replica exchanges attempts were performed to swap replicas between the adjacent states. To construct the final state (window) for umbrella sampling, we applied two parabolic restraints that are located at $\phi_0^{(G53)} = 270^\circ$ and $\psi_0^{(D52)} = -30^0$ to restrain the Ubiquitin molecule in the β-I state. Similarly, the free energy change of applying these restraints on the free Ubiquitin molecule in the β-I state was estimated by FEP with replica exchange.

Then the two endpoint states were connected by adding 31 umbrella sampling windows evenly distributed between the initial and final states along bridge B (or bridge A). Namely, there are totally 33 windows in the umbrella sampling simulations. The force constant of each parabolic restraint is 200 kJ/mol/rad^2. The MD simulation of each umbrella sampling window lasted totally 300 ns. Replica exchange was introduced to accelerate the convergence. After every 1 ps MD simulation, replica exchanges attempts were performed to swap replicas between the adjacent umbrella sampling windows. The data generated from FEP and umbrella sampling were analyzed by using the UWHAM software package.[9]

The final result and the uncertainties were estimated by dividing the whole trajectories into 10 equally long blocks and calculating the mean of the estimates of each block and the standard error of the mean, respectively.

R-FEP-R 2.0

Dihedral Angle Restraints

Table S1: Reference values of dihedral angle restraints applied in R-FEP-R 2.0

atom numbers	type-II	type-I	notes
827, 826, 828, 830	0		
815, 826, 828, 830	180		
826, 828, 830, 831	50		
826, 828, 830, 832	170		
828, 830, 832, 833	−170		
828, 830, 832, 834	70		
828, 830, 832, 835	−50		
830, 832, 835, 836	90		
830, 832, 835, 837	−90		
826, 828, 830, 838	−70		$\phi^{(D52)}$
839, 838, 840, 841	180		
828, 830, 838, 840	150	−30	$\psi^{(D52)}$
830, 838, 840, 841	0		
842, 845, 847, 848	0		
842, 845, 847, 849	180		
843, 842, 845, 847	120		
844, 842, 845, 847	−120		
838, 840, 842, 845	90	−90	$\phi^{(G53)}$
840, 842, 845, 847	0		$\psi^{(G53)}$
839, 838, 840, 842	0		
830, 838, 840, 842	180		

Table S1 lists the reference values of the harmonic dihedral restraints applied to the atoms in the dual sets. They were chosen based on the most populated values of each dihedral angle observed in the initial state or the final state. Because the transition between the type-I and type-II β-turn only causes little disturbances to the surrounding side chains and peptides, the reference values of the harmonic dihedral restrains on the $\phi^{(G53)}$ and $\psi^{(D52)}$ are the only differences between the restraints applied to the two dual sets. The force constant of each parabolic dihedral angle restraint is 1000 kJ/mol/rad^2.

Setup of FEP Simulations

There are 5 FEP simulations in R-FEP-R 2.0. The number of states and chosen λ values of each FEP are listed in Table. S2.

Table S2: Number of states and λ values in R-FEP-R 2.0. In the FEP simulations that estimate ΔG_0 and ΔG_1, the restraints are fully turned on at the $\lambda = 1$ state. In the FEP simulations that estimate ΔG_{vr}^{CL} and ΔG_{rv}^{CL}, the bond length, bond angle, improper dihedral angle interactions, and proper dihedral angle restraints between the atoms N840 and C842 are fully broken at the $\lambda = 1$ state. Note that λ_i in Eq.(11) in the main text has been changed to $(1 - \lambda_i)$ in the FEP simulations that estimate ΔG_{vr}^{CL} and ΔG_{rv}^{CL}.

	# of states	λ values
ΔG_0 and ΔG_1	20	0, 0.001, 0.003, 0.006, 0.01, 0.02, 0.04, 0.07, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0
ΔG_D	24	0, 0.001, 0.004, 0.01, 0.03, 0.06, 0.1, 0.15, 0.2, 0.29, 0.38, 0.46, 0.54, 0.62, 0.71, 0.8, 0.85, 0.9, 0.04, 0.97, 0.99, 0.996, 0.999, 1.0
ΔG_{vr}^{CL} and ΔG_{rv}^{CL}	20	0, 0.15, 0.3, 0.45, 0.6, 0.7, 0.8, 0.85, 0.9, 0.93, 0.95, 0.96, 0.97, 0.98, 0.985, 0.99, 0.994, 0.997, 0.999, 1.0

All FEP simulations were run with replica exchange. After every 1 ps MD simulation, replica exchanges attempts were performed to swap replicas between the adjacent states. The FEP simulations run to estimate ΔG_0, ΔG_1, ΔG_D and ΔG_{rv}^{CL} all lasted 100 ns at each state. However, the FEP for the leg ΔG_{vr}^{CL} converges much slower compared with the others. We ran 300 ns simulation at each state. Only the data generated during the last 100 ns were used to estimate ΔG_{vr}^{CL}. The data generated from FEP are analyzed by using the UWHAM software package.[9] The final results and the uncertainties were estimated by dividing the whole trajectories into 10 equally long blocks and calculating the mean of the estimates of each block and the standard error of the mean, respectively.

Convergence of FEP Simulations

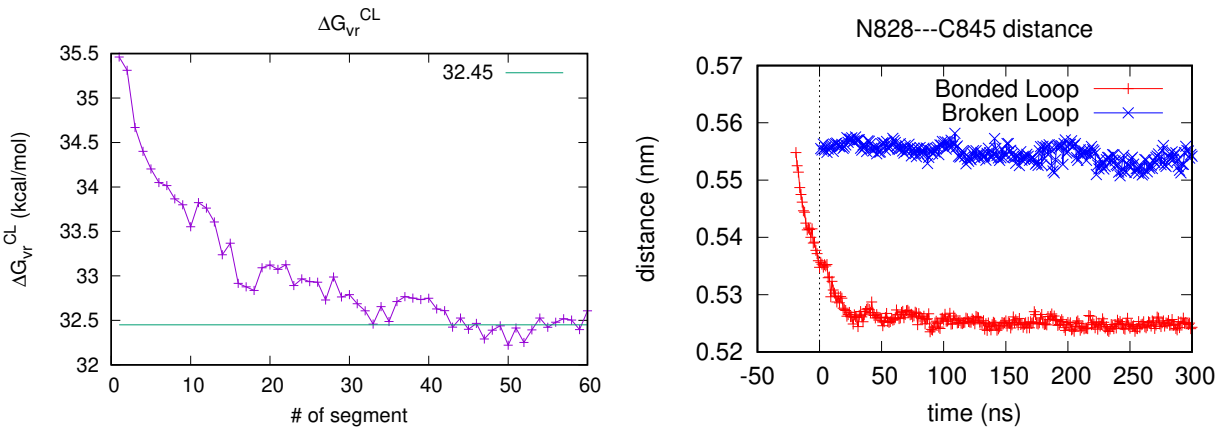

Figure S2: Convergence of the FEP simulation that estimates ΔG_{vr}^{CL}, and the distance between atoms N828 and C845 during the FEP simulation.

The FEP simulation that estimates ΔG_{vr}^{CL} is the most difficult to converge. To confirm the convergence of FEP simulations, we divide the whole trajectory at each state into multiple 5 ns long segments (blocks) and examine the dependence of the estimates of the free energy change ΔG_{vr}^{CL} on the simulation time. In the left panel of Fig.S2, each data point is an estimate of the free energy change ΔG_{vr}^{CL} estimated from a group of 5 ns long trajectories at each state during the same time interval. The left panel of Fig.S2 shows that the estimate of ΔG_{vr}^{CL} continues decreasing until reaching a plateau around $40 \times 5 = 200$ ns. Similar analyses for the other 4 FEP simulations in R-FEP-R 2.0 are shown in Fig.S3

The right panel of Fig.S2 shows the change of the distance between the atoms N828 and C845 at the endpoint states during the FEP simulation that estimates ΔG_{vr}^{CL}. The atom N828 and C845 are the two atoms in the shared set to which the two ends of the dual sets are attached. For the state that the loop is fully closed, the data points during the equilibration period (simulation time < 0) are included. As can be seen, the distance between these two atoms decreased from 0.555 ± 0.001 nm to 0.526 ± 0.001 nm during the first 50 ns simulation at the state that the loop is fully closed.

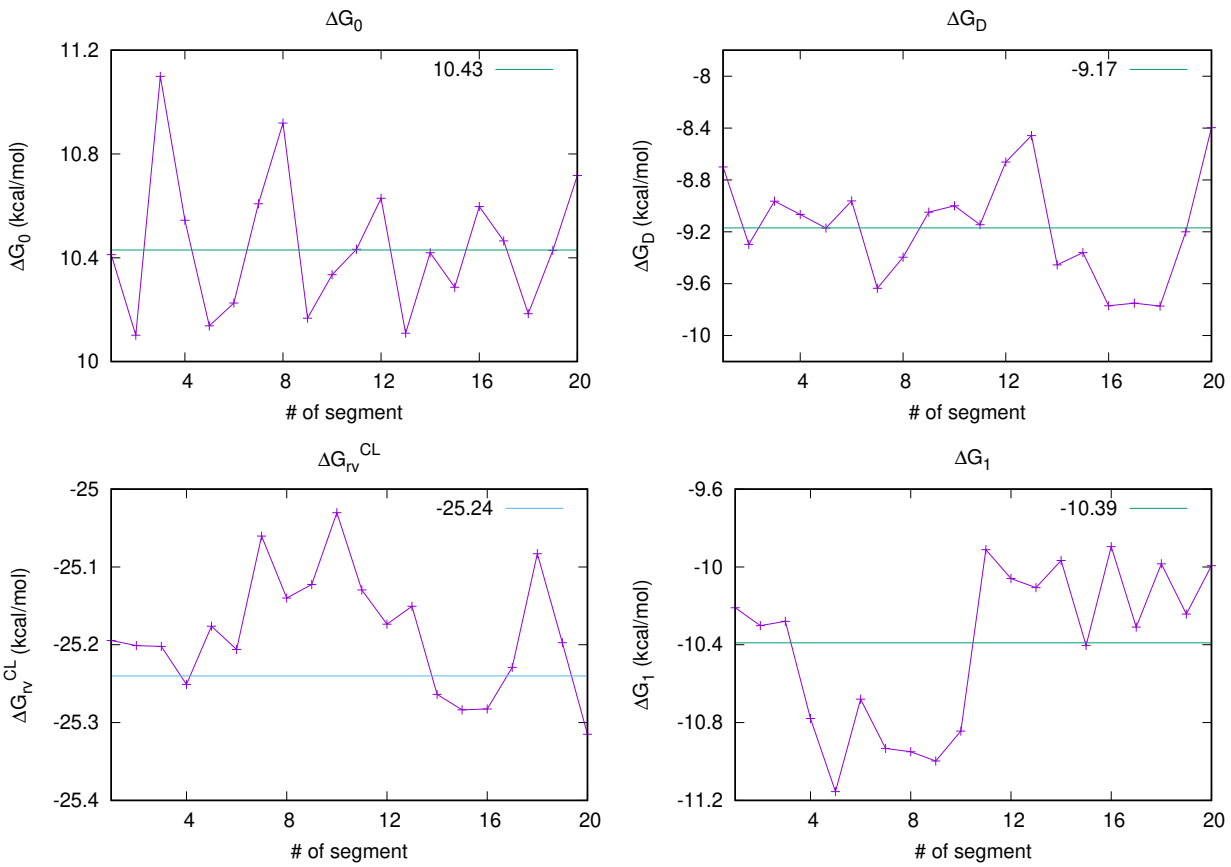

Figure S3: Convergence of the FEP simulations that estimate ΔG_0, ΔG_1, ΔG_D and ΔG_{rv}^{CL}

References

(1) 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.

(2) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(3) 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., Funct., Bioinf.* **2010**, *78*, 1950–1958.

(4) Parrinello, M.; Rahman, A. Polymorphic Transitions In Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(5) Gunsteren, W. F. V.; Berendsen, H. J. C. A Leap-frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1988**, *1*, 173–185.

(6) 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*, 1463–1472.

(7) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(8) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(9) Zhang, B. W.; Arasteh, S.; Levy, R. M. The UWHAM and SWHAM Software Package. *Sci. Rep.* **2019**, *9*, 2803.