Supporting Information for Atomistic simulations indicate the functional loop-to-coiled-coil transition in influenza hemagglutinin is not downhill

Xingcheng Lin[1,2], Jeffrey K. Noel[3], Qinghua Wang[4], Jianpeng Ma[1,4,7], and José N. Onuchic*[1,2,5,6]

[1]Center for Theoretical Biological Physics, Rice University, Houston, TX 77030
[2]Department of Physics and Astronomy, Rice University, Houston, TX 77005
[3]Crystallography, Max Delbrück Center for Molecular Medicine, Berlin, Germany 13125
[4]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030
[5]Departments of Chemistry, Rice University, Houston, TX 77005
[6]Departments of BioSciences, Rice University, Houston, TX 77005
[7]Department of Bioengineering, Rice University, Houston, TX 77030
*Email: jonuchic@rice.edu

Contents

www.pnas.org/cgi/doi/10.1073/pnas.1805442115

1 Continuous simulated tempering

The CST method[1, 2] was employed to accelerate the sampling of the B-loop conformational space. CST speeds up the sampling of biomolecular dynamics by continuously changing the temperature of a single copy of a simulation. The temperature is guided by a Langevin equation,

$$\frac{d(1/\beta)}{dt} = E - \langle E(\beta)\rangle - \frac{\partial \ln w(\beta)}{\partial \beta} + \frac{\sqrt{2}}{\beta}\xi \tag{1}$$

where ξ is Gaussian white noise and $\langle \xi(t) \cdot \xi(t')\rangle = \delta(t - t')$, $\beta = 1/k_{\mathrm{B}}T$, with k_{B} being the Boltzmann constant and T the effective temperature. E is the instantaneous potential energy of the system in the current configuration. $\langle E(\beta)\rangle$ is the ensemble averaged potential energy at a specific temperature. This Langevin equation guarantees detailed balance of the simulated system. The converged system will eventually have a phase space distribution satisfying a generalized canonical ensemble,

$$p(\beta, X) \propto \frac{exp[-\beta E(X)] \cdot w(\beta)}{Z(\beta)} \tag{2}$$

where $w(\beta)$ is a free parameter. While any distribution $w(\beta)$ is valid for CST, the distribution $w(\beta) = 1/\beta$ has been suggested as optimal for molecular sampling [1], and it is used in our simulations.

2 Bootstrap sampling and temperature dependence of the free energy profile

The free energy profile was calculated from the first set of simulations, which used Continuous Simulated Tempering (CST) [1, 2]. The profile at 350 K is similar to that of 310 K (Figure S2A). In order to assess the level of equilibration in the sampling, the simulation data from umbrella-biased sampling was truncated in half and two new free energy profiles were recalculated, one with each half (Figure S2B). Although there is some discrepancy in the two free energy profiles, importantly, they both show the final state of the B-loop is unstable. In addition, we performed a bootstrap sampling based on the 10 evenly divided data of our simulation. The mean and standard deviation of our free-energy diagram were calculated from a total of 48 bootstrapped samples and are shown in (Figrure S2C). The plot indicates the original free energy curve is within the standard deviation of the bootstrap sampling.

Further, we track the change of the B-loop conformations during simulation by plotting the associated reaction coordinate (N_{hb}) as a function of time during our simulations (Figure S3). The figure shows in both the simulations with and without umbrella potential, the B-loop has accessed Basin C and Basin D for multiple

times. Thus, these simulations can be connected for a reliable free energy profile.

3 pKa calculation and free energy perturbation analysis

The pKa values of titratable residues in the post-fusion B-loop shares the same pattern as the pre-fusion state (Table S2), except for one residue, Glu67, which has a lower than 4.0 pKa, due to the formation of a salt-bridge interaction between it and Lys62. Glu67 is located in the middle of the B-loop, so its charge state could either stabilize Basin D or Basin G in our free energy plot (main text Figure 2). Because it is intractable to change the protonation state of this residue along with our thermodynamic simulation, we performed a free energy perturbation (FEP) calculation [4] to get a hint of whether a negatively-charge Glu67 would be stabilizing to basin G. Because FEP requires a very high degree of sampling, we used 1% of the change in potential energy in the FEP analysis for convergence. FEP (Figure S4) suggests that the addition of a negative charge to Glu67 will not change the free energy of Basin D as compared to Basin G, but instead would likely stabilize the pre-fusion state even further. Therefore, the protonation of this residue should not affect the main conclusion from our thermodynamic sampling. In sum, our protonation scheme, while below the range of pH probed by experiments, is set up in favor of stabilizing the post-fusion B-loop structure so that the B-loop transition here should be faster than at pH = 4.0. Our final free energy plot (main text Figure 2) demonstrates that even in this situation, the B-loop is unstable by itself in the post-fusion state.

4 Characterizing the "trapped" basin (Basin T)

The CST simulations reveal a trapped structure ensemble at Basin T (Figure 2 of main text). These structures have less formed hydrogen bonds relevant for the coiled-coil structure than the B-loop of the pre-fusion HA_2. The central structure clustered from this basin shows a β-sheet formation within one monomer of the B-loop (Figure S9). Note that the B-loop monomer shows significant tendency to form β-structure (Figure S8). In addition, the surface representation of the clustered trapped structure exhibits a close interaction between the B-loop and S3, with the B-loop packed in the grooves formed in between S3 helices.

5 Supplemental Figures S1-S10 and Table S1-S2

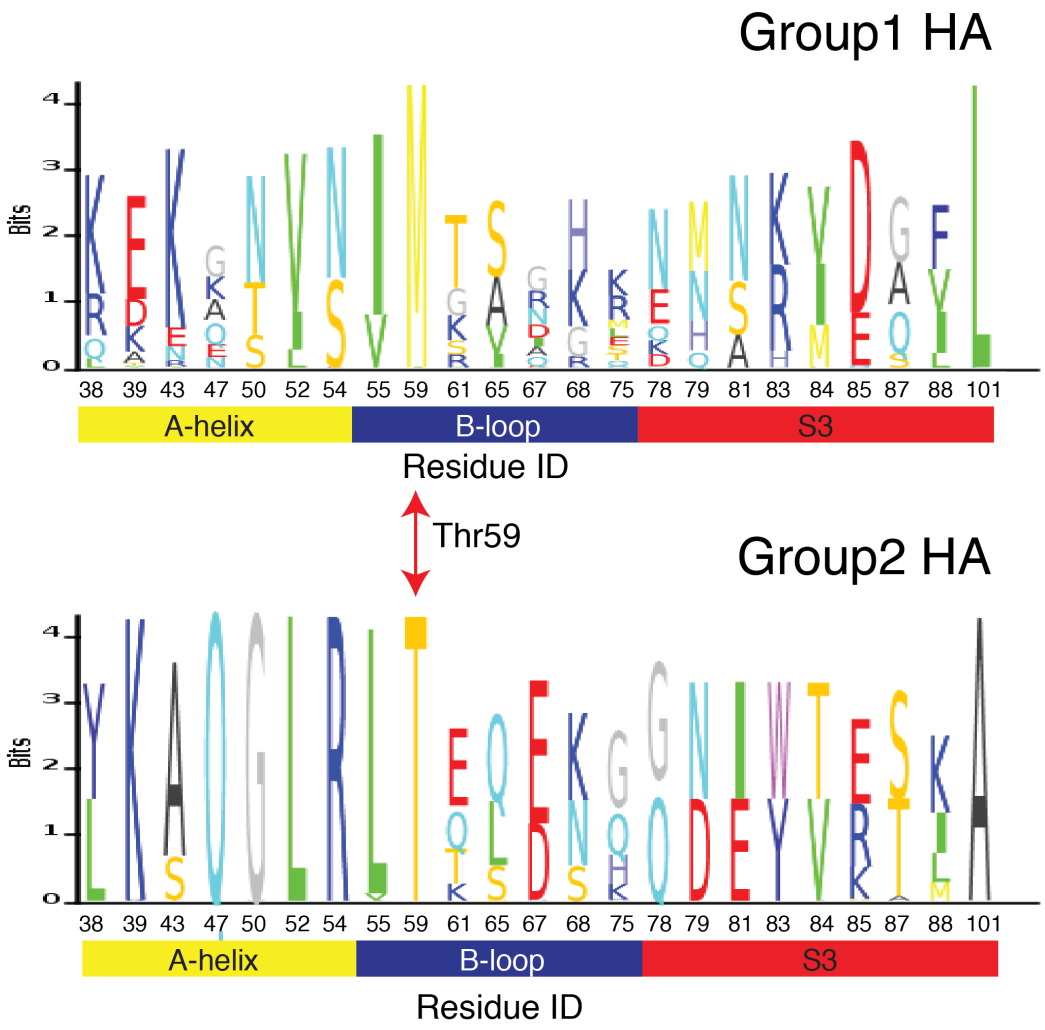

Figure S1: Sequence Logo between Group 1 and Group 2 HA. Only sites with 1-$H^2(P,Q) < 0.5$ were selected (described in main text). A comparison of the sequence logo of the B-loop (shown in the blue bar) between Group 1 and Group 2 HAs reveals a stark difference in position 59, where all Group 1 HAs share methionine while all Group 2 HAs share threonine. It is also the most conserved site of the B-loop region. This hints at a switch in the identity of the residue at this location when these two groups diverged from each other.

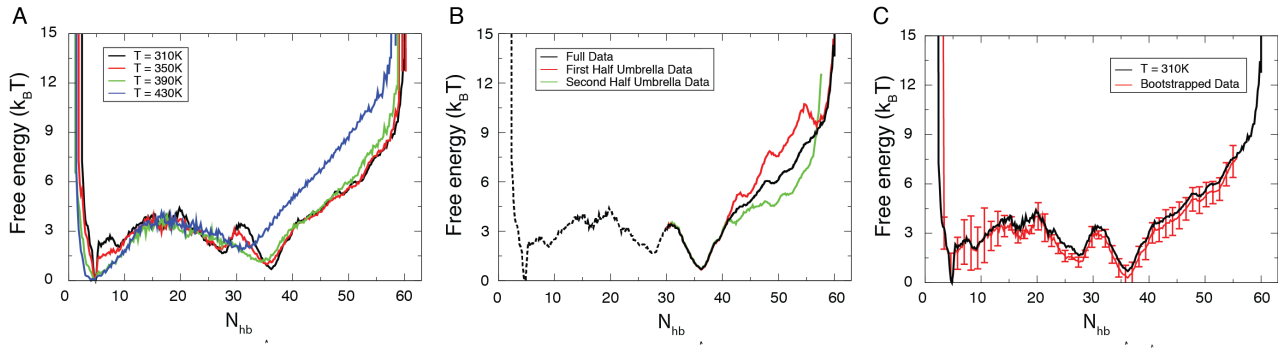

Figure S2: The free energy profile calculated from the CST simulations. (A) The data were binned into different temperatures of 310K, 350K, 390K and 430K. (B) We truncated the data in umbrella sampling in two and recalculated the free energy profile at T = 310K based on the truncated data. The red curve stands for the profile with the first half of the umbrella simulations and the green curve stands for the profile with the second half of the umbrella simulations. The black curve stands for the profile by using all the data in the umbrella simulations. The free energy surface accessed from simulations with umbrella bias are plotted in solid lines, while that from unbiased simulation is plotted in dashed lines. (C) The free energy diagram calculated from the bootstrap sampling based on our existing simulation data. Black: The free energy calculated from the raw data. Red: The mean and standard deviation calculated from 48 bootstrapped data sets.

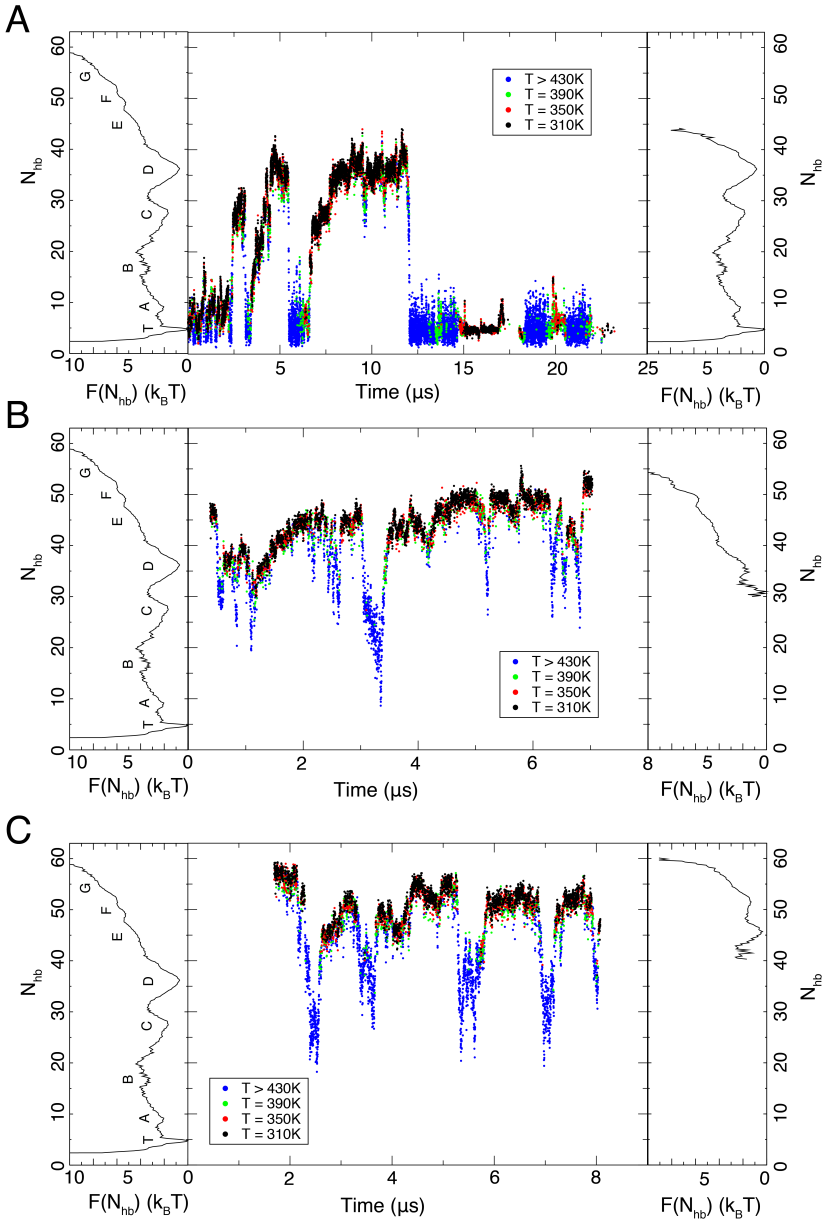

Figure S3: N_{hb} as a function of time in our simulations. In both the cases of unbiased simulation (A) and simulations biased by umbrella potential referenced at 55.0 (B) and at 60.0 (C), the B-loop was able to access Basin C and Basin D for multiple times. In the left hand side of each panel, we show the complete free energy diagram at T = 310K after thermodynamically reweighting all the data from simulations. In the right hand side, we show the free energy diagram at T = 310K after thermodynamically reweighting the data from each panel separately.

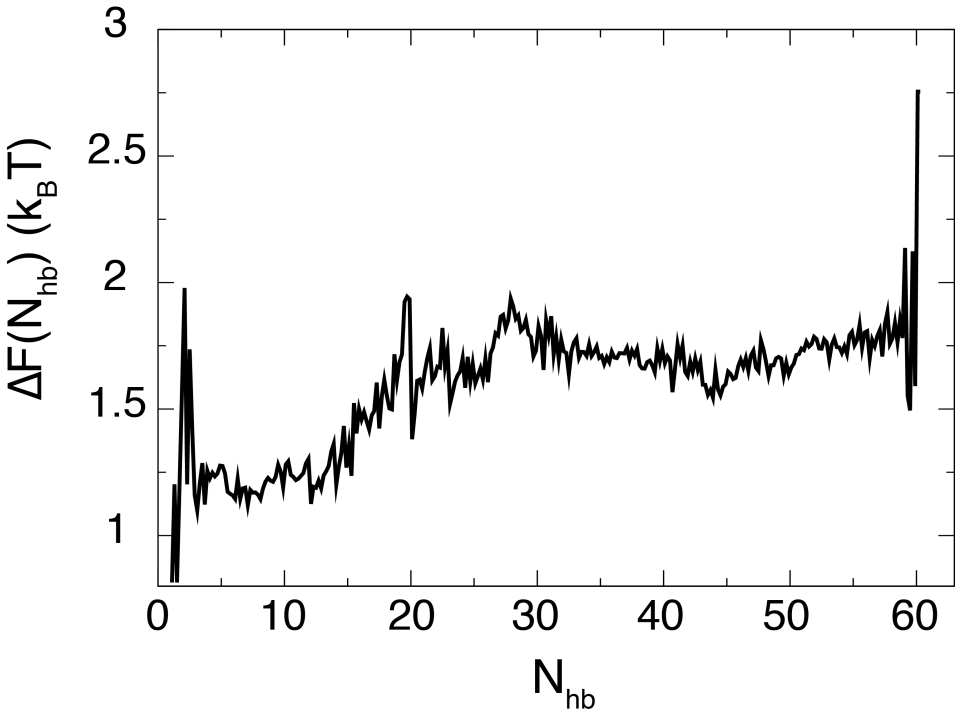

Figure S4: The calculation of the change of free energy by recovering the negative charge of three Glu67s in our thermodynamic simulations. For numerical stability, we only use 1% of the change of potential energy due to the deprotonation of Glu67. The calculation shows a further stabilization in the pre-fusion B-loop relative to Basin G, while the free energy difference between Basin D and Basin G is negligible.

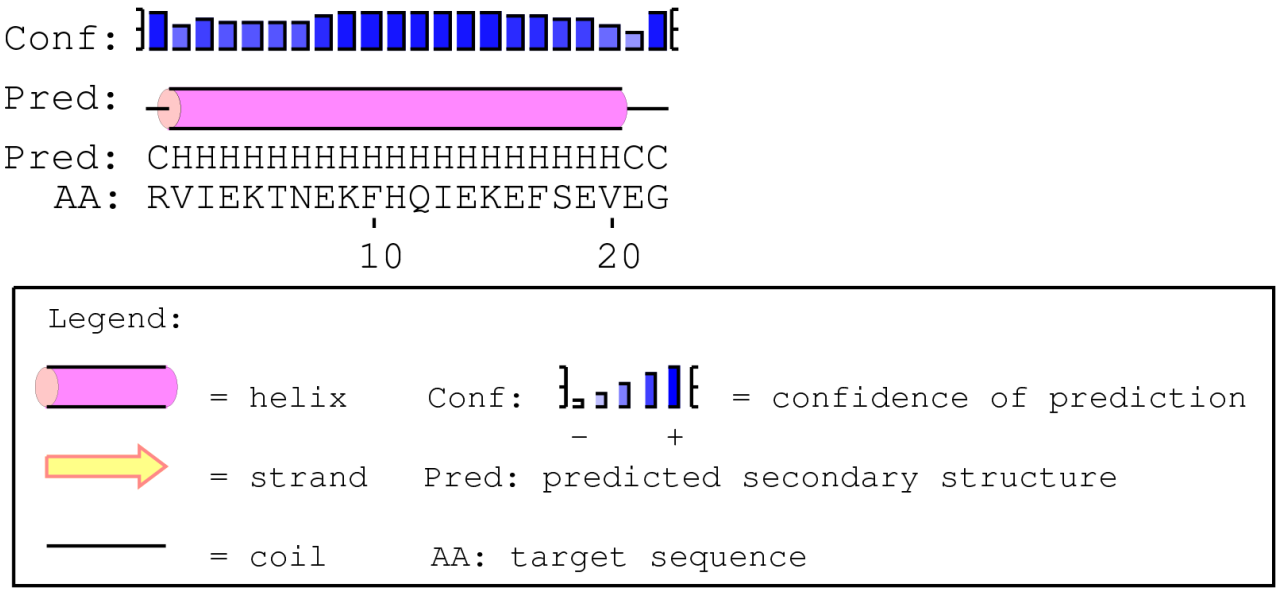

Figure S5: The secondary structure prediction of the B-loop. Psipred [5, 6], a secondary structure prediction tool based on bioinformatic study, predicts a mostly helical B-loop.

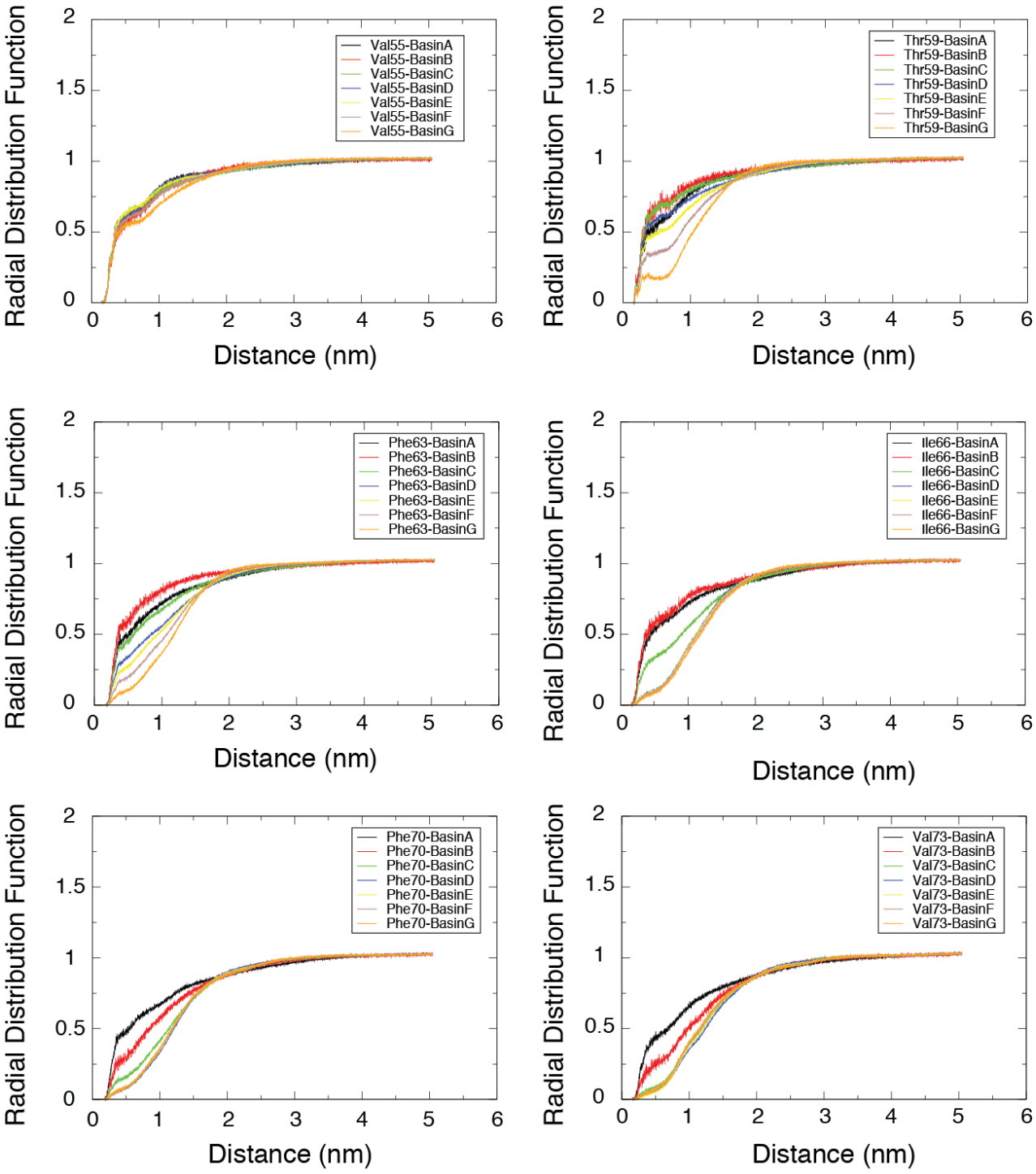

Figure S6: The calculated RDF functions of water around the side chain atoms of each a/d residues of the B-loop to assess the dehydration of water molecules in the core residues of the B-loop. The RDF represents the average density of solvent particles within a certain distance. Excerpts of this data are shown in Main Text Figure 3.

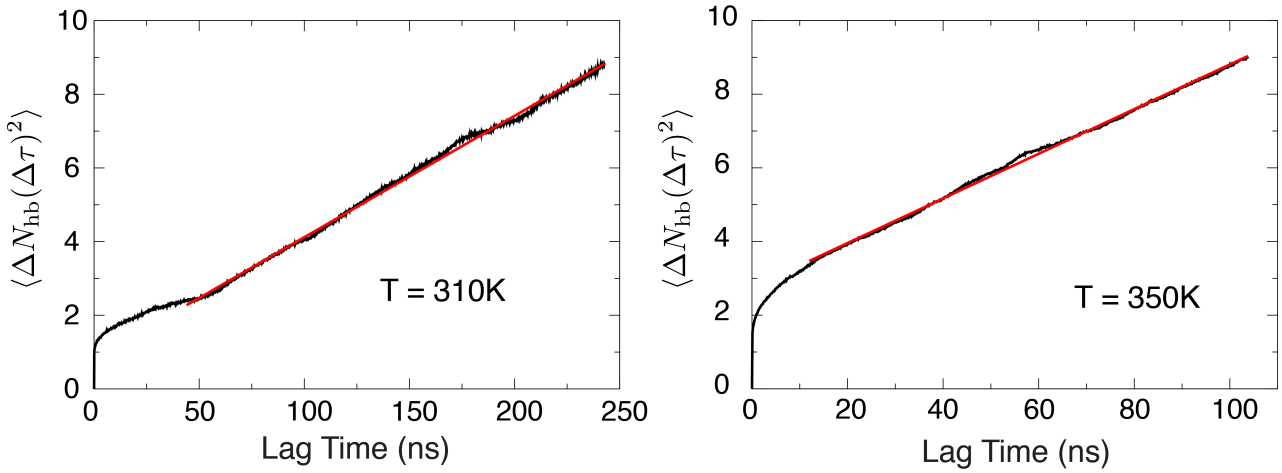

Figure S7: The mean square displacement of N_{hb} as a function of lag time ($\langle \Delta N_{\rm hb}(\Delta\tau)^2 \rangle$) calculated from many independent 300 ns simulations in different basins. As shown in Table S1, there is a large variation in the calculated diffusion coefficient from each basin. To reduce the error, we approximate a constant diffusion coefficient for the whole profile by averaging $\Delta N_{hb}(\Delta\tau)$ at each $\Delta\tau$ over all the 300 ns simulations. The final plot of two simulation temperatures (310K and 350K) for estimating diffusion coefficient is shown here.

Temperature (K)	Intermediates	Clusters	Diffusion Coefficient (/μs)
310 / 350	A	1	0.372 / 1.81
310 / 350	A	2	25.9 / 10.6
310 / 350	A	3	11.9 / 8.13
310 / 350	B	1	5.65 / 60.1
310 / 350	B	2	1.59 / 83.7
310 / 350	B	3	1.02 / 43.1
310 / 350	C	1	6.69 / 10.1
310 / 350	C	2	61.5 / 125
310 / 350	C	3	40.3 / 71.3
310 / 350	D	1	4.22 / 24.8
310 / 350	D	2	5.22 / 77.1
310 / 350	D	3	186 / 25.7
310 / 350	E	1	0.434 / 28.3
310 / 350	E	2	NA /
310 / 350	E	3	NA /
310 / 350	F	1	6.94 / 0.983
310 / 350	F	2	NA /
310 / 350	F	3	NA /
Averaged Value			
310 / 350	NA	NA	**16.5 / 30.3**

Table S1: Summary of the diffusion coefficients calculated from 300 ns constant temperature simulations

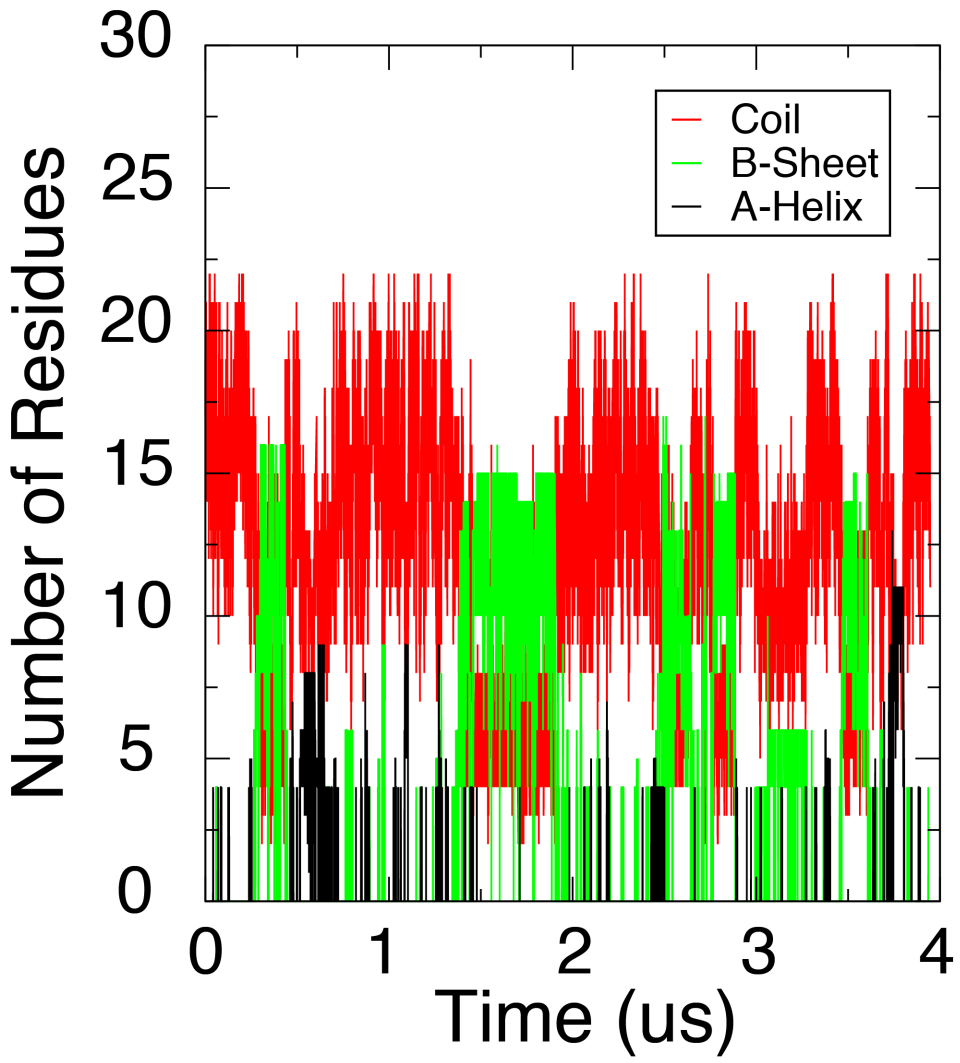

Figure S8: The secondary structure of a single B-loop monomer during 4 μs of simulation. CST enhanced sampling was used, and plot presents data combined from 300 K to 400 K. The B-loop monomer shows β-sheet structure in about 20% of the simulation.

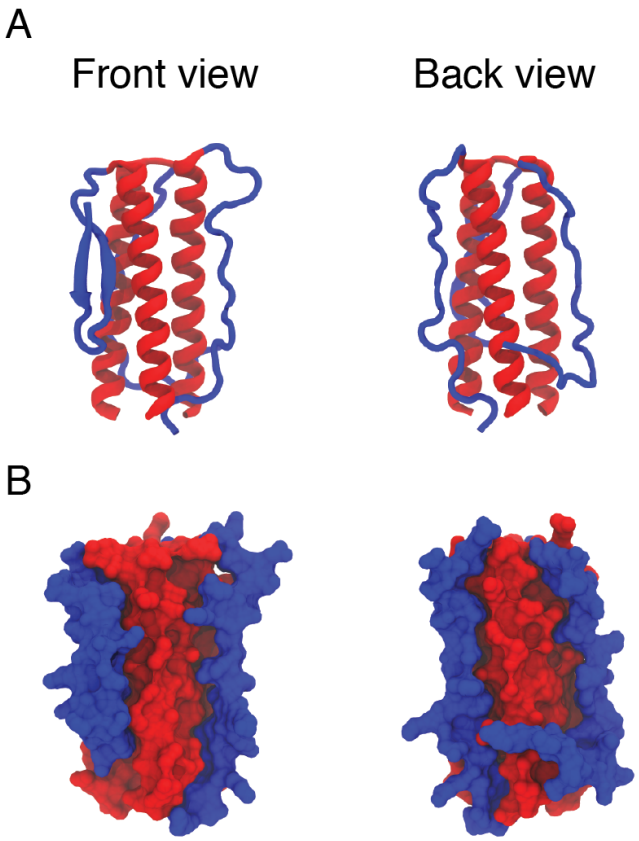

Figure S9: The front and back views of the structure clustered from the Basin T in the free energy diagram of Set 1 simulations (main text Table 1 and Figure 2). (A) Cartoon representation. (B) Surface representation.

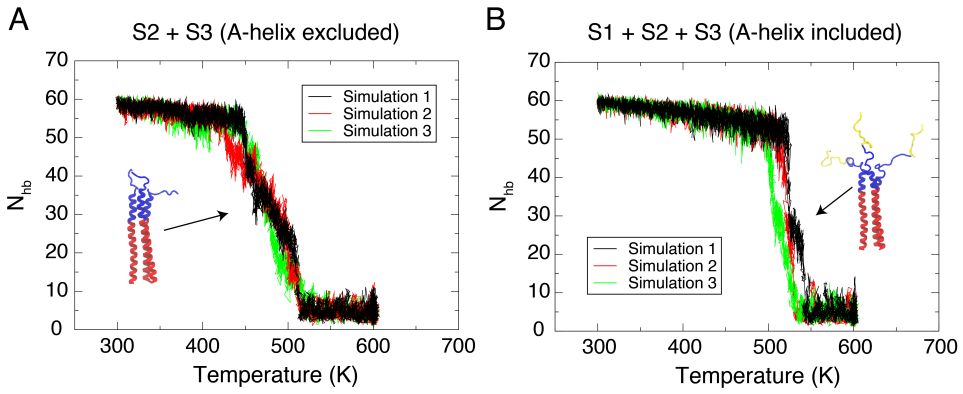

Figure S10: The individual melting curves for the S2+S3 and the S1+S2+S3 systems. Three identical simulation setups with different random seeds were used for each. N_{hb} of the B-loop is used to show the portion of the B-loop structure remaining as coiled-coil. In each of the three simulations, intermediate states resembling the structural ensemble of Basin D in Figure 2 were observed. (A) The simulations for the S2+S3 system. (B) The simulations for the S1+S2+S3 system.

Residues	Pre-fusion A[a]	Pre-fusion B	Pre-fusion C	Post-fusion A	Post-fusion B	Post-fusion C
Glu_{57}	4.40	4.40	4.38	4.63	4.65	4.65
Glu_{61}	4.70	4.70	4.70	4.52	4.55	4.54
His_{64}	6.63	6.64	6.64	6.56	7.39	6.54
Glu_{67}	4.67	4.66	4.66	3.65	2.77	3.60
Glu_{69}	4.59	4.59	4.59	4.46	4.41	4.61
Glu_{72}	4.15	4.16	4.16	4.64	4.62	4.47
Glu_{74}	2.43	2.45	2.37	7.09	7.11	7.34
Asp_{79}	4.04	4.06	4.11	3.74	3.80	3.74
Glu_{81}	4.43	4.39	4.40	5.19	6.22	5.20
Glu_{85}	3.33	3.29	3.30	3.92	3.78	3.92
Asp_{86}	3.78	3.77	3.80	4.24	4.24	4.21
Asp_{90}	2.99	2.97	3.08	4.03	4.02	4.00
Glu_{97}	3.58	3.57	3.58	4.67	4.16	4.22
Glu_{103}	4.73	4.73	4.73	4.60	4.61	4.61

[a] A means the Chain A

Table S2: The pKa values calculated from PROPKA for the titratable residues in the pre-fusion and post-fusion B-loop simulation system.

References

[1] Zhang, C & Ma, J. (2010) Enhanced sampling and applications in protein folding in explicit solvent. *The Journal of Chemical Physics* **132**, 244101.

[2] Zhang, C & Ma, J. (2012) Folding helical proteins in explicit solvent using dihedral-biased tempering. *Proceedings of the National Academy of Sciences* **109**, 8139–8144.

[3] Li, H, Robertson, A. D, & Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics* **61**, 704–721.

[4] Zwanzig, R. W. (1954) HighTemperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **22**, 1420–1426.

[5] Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* **292**, 195–202.

[6] Buchan, D. W. A, Minneci, F, Nugent, T. C. O, Bryson, K, & Jones, D. T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research* **41**, W349–W357.

[7] Kabsch, W & Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.

[8] Lin, X, Eddy, N. R, Noel, J. K, Whitford, P. C, Wang, Q, Ma, J, & Onuchic, J. N. (2014) Order and disorder control the functional rearrangement of influenza hemagglutinin. *Proceedings of the National Academy of Sciences* **111**, 12049–12054.

[9] Lin, X, Noel, J. K, Wang, Q, Ma, J, & Onuchic, J. N. (2016) Lowered pH Leads to Fusion Peptide Release and a Highly Dynamic Intermediate of Influenza Hemagglutinin. *The Journal of Physical Chemistry B* **120**, 9654–9660.