SUPPORT INFORMATION

Free Energy Guided Sampling

Ting Zhou and Amedeo Caflisch*

Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland

E-mail: caflisch@bioc.uzh.ch

Phone: (+41 44) 635 55 21. Fax: (+41 44) 635 68 62

References

(1) Krivov, S. V.; Karplus, M. *J Phys Chem B* **2006**, *110*, 12689–12698.

(2) Seeber, M.; Felline, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F. *J Comput Chem* **2011**, *32*, 1183–1194.

(3) Krivov, S. V.; Muff, S.; Caflisch, A.; Karplus, M. *J Phys Chem B* **2008**, *112*, 8701–8714.

(4) Andersen, C. A. F.; Palmer, A. G.; Brunak, S.; Rost, B. *Structure* **2002**, *10*, 175–184.

*To whom correspondence should be addressed

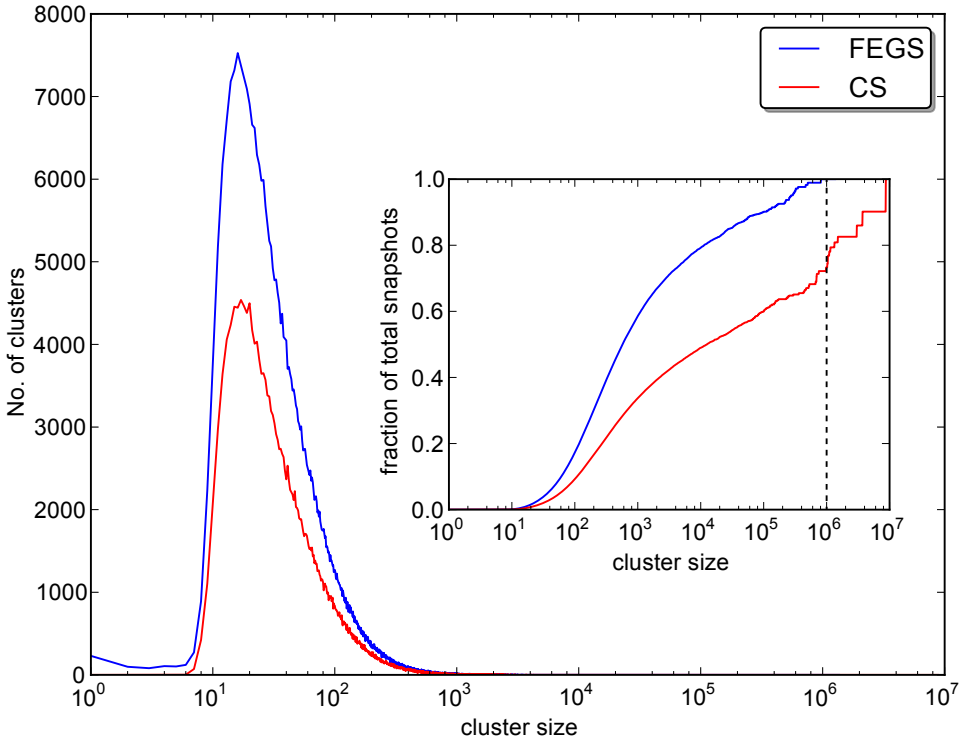


Figure S1: Comparison of distribution of the cluster size of FEGS and CS on Beta3S. The most frequent size of clusters in both sampling methods are similar around 20. However, the CS tends to generate clusters that contain a large amount of snapshots. In the CS, over 20% of snapshots belong to clusters that contain more than 10^6 snapshots (see the inset), whereas there are not any large cluster in the FEGS. This indicates that FEGS escapes enthalpy traps more easily than CS.

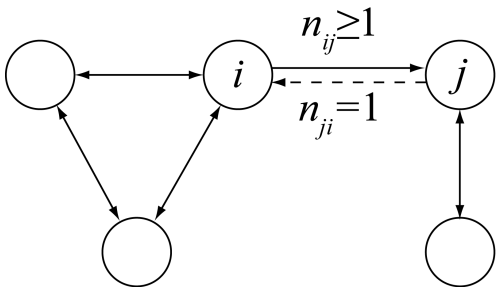


Figure S2: Schematic illustration of manually added backward transition for generating an ergodic transition network. The cycles are the mesostates in transition network. The transition from j to i (dashed arrow) is added to make the whole transition network strongly connected.

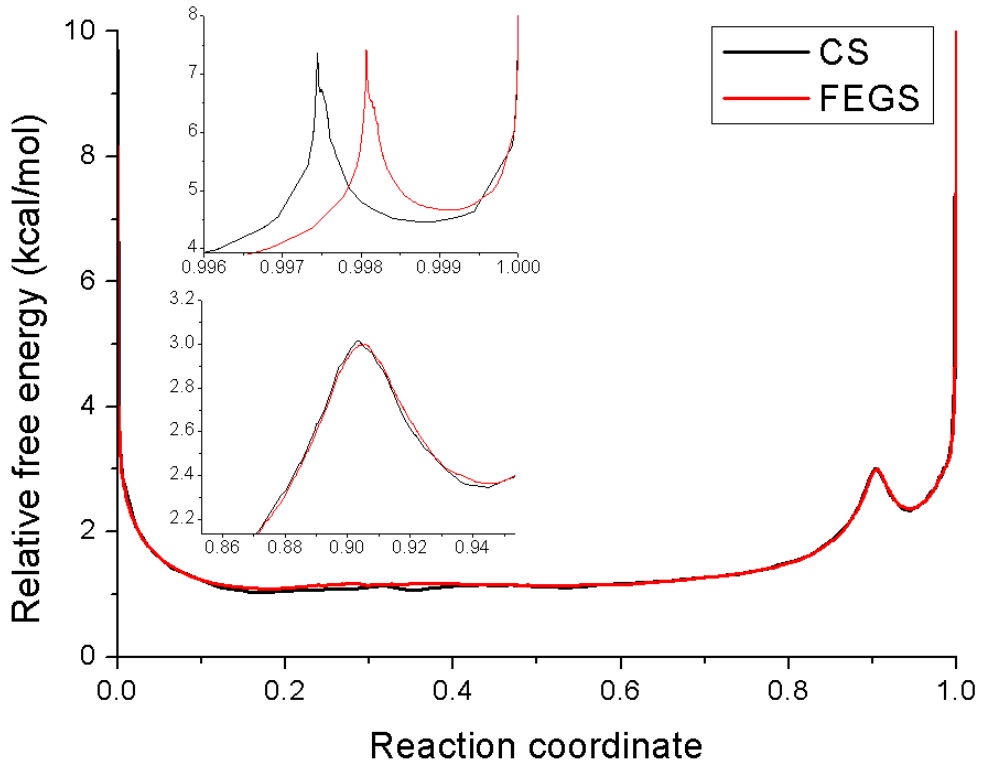


Figure S3: cFEPs of the alanine dipeptide calculated from the transition networks sampled by CS and FEGS. The two insets are detailed zoom-in cFEP at two barriers on the original cFEP. There were only marginal differences (0.04 kcal/mol in barrier height and 0.0006 relative partition function in the barrier position) in cFEPs between two sampling methods.

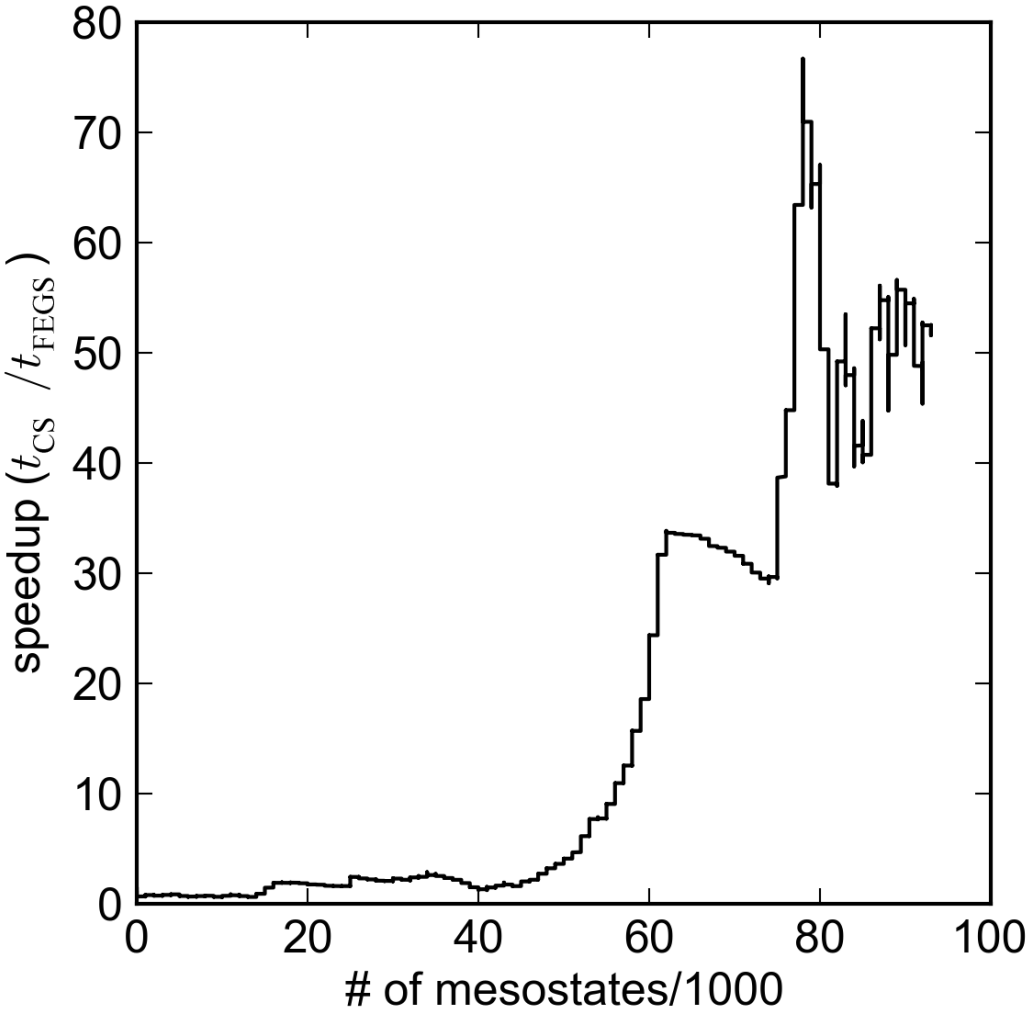


Figure S4: The ratio of total simulation time needed by CS and FEGS to visit a certain number of clusters of the alanine dipeptide. The conformations are binned based on the two dihedral angles ϕ, ψ. The bin size on both dihedral angles (axes) is 1°, so that the total number of bins is 360^2.

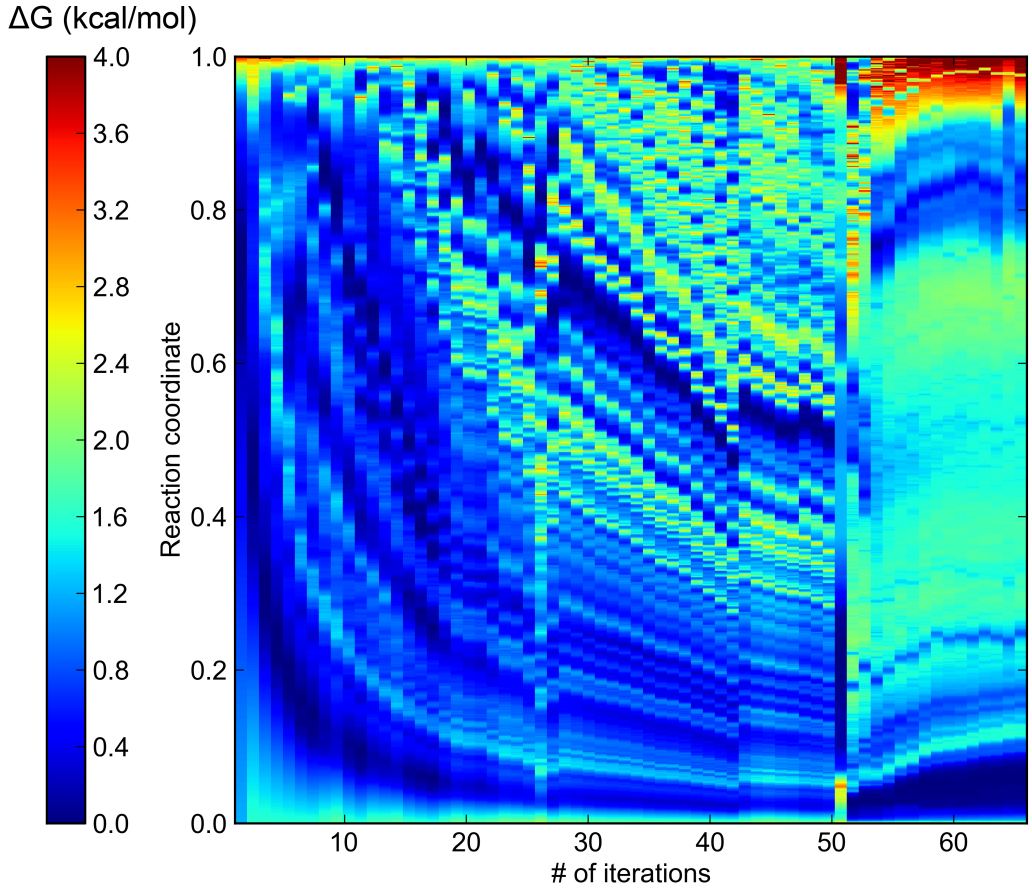


Figure S5: Evolution of the cut-based free energy profile of Beta3s. The vertical-axis is the natural reaction coordinate sorted by mfpt in cFEPs.[1] The low to high value of free energy are depicted with the spectrum from blue to red. The horizontal axis is the number of iterations. In the first 50 iterations the sampling was in the exploring stage. After the first 50 iterations the free energy converged rapidly in the refining stage.

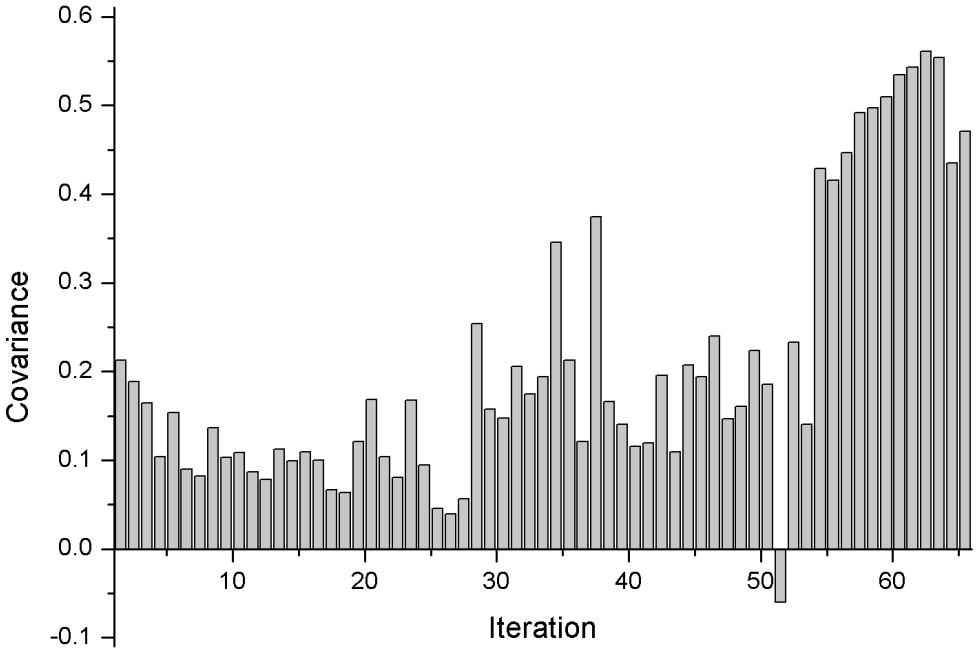


Figure S6: The covariances between two adjacent cFEPs of Beta3s. The covariance was calculated by $\mathrm{Cov}(\mathrm{cFEP}_n^{\mathrm{interp}}, \mathrm{cFEP}_{n+1}^{\mathrm{interp}}) = \mathrm{E}(\mathrm{cFEP}_n^{\mathrm{interp}} \cdot \mathrm{cFEP}_{n+1}^{\mathrm{interp}}) - \mathrm{E}(\mathrm{cFEP}_n^{\mathrm{interp}}) \cdot \mathrm{E}(\mathrm{cFEP}_{n+1}^{\mathrm{interp}})$ where the $\mathrm{cFEP}_n^{\mathrm{interp}}$ is the linearly interpolated cFEP of nth iteration used for checking the convergence, $\mathrm{E}(\cdot)$ is the operator of the expected value. The spacing between adjacent interpolated points is 0.001 of relative partition function.

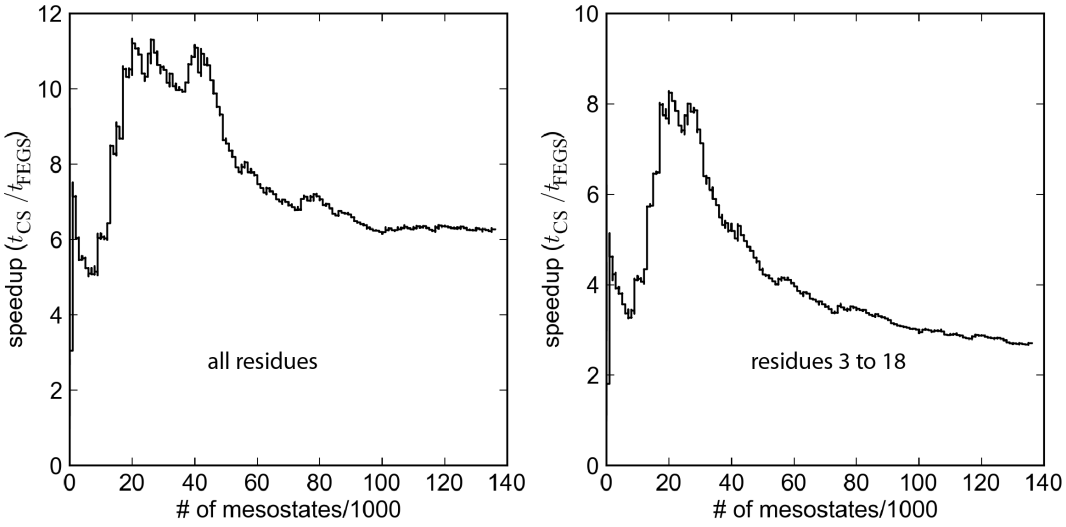
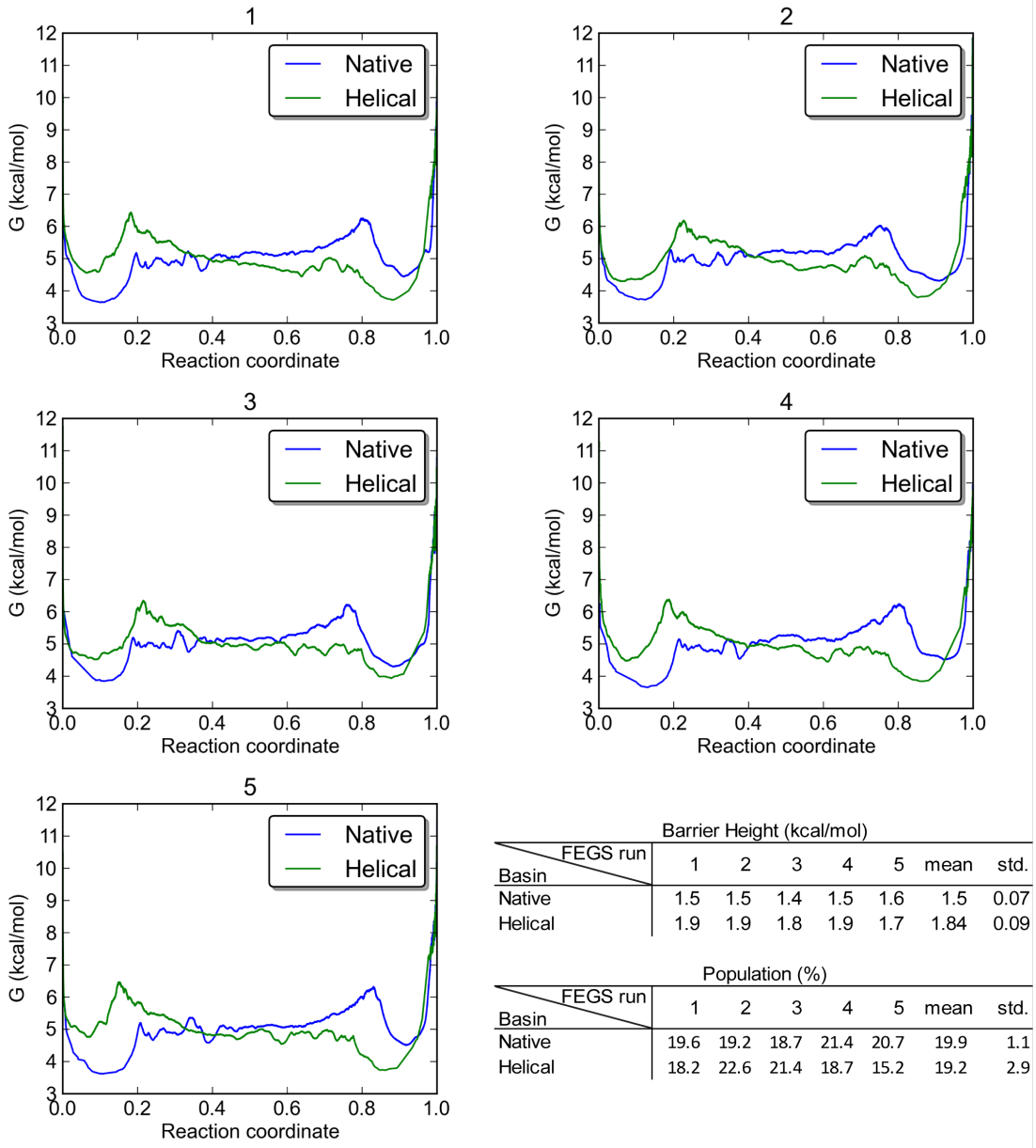


Figure S7: The ratio of total simulation time needed by CS and FEGS to visit a certain number of clusters of the 20-residue peptide Beta3s. Importantly the actual speedup is larger than the ratio of total simulation time because FEGS made use of 100 parallel runs while CS is usually carried out by up to 10 runs. The sequential leader-like clustering algorithm was used with a threshold of 2.5 Å on the pairwise coordinate root mean square deviation of all unsymmetrical heavy atoms (left panel) and the unsymmetrical heavy atoms of residues 3 to 18 (right panel). The measured speedup of FEGS depends on the granularity with which the conformational space is decomposed in clustering process. The speedup of FEGS is more significant if a finer granularity is used. Note that despite the same threshold of 2.5 Å the all-residue RMSD has a finer granularity than the RMSD of residues 3 to 18 because of the fluctuations at the termini (residues 1, 2, 19, and 20), and because of the correlation between number of atoms and value of RMSD.



Barrier Height (kcal/mol)

Basin \ FEGS run	1	2	3	4	5	mean	std.
Native	1.5	1.5	1.4	1.5	1.6	1.5	0.07
Helical	1.9	1.9	1.8	1.9	1.7	1.84	0.09

Population (%)

Basin \ FEGS run	1	2	3	4	5	mean	std.
Native	19.6	19.2	18.7	21.4	20.7	19.9	1.1
Helical	18.2	22.6	21.4	18.7	15.2	19.2	2.9

Figure S8: cFEPs of Beta3s in the reproducibility test which consisted of five independent FEGS runs of 24.5 μs each started from the native structure. Clustering was performed by the sequential leader-like clustering algorithm[2] with a threshold of 2.5 Å on the pairwise coordinate root mean square deviation of the unsymmetrical heavy atoms of residues 3 to 18.[3] The reference structure for the cFEPs (i.e., structure with reaction coordinate = 0) is the native (blue) or non-native helical (green). The barrier height and population of basins are shown in the table at the bottom-right.

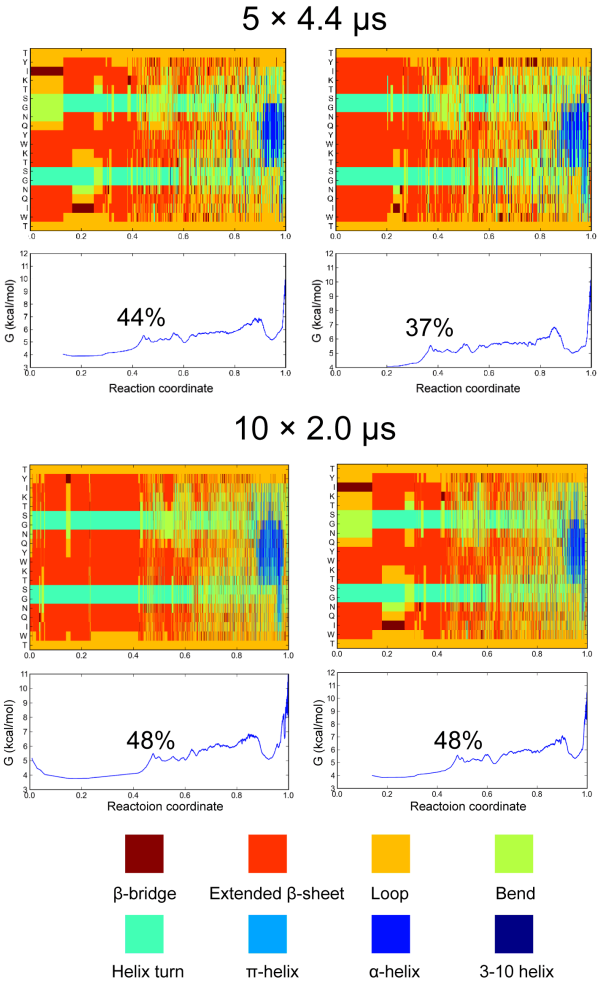


Figure S9: Comparison of cFEPs generated by CS of Beta3s with different trajectory length. Each of the two cFEPs in the top was generated from five 4.4-μs trajectories. The cFEPs in the bottom were generated each from ten 2.0-μs trajectories. The population of the native basin (denoted in percentage on each cFEP) derived from short trajectories (i.e., 2.0 μs) is larger than the one derived from long trajectories (i.e., 4.4 μs), which shows the dependence of CS on the starting structure. The upper part of each panel (with the sequence of the Beta3s on the y-axis) shows the colored DSSP[4] strings of the cluster representatives, which are arranged according to the reaction coordinate of the cFEP. The legend of colors for different secondary structure elements in the traces are indicated at the bottom panel.