Supporting information for: Reduction of All-Atom Protein Folding Dynamics to One-Dimensional Diffusion

Wenwei Zheng and Robert Best*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Insitutes of Health, Bethesda, MD 20892-0520

* robertbe@helix.nih.gov

SUPPORTING TABLES

TABLE S1. Details of simulations[2, 3] used in analysis. All simulations used the CHARMM 22* force field, which is a variant of CHARMM 22[4] with backbone adjustments [5]. Shown are the number of residues N_{res}, the duration of each simulation t_{sim}, the boundaries of Q used to define transition paths (Q_{u} on the unfolded side and Q_{f} on the folded side of the barrier), the number of transition paths N_{TP}, the curvatures of the unfolded state (ω_{u}), barrier ($\omega_{\ddagger}$) and folded state (ω_{f}), the barriers for folding (ΔG_{f}) and unfolding (ΔG_{u}), the averaged constant diffusion coefficient D_c, and the resulting Kramers prefactor $k_0 = D_c \omega_{\mathrm{u}} \omega_{\ddagger} / (2\pi k_{\mathrm{B}} T)$.

Protein (N_{res})	t_{sim}	Q_{u}	Q_{f}	N_{TP}	ω_{u}	$\omega_{\ddagger}$	ω_{f}	ΔG_{f}	ΔG_{u}	D_c	k_0
	μs				$\sqrt{k_{\mathrm{B}}T}$	$\sqrt{k_{\mathrm{B}}T}$	$\sqrt{k_{\mathrm{B}}T}$	$k_{\mathrm{B}}T$	$k_{\mathrm{B}}T$	$\mu\mathrm{s}^{-1}$	$\mu\mathrm{s}^{-1}$
CLN025 (10)	106	0.03	0.80	104	44.6	7.8	11.3	3.89	3.78	1.75	96.6
Trp Cage (20)	208	0.10	0.95	24	8.9	9.6	30.8	3.93	3.61	0.34	4.58
BBA (28)	325	0.05	0.77	24	13.4	7.4	12.7	2.44	0.94	0.084	1.33
NLE Villin (35)	125	0.28	0.94	86	7.6	9.9	20.3	2.32	2.08	0.38	4.52
GTT WW (35)	1137	0.05	0.92	23	17.9	16.1	31.0	4.51	6.21	0.13	6.09
mini-NTL9 (39)	2936	0.12	0.96	34	11.3	8.7	36.4	3.20	5.75	0.10	1.55
NuG2 (56)	1155	0.30	0.92	27	7.6	15.7	31.2	3.03	4.08	0.036	0.68
α_3D (73)	707	0.19	0.70	24	12.4	11.7	22.6	2.38	3.03	0.036	0.82
Ubiquitin (76)	7839	0.10	0.95	28	16.1	11.8	36.5	6.12	6.45	0.12	3.74
λ-repressor (80)	643	0.40	0.90	10	10.2	9.9	20.6	2.02	1.12	0.028	0.45

TABLE S2. Correlations of landscape parameters and Kramers prefactors k_0 with protein length. For 10 data points, the critical value of t at the two-sided 5 % significance level is $|t| = 2.23$.

Parameter	Spearman Correlation	Student t
Q_{u}	0.81	3.97
$Q_{\ddagger}$	0.08	0.22
Q_{f}	0.01	0.03
ω_{u}	-0.20	-0.58
$\omega_{\ddagger}$	0.50	1.65
ω_{f}	0.50	1.65
ΔG_{f}	-0.04	-0.82
ΔG_{u}	0.01	0.33
D_c	-0.75	-3.19
k_0	-0.84	-4.04

SUPPORTING FIGURES

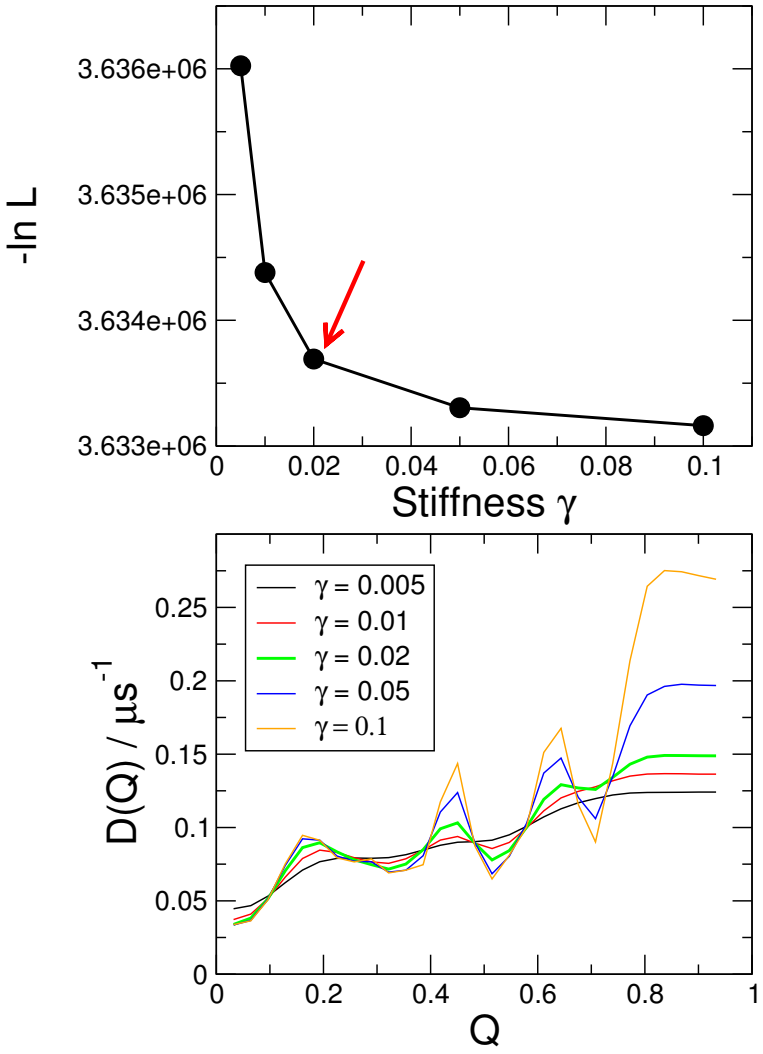

FIG. S1. Choice of stiffness parameter γ. Top: variation of the negative log-likelihood of the diffusion model as a function of γ. Red arrow indicates the chosen $\gamma = 0.02$. Bottom: Position-dependent diffusion coefficient profiles $D(Q)$ obtained using γ values in top plot.

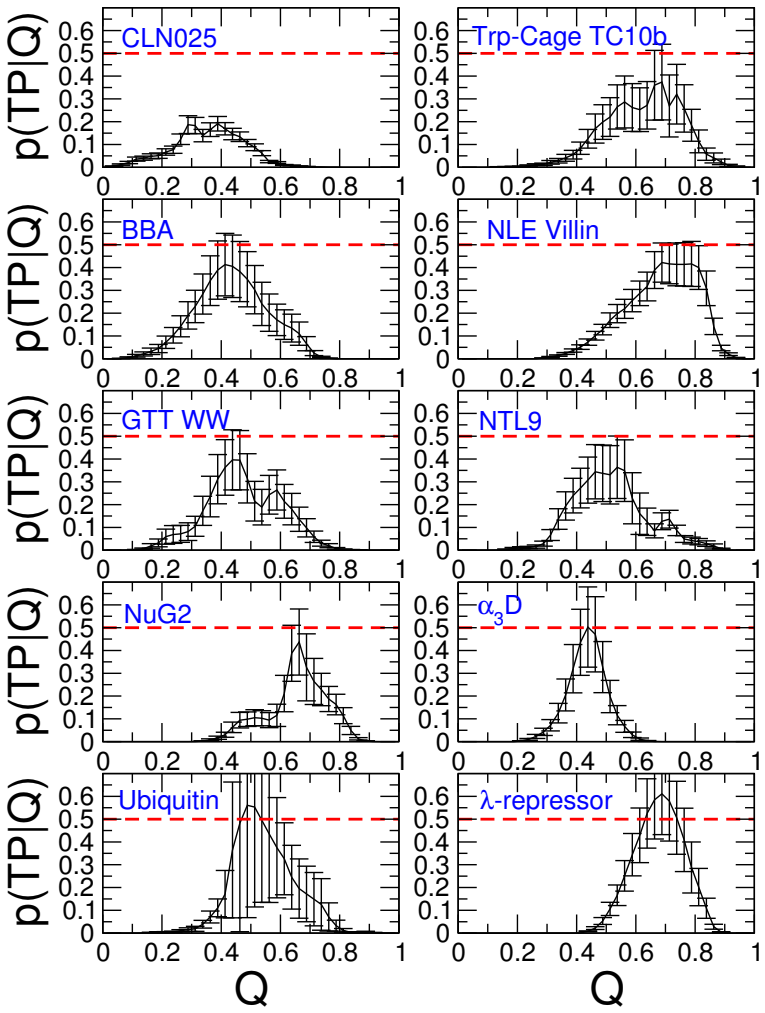

FIG. S2. Bayesian criterion for reaction coordinate quality. For each of the ten proteins considered, we plot the quantity $p(TP|Q)$ – the probability of being on a transition path given a particular value of Q. Errors are determined by block error analysis, and the broken red line shows the theoretical maximum $p(TP|Q) = 0.5$. In some cases, $p(TP|Q)$ exceeds this value, due to limited statistics.

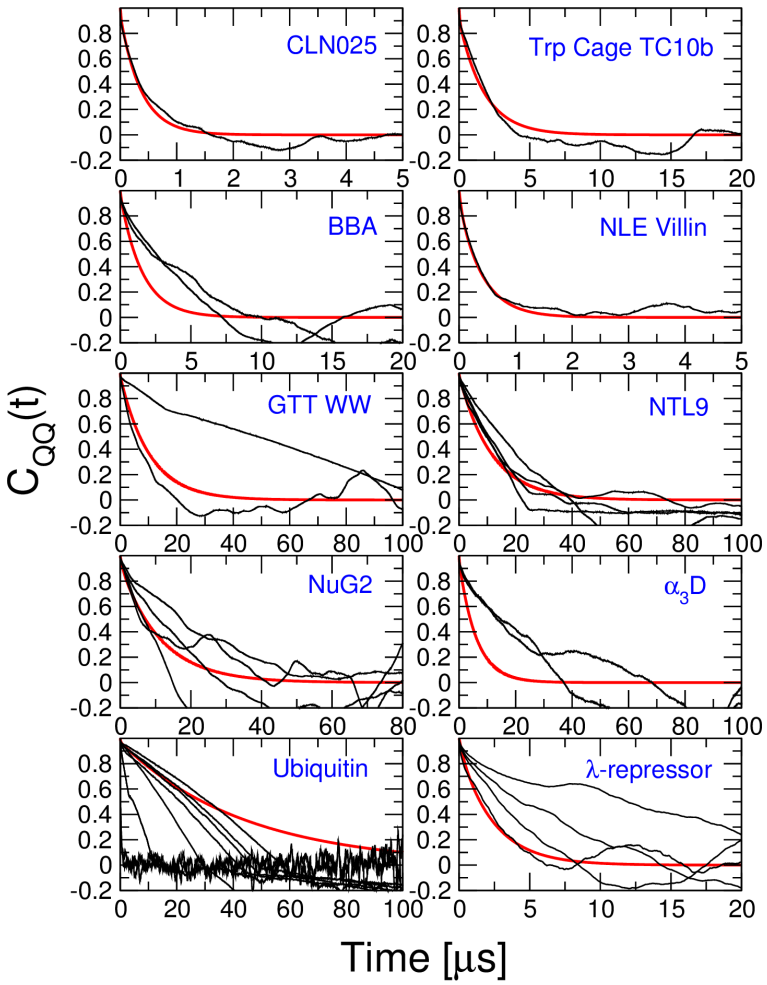

FIG. S3. Autocorrelation functions for Q computed from the diffusion model (red) and from the original trajectories (black). Individual correlation functions are shown for each separate trajectory (where applicable), to give an idea of the uncertainty. For ubiquitin, two of the trajectories contained no folding or unfolding events, and hence $C_{QQ}(t)$ decays very rapidly.

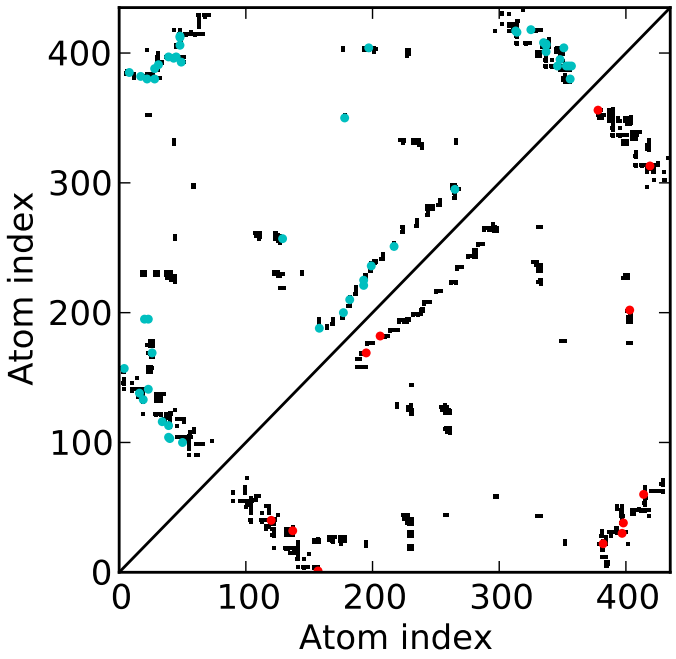

FIG. S4. The most positive (w_i >0.01 shown in cyan dots) and negative weights (w_i <-0.01 shown in red dots) in calculating $Q_{\rm opt}$ in NuG2.

REFERENCES

[1] Best, R. B.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732–6737.

[2] Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517–520.

[3] Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5915–5920.

[4] Mackerell, A. D., Jr. et al. *J. Phys. Chem. B* **2000**, *102*, 3586–3616.

[5] Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Biophys. J.* **2011**, *100*, L47–L49.