SUPPLEMENTARY INFORMATION

Efficient sampling of high-dimensional free-energy landscapes with Parallel Bias Metadynamics

Additional details of the simulations

Model system. To verify the absence of systematic errors in the free energy reconstructed by Parallel Bias Metadynamics (PBMetaD) and benchmark its efficiency, we used a model system whose free energy is a function of two dimensionless collective variables S_1 and S_2:

$$F(S_1, S_2) = -k_B T \cdot \log\left\{\sum_{i=1}^{4} e^{-\frac{1}{2}k^i\left[\left(S_1 - S_1^i\right)^2 + \left(S_2 - S_2^i\right)^2\right]}\right\}. \quad \text{(S1)}$$

The free energy defined by Eq. S1 is characterized by the presence of four local minima separated by high free-energy barriers (Fig. 1A). The parameters S_1^i, S_2^i, and k^i are defined in Tab. T1. The estimates of the mono-dimensional free energies $\tilde{F}(S_1)$ and $\tilde{F}(S_2)$ obtained by PBMetaD or Bias Exchange Metadynamics (BEM) were compared to the exact analytical functions:

$$F(S_1) = -k_B T \cdot \log \int dS_2\, e^{-\frac{F(S_1,S_2)}{k_B T}} = -k_B T \cdot \log\left\{\sum_{i=1}^{4} e^{-\frac{1}{2}k^i\left(S_1 - S_1^i\right)^2} \cdot \sqrt{\frac{2\pi}{k^i}}\right\}$$

$$F(S_2) = -k_B T \cdot \log \int dS_1\, e^{-\frac{F(S_1,S_2)}{k_B T}} = -k_B T \cdot \log\left\{\sum_{i=1}^{4} e^{-\frac{1}{2}k^i\left(S_2 - S_2^i\right)^2} \cdot \sqrt{\frac{2\pi}{k^i}}\right\}, \quad \text{(S2)}$$

using the Root Mean Square Deviation (RMSD) metrics:

$$RMSD\left(\tilde{F}, F\right) = \sqrt{\frac{1}{\Omega}\int_{\Omega} dS\left[\left(\tilde{F}(S) - \langle\tilde{F}(S)\rangle\right) - \left(F(S) - \langle F(S)\rangle\right)\right]^2}, \quad \text{(S3)}$$

where S is either S_1 or S_2, $\langle F(S)\rangle$ and $\langle\tilde{F}(S)\rangle$ are the exact and estimated free energies averaged over the region Ω. Since sampling in PBMetaD and BEM is limited to relevant regions of the CVs space by the temperature parameter ΔT of well-tempered metadynamics (WTMetaD), Ω was defined as the region of the CVs space within 20 $k_B T$ of the global minimum in Eq. S2.

To calculate the full two-dimensional free energy from a PBMetaD simulation, we used a simple rewighting procedure. Since the PBMetaD bias potential $V_{PB}(S_1, S_2, t)$ becomes quasi-static in the long-time limit, we discarded the initial transient (20% of the entire run) and considered the bias potential as static for the remaining of the simulation. We then used the standard Torrie-Valleau umbrella sampling reweighting[1] to recover the unbiased probability distribution by assigning the following weight to each conformation:

$$w(S_1, S_2) \propto e^{+\frac{V_{PB}(S_1, S_2, \bar{t})}{k_B T}}, \quad \text{(S4)}$$

where $V_{PB}(S_1, S_2, \bar{t})$ is the PBMetaD bias potential at the end of the simulation ($\bar{t} = 5 \cdot 10^6$ MC steps). The two-dimensional free energy obtained using the weights in Eq. S4 was in excellent agreement with the reference landscape (Fig. S4). A similar approach has been used before with standard WTMetaD simulations[2].

Tryptophan-cage miniprotein.

The standard NMR structure (PDB entry 1L2Y)[3] was used for the all of the simulations. The GBSA implicit solvation model[4-5] implemented in GROMACS 5.05 was ported to GROMACS 4.5.7[6] and used with the AMBER99SB-ILDN force field[7]. The isolated protein was simulated without periodic boundary conditions and a 2.0 nm cutoff for Lennard-Jones and electrostatic interactions. All enhanced sampling calculations were implemented using standard CVs within the PLUMED plugin (version 2.2.0)[8]. The CVs used (as defined by their PLUMED keyword) and any relevant paramters are in the Table T2 below.

References

1. Torrie, G. M.; Valleau, J. P. *J Comput Phys* **1977,** *23*, 187-199.

2. Branduardi, D.; Bussi, G.; Parrinello, M. *J Chem Theory Comput* **2012,** *8*, 2247-2254.

3. Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat Struct Biol* **2002,** *9*, 425-30.

4. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry* **1996,** *100*, 19824-19839.

5. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *The Journal of Physical Chemistry A* **1997,** *101*, 3005-3014.

6. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J Chem Theory Comput* **2008,** *4*, 435-447.

7. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010,** *78*, 1950-8.

8. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput Phys Commun* **2014,** *185*, 604-613.

List of Supplementary Tables

Tab. T1. Parameters of the potential used in the two-dimensional model system.

# minimum	S_1^i	S_2^i	k^i
1	1.0	5.0	0.1
2	20.0	7.0	0.2
3	5.0	30.0	0.2
4	30.0	35.0	0.05

Tab. T2. Parameters used in PLUMED for monitoring CVs and biasing simulations

CV Description	CV Name	Parameters	Notes
Hydrophobic Contacts	COORDINATION	N=8,M=12,R_0=0.5 nm	C-gamma atoms of residues LEU2,TRP6,LEU7,PRO12,PRO17,PRO18,PRO283
Dihedral Correlation	DIHCOR	Residues 2--19	All psi / phi residues
Hydrogen Bonds	COORDINATION	N=8,M=12,R_0=0.25 nm	All backbone O and H atoms. Neighbor list used with stride of 5 steps
Alpha Helicity	ALPHABETA	Φ^0 = -1.0 rad , Ψ^0= -0.82 rad	
Beta Similarity	ALPHABETA	Φ^0 = -1.396 rad , Ψ^0= 2.618	
Radius of Gyration	GYRATION		All alpha carbons

List of Supplementary Figures

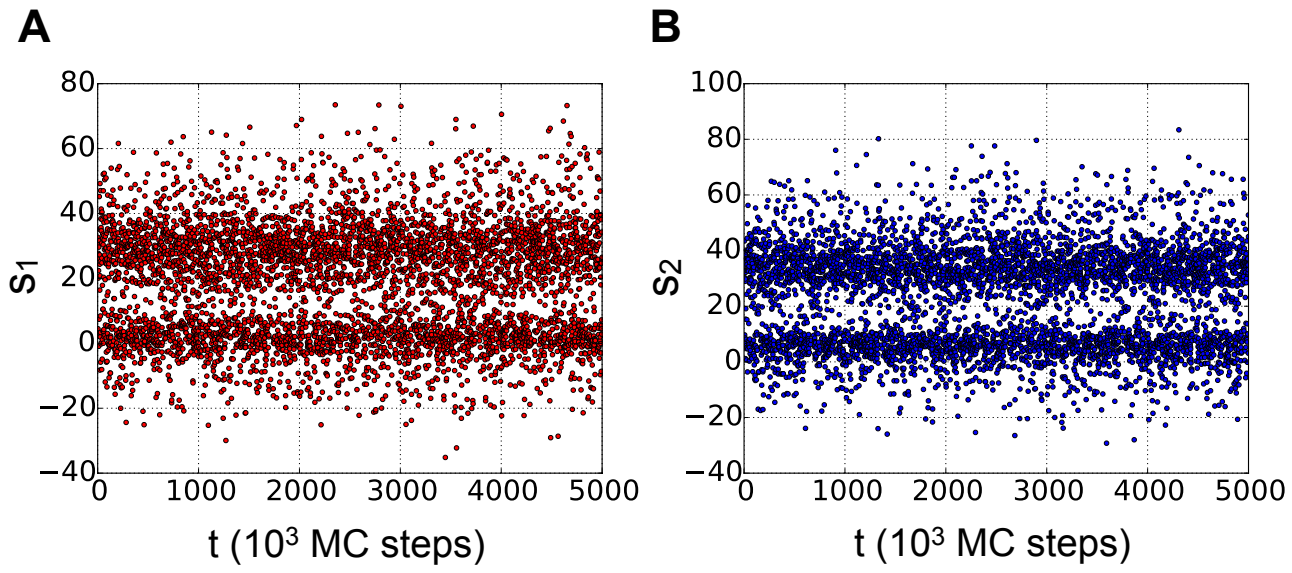

Fig. S1. Time series of S_1 (A) and S_2 (B) in a typical model system Monte Carlo (MC) simulation. Due to the effect of the PBMetaD bias potential, the system explores all the four local minima in the free-energy landscape and rapidly achieves a diffusive behavior in the entire CVs space.

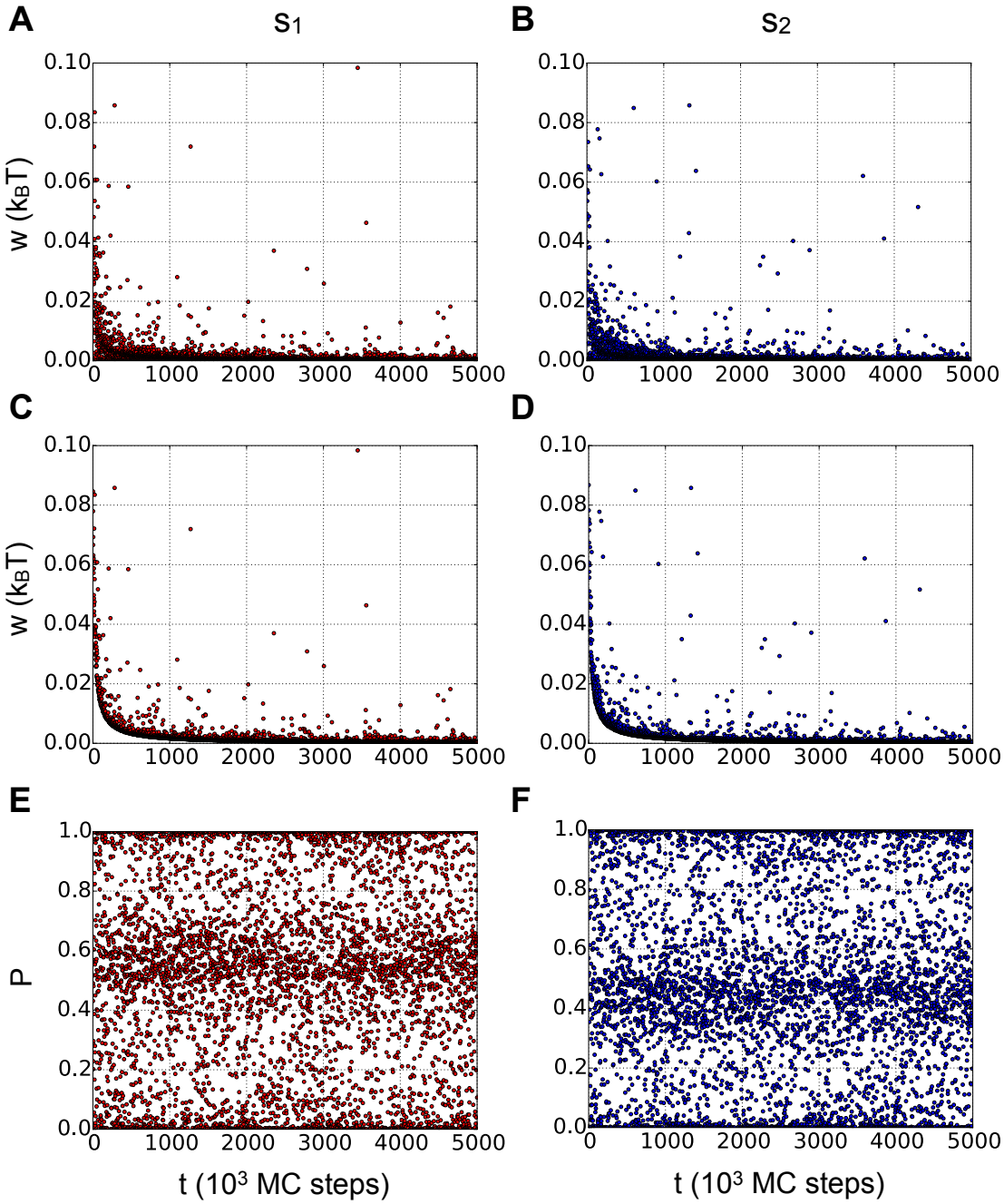

Fig. S2. Time series of the Gaussian heights for S_1 (A) and S_2 (B) in a typical model system MC simulation, decoupling the standard reduction of WTMetaD (C, D) from the additional conditional weight of PBMetaD (E, F).

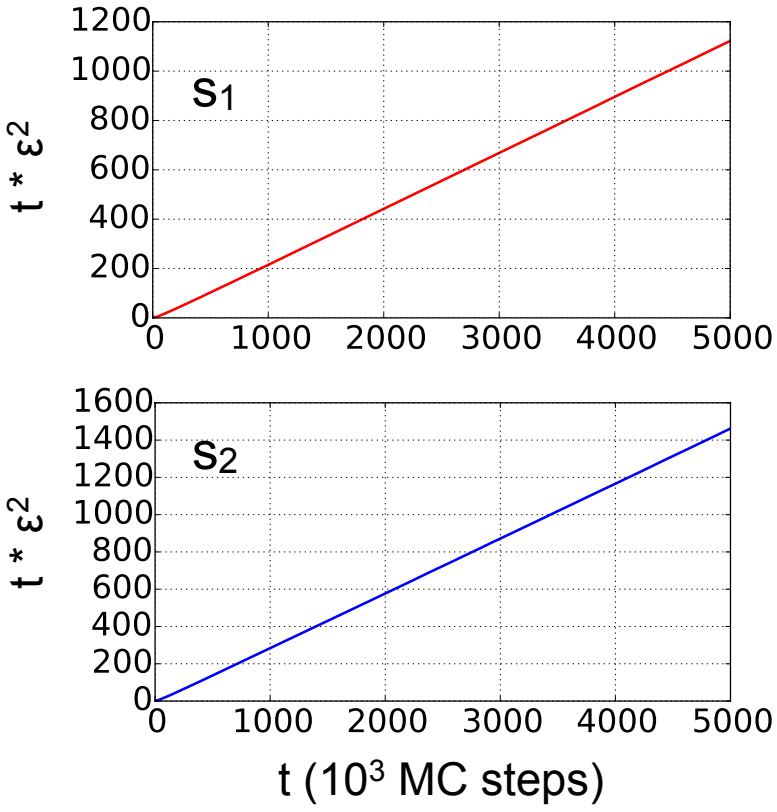

Fig. S3. Systematic errors in the free-energy reconstruction of a model system. Deviations of the PBMetaD free energies from the analytical solutions as a function of time, when the Gaussian heights are rescaled using only the standard recipe of WTMetaD (Eq. 3) instead of Eq. 12, which contains the additional conditional weights.

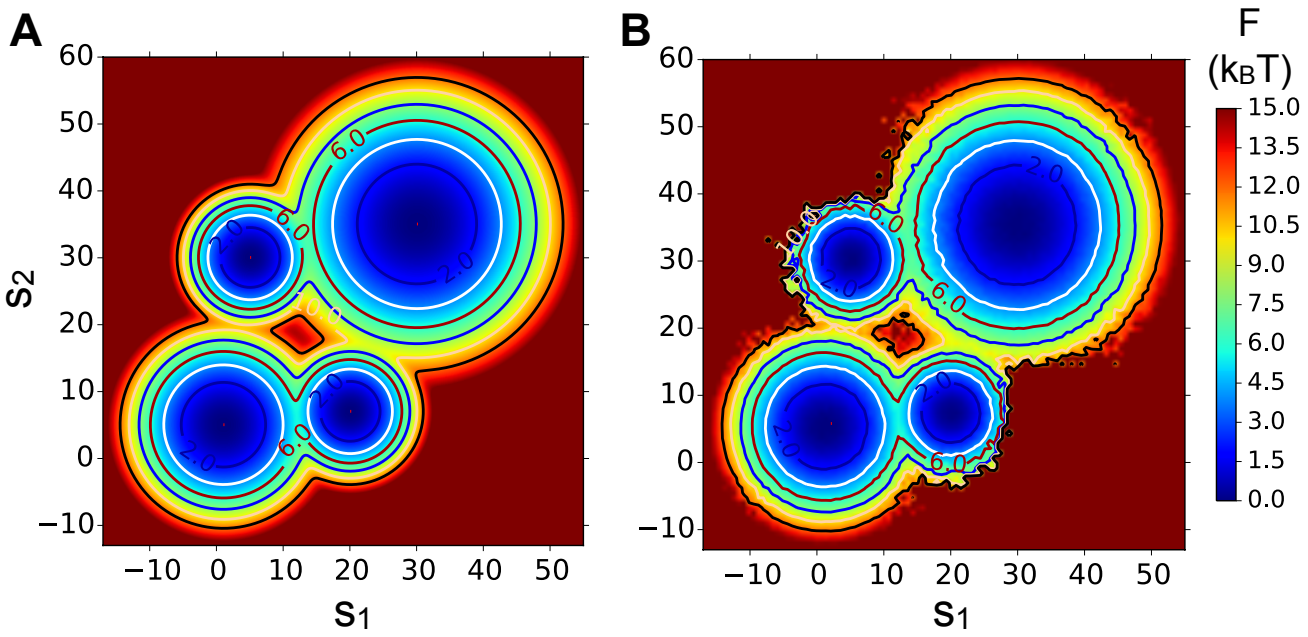

Fig. S4. Recovering the two-dimensional free energy (A) from the PBMetaD simulation of a model system by simple Torrie-Valleau reweighting (B).

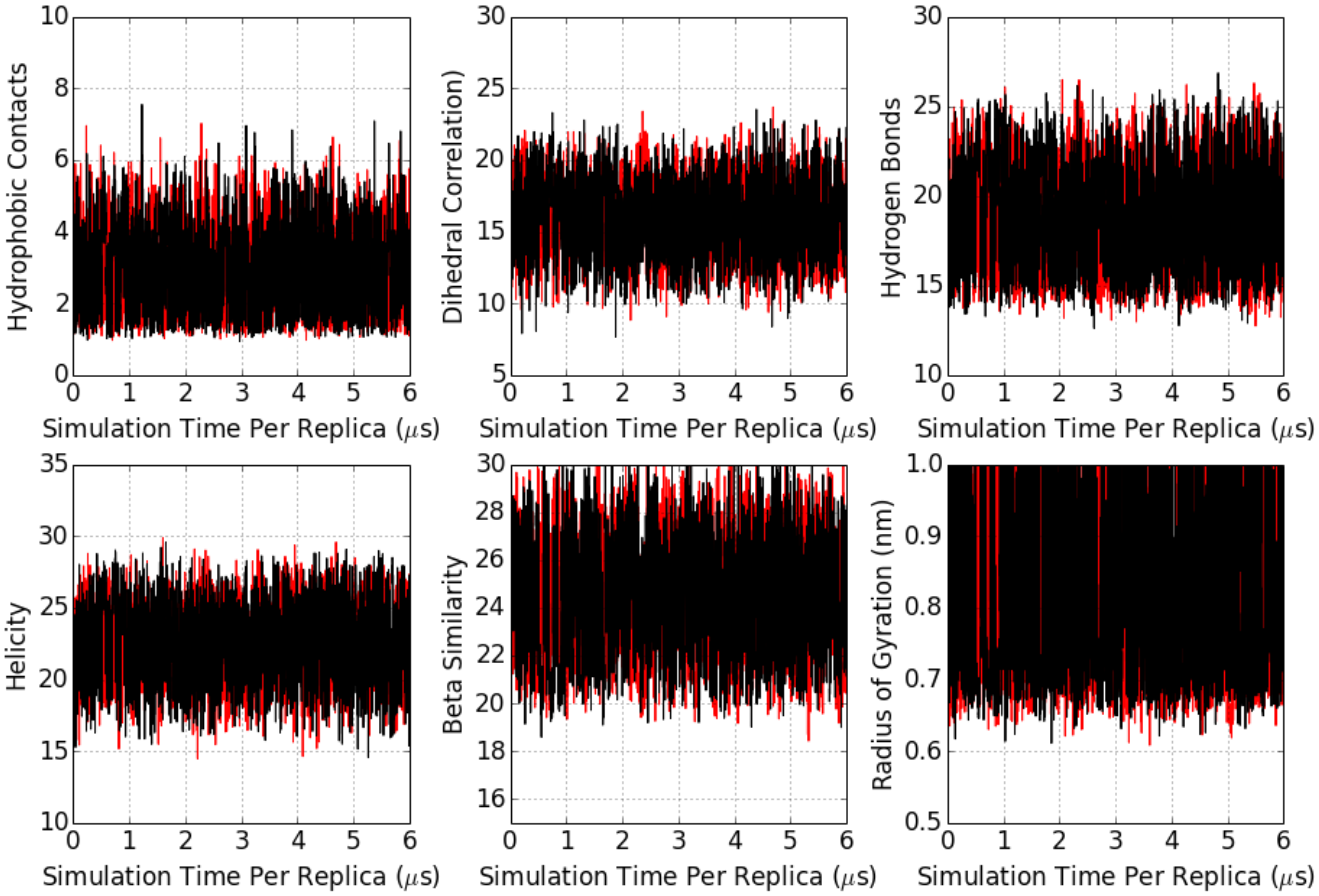

Fig. S5. Assessing convergence of PT simulation I. Analysis of the diffusion of the demuxed trajectories in the reduced dimensional space of the six monitored CVs. Two representative examples (black and red) from the ~75 μs TRPC PT simulation are shown. The other 10 replicas show similar behavior but are omitted for the sake of clarity. Note that owing to the extremely fast diffusion in the CV space, the plots below might be confused with those that are obtained from the (discontinuous) trajectories at constant temperature. However, the data shown are indeed for the demuxed trajectories. We carefully verified that the timescale for changing, for example, from the highest to lowest radius of gyration values in this model (~20 ps) is commensurate with the exchange rate (~20%) and, number of replicas (12), and exchange attempt frequench (200 ps).

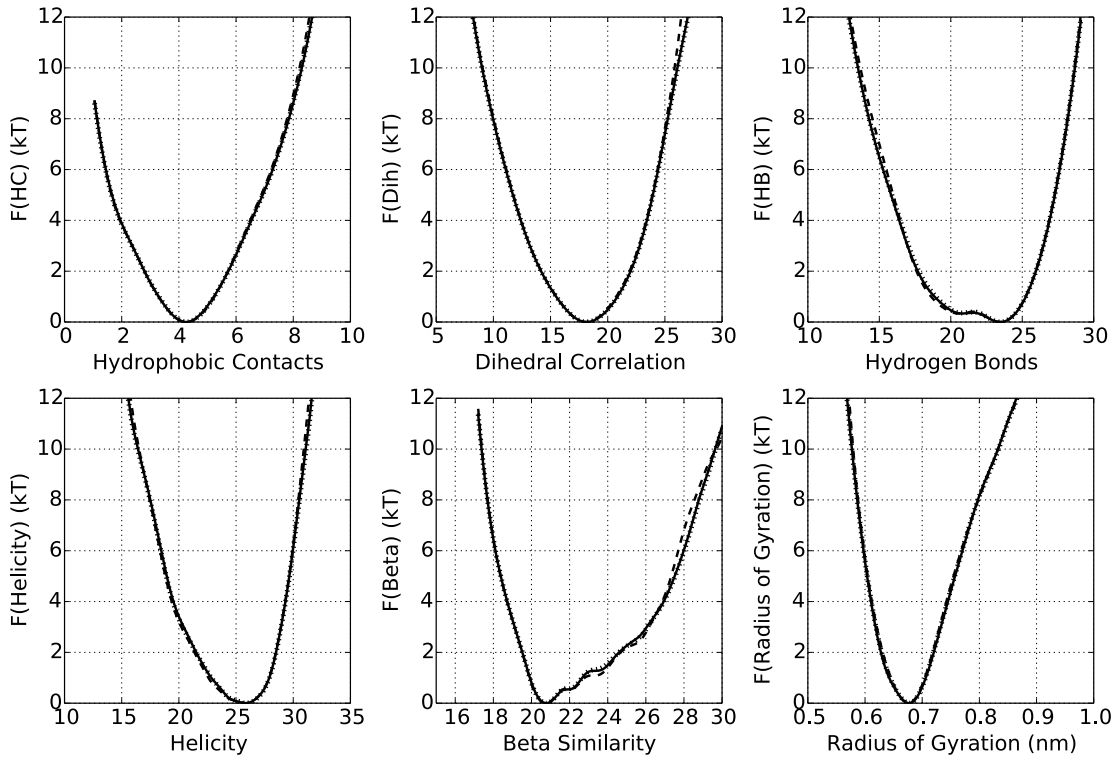

Fig. S6. Assessing convergence of PT simulation II. Each panel shows the free energy estimate for each of the six monitored CVs (which will be biased in the subsequent PBMetaD and BEM simulations) at 50% completion (dashed line), 80% completion (dotted line) and final (solid line). Even at 50% completion of the simulation, the free energies are indistinguishable within a fraction of $k_B T$.

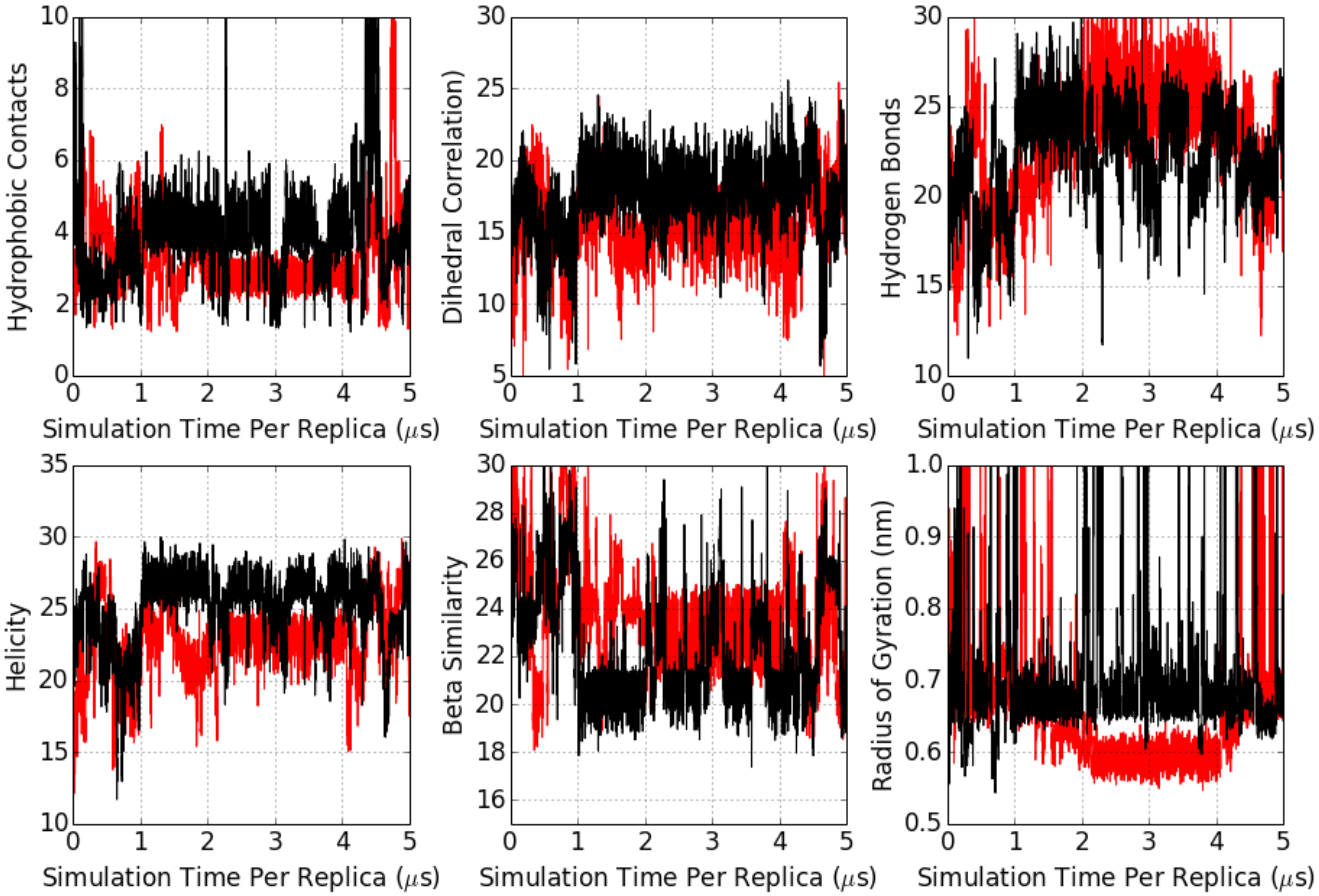

Fig. S7. Assessing convergence of PBMetaD simulation I. Analysis of the diffusion of the multiple walker trajectories in the reduced dimensional space of the six biased CVs. Two representative examples (black and red) from the 30 μs TRPC PBMetaD simulation. The other 4 walkers show similar behavior but are omitted for the sake of clarity.

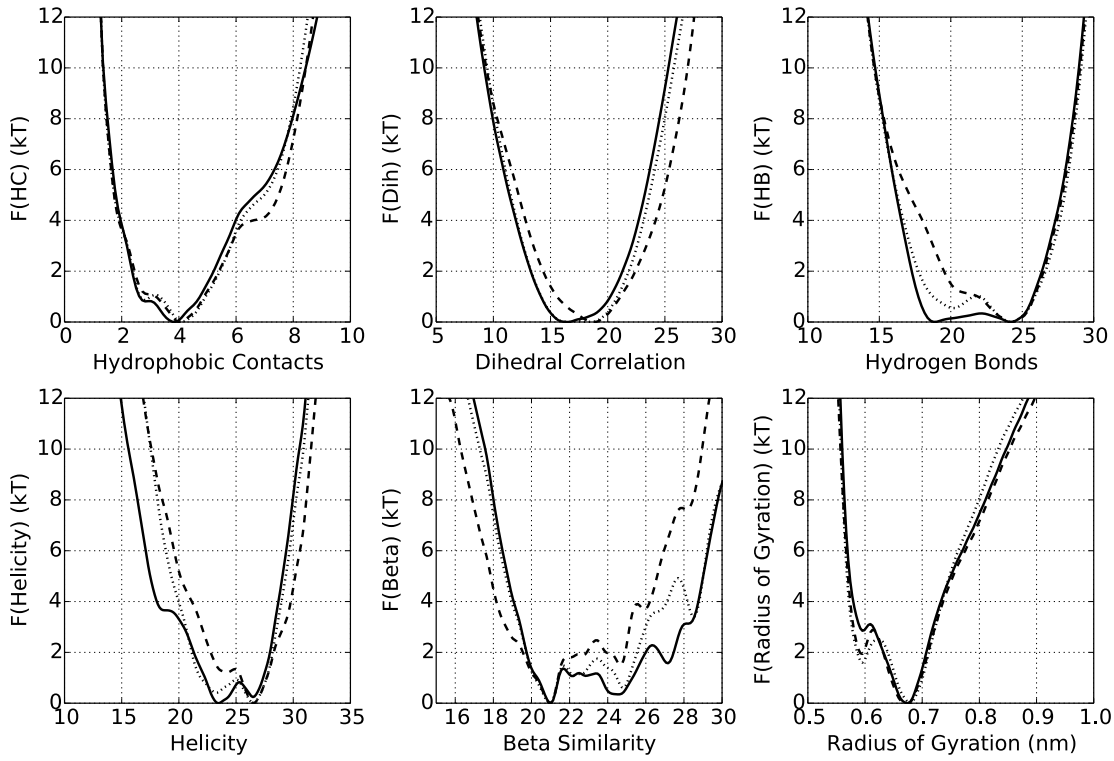

Fig. S8. Assessing convergence of PBMetaD simulation II. Each panel shows the free-energy estimate from the PBMetaD TRPC simulation for each of the six biased CVs at 50% completion (dashed line), 80% completion (dotted line) and final (solid line).

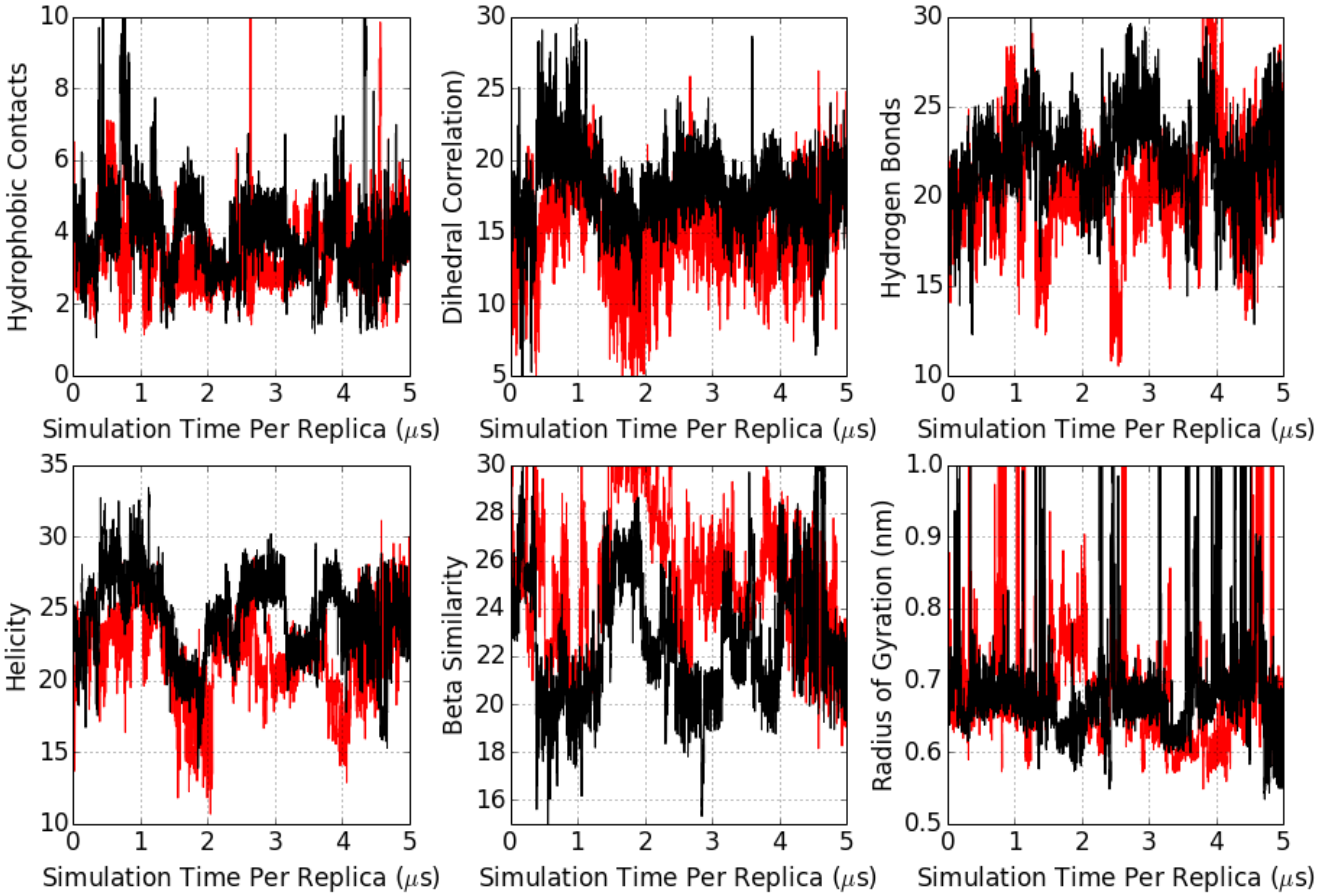

Fig. S9. Assessing convergence of BEM100 simulation I. Analysis of the diffusion of the demuxed trajectories in the reduced dimensional space of the six biased CVs. Two representative examples (black and red) from the ~30 μs TRPC BEM100 simulation are shown. The other 4 replicas show similar behavior but are omitted for the sake of clarity.

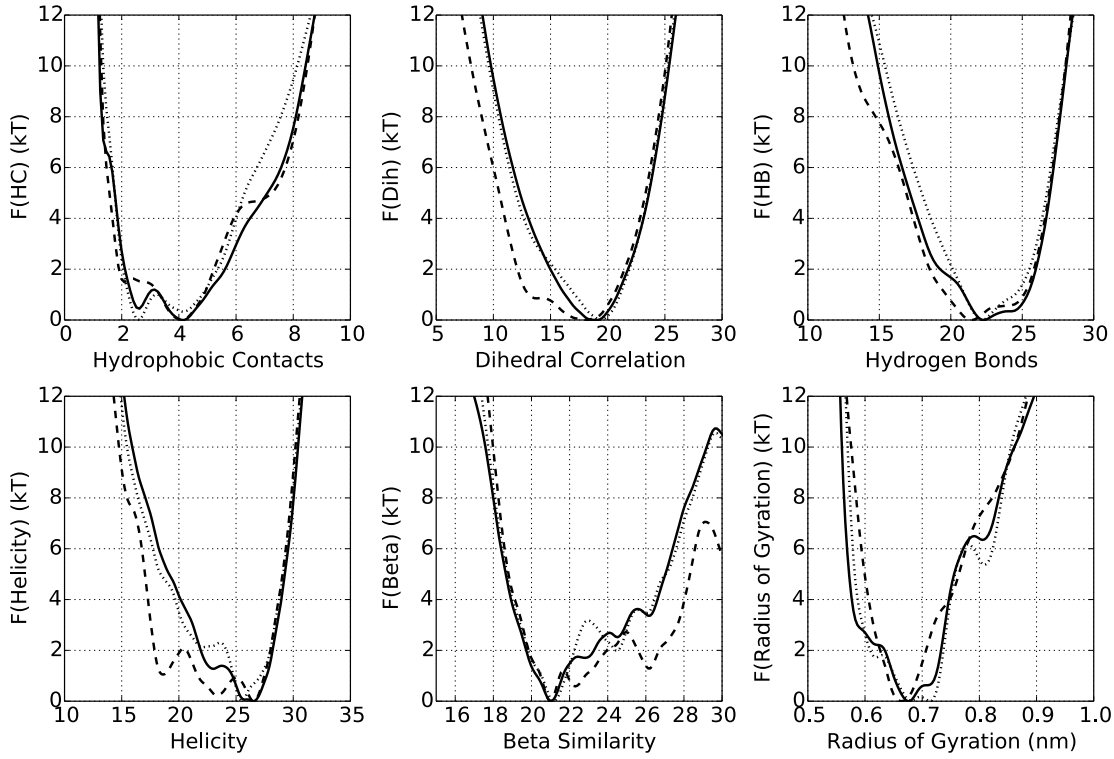

Fig. S10. Assessing convergence of BEM100 simulation II. Each panel shows the free-energy estimate from the BEM100 TRPC simulation for each of the six biased CVs at 50% completion (dashed line), 80% completion (dotted line) and final (solid line).

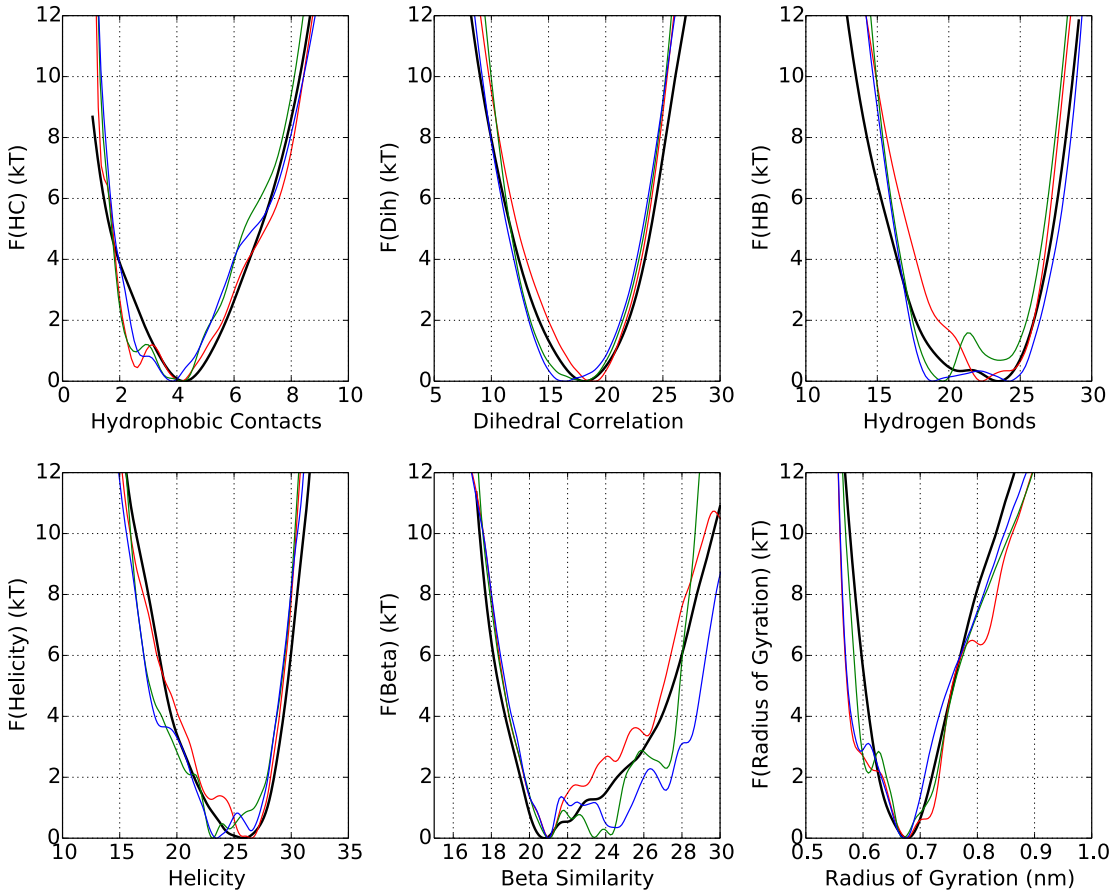

Fig. S11. Comparison of the final free-energy profiles from PT (black), BEM100 (red), BEM1000 (green), and PBMetaD (blue) for each of the six CVs.

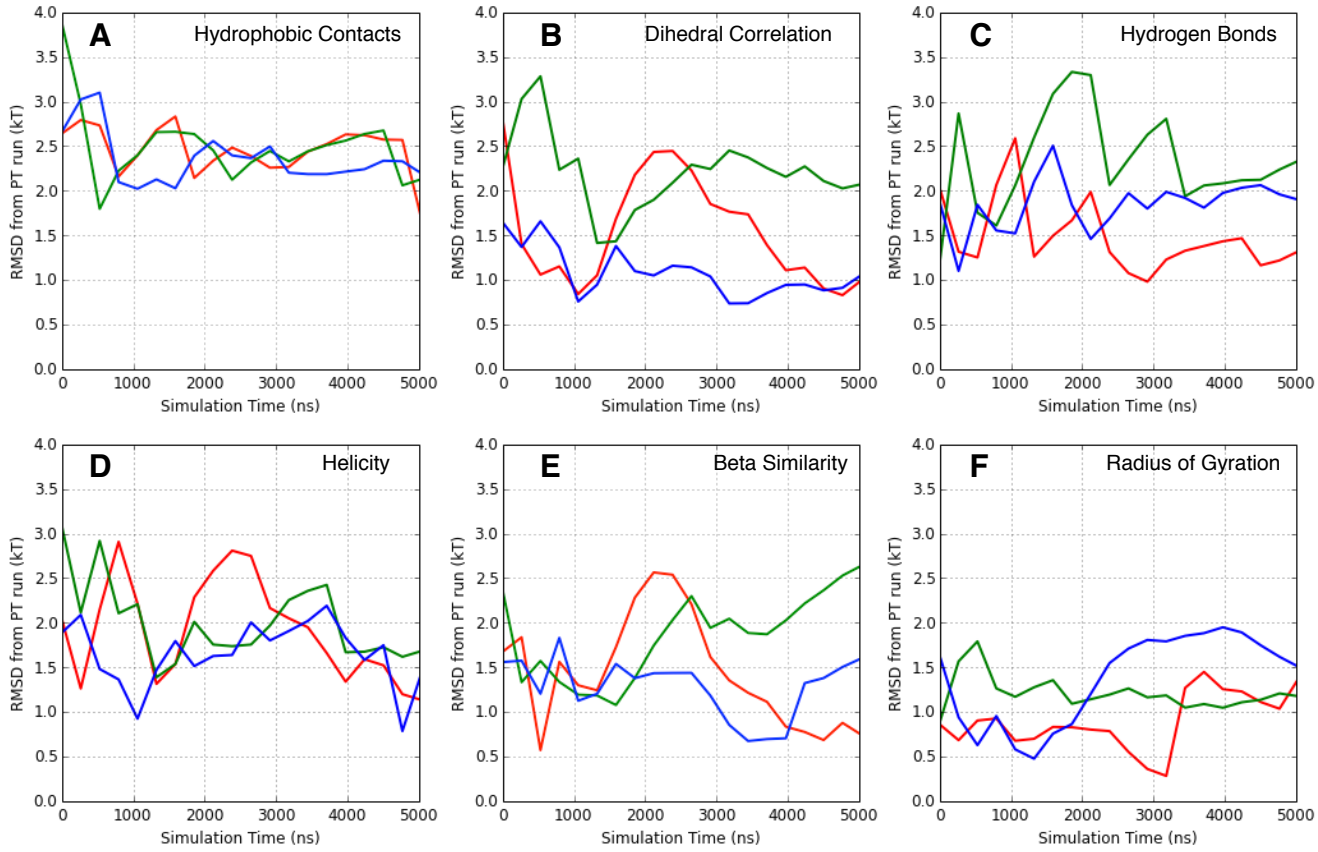

Fig. S12. Convergence of the deviation between the monodimensional free-energies obtained from PT and PBMetaD (blue), BEM100 (red), and BEM1000 (green). The convergence metric is the RMSD applied in the same manner as in the model system. The simulation time represents MD simulation time per replica in the implicit solvent model, with each curve representing 30 μs of aggregate sampling. The six panels represent the biased CVs: A) hydrophobic contacts, B) dihedral correlation, C) hydrogen bonds, D) alpha helicity, E) beta similarity, F) radius of gyration