Supporting Information

Wereszczynski and McCammon 10.1073/pnas.1117441109

SI Text

Comparison of Convention Molecular Dynamics (cMD), Accelerated Molecular Dynamics (aMD), and Adaptive Biasing Force (ABF) simulations. The calculation of potentials of mean force (PMF) presented in Fig. 3 allows for the opportunity to reflect on the sampling observed in the cMD and aMD simulations (Fig. 2). Based upon experimental evidence, we expected to see an opening motion of dimers not bound to nucleotides, and a closing motion of those in the two-nucleotide states. In simulations of the apo state, the cMD simulations sampled regions local to their initial conformations, which when compared to the PMF are near local energy minima (Fig. S4). However, it was curious to observe that in aMD simulations of the fully closed state the dimer remained closed, whereas in the open state it transitioned to closed conformations quickly, where it remained. Estimation of the position-dependent diffusion coefficient (Fig. S3) suggests that this sampling may be a kinetic effect: the diffusivity of Get3 along the full-correlation analysis (FCA) mode 1/4 space may be as much as an order of magnitude slower in closed conformations then in open ones. Therefore, the likely explanation for the behavior observed in apo state aMD calculations is that the sampling time required to transition from closed to open states is likely much greater then for the reverse process, and that the timescale sampled here is sufficient to observe closing but not opening. In the open/two-nucleotide simulations, it was observed that the aMD (and to a lesser extent the cMD) projections displayed an opening motion of the dimer for simulations initiated in the open conformation, which is the opposite of what would be expected from experimental results. The free energy landscapes of these systems both show local minima at high FCA 1 projection values (the wide-open states), thus the simulations followed the initial free energy gradient they experienced and became trapped in high-energy minima. These results highlight that, although aMD simulations allow for the observation of long timescale event that are beyond the reach of conventional simulations, they can still suffer from finite sampling effects in large biomolecular systems. Complementing aMD with free energy calculations appears to be an effective strategy for rigorously describing the thermodynamics or large biomolecular transitions, as the combination can be used to overcome limitations inherent to each method.

Estimation of Position-Dependent Diffusion Coefficient. The position-dependent diffusion coefficient (Fig. S3) was estimated by computing the diffusion coefficient of each ABF window over the final 5 ns of simulation, and interpolating between values on the FCA mode 1/4 space using the SciPy linear radial basis function with a radius of five and a smoothing parameter of one.

SI Materials and Methods. Models of the fully closed and open Get3 dimer were constructed from structures with Protein Data Bank ID 2WOJ and 3H84 (1, 2), with missing protein segments built from segments resolved in the other structure when available, and modeled with the program PRIME (3). Protein parameters were derived from the AMBER99SB force field (4), with ADP and ATP parameters from Meagher et al. and zinc ion and zinc-coordinating cysteine parameters from the zinc AMBER force field (5, 6). ADP systems did not include any additional inorganic phosphates. Systems were solvated in an orthorhombic transferable intermolecular potential three point (TIP3P) water box with 150 mM NaCl concentration such that there was a minimum 10-Å buffer between protein heavy atoms and the box boundaries (7, 8). All simulations were performed with NAMD 2.8 (9, 10). Long-range electrostatics were treated with particle-mesh Ewald using a maximum grid spacing of 1 Å in each dimension, along with a cubic spline for cMD and aMD simulations and a fifth-order spline for ABF calculations (11). Temperature was maintained at 300 K in all simulations through the use of Langevin dynamics with a damping coefficient of 2 ps^{-1}, and in constant number, pressure, and temperature (NPT) simulations, pressure was controlled by a Langevin piston with a target pressure of 1.01325 bar, a 100-fs piston period, a 50-ps damping timescale, and a piston temperature of 300 K (12, 13). Following 50,000 steps of minimization, restraints on protein heavy atoms beginning at 10 $kcal/(mol \cdot Å^2)$ were gradually released for the first 1 ns of cMD simulation, while the system was heated by temperature rescaling over the first 3 ps. Both cMD and aMD simulations were run for a total of 100 ns each, with aMD simulations seeded from snapshots taken at 10, 15, and 20 ns into the cMD simulations. Conventional MD simulations were performed in the NPT ensemble, whereas aMD were done in the constant number, volume and temperature (NVT) ensemble. The boosting parameters for aMD simulations were $E_{dihed} = 11{,}500$, $alpha_{dihed} = 500$, $E_{Total} = 18{,}500$, and $alpha_{Total} = 18{,}500$ kcal/mol (14). A restraint of 200 $kcal/(mol \cdot rad^2)$ was applied to the ϕ and ψ backbone dihedrals of protein elements in well-defined secondary structure elements in the aMD simulations using the "SSRestraints" plug-in to visual molecular dynamics (VMD) (15). Analysis of simulations was performed with a combination of VMD (16), Gromacs (17), NumPy (18), SciPy (19), and matplotlib (20).

Free energy calculations were performed using the ABF method (21, 22). The dimensions that were projected along in the cMD and aMD simulations, and were biased along in ABF calculations, were generated from a FCA over all C^α atoms, excluding the terminal five residues, of the six cMD simulations (23). The space sampled in aMD calculations was divided into a series of 42 overlapping subspaces, "windows," for ABF calculations (see Fig. S6). For each window, calculations were performed in three phases: targeted MD (24), equilibration, and ABF. Before calculations, molecular mechanics/generalized born analysis was performed on the aMD trajectories (25), and the low energy structure in each window was chosen as the initial target for each window. Windows were initialized beginning in the fully closed conformation, and a 1 ns targeted MD simulation in the NPT ensemble, with a force constant of 1,000 $kcal/(mol \cdot Å^2)$ on all protein heavy atoms, was used to transform the protein close to the desired conformation. Then a 4-ns equilibration was performed, with the first 3 ns in the NPT ensemble before a switch was made to the NVT ensemble for the final 1 ns. ABF calculations were performed with a bin spacing of 5 Å in each dimension, and a half-harmonic force of 1 $kcal/(mol \cdot Å^2)$ was applied at the boundaries to ensure the system remained in the desired window. The ABF force was only applied after 10,000 samples of the mean force were generated in each bin. Each window was simulated for a minimum of 20 ns, at which time convergence was checked by comparing the mean forces to the ones resolved 5 ns prior, and simulations were ended if the root mean square of the ABF forces was less than 0.2 $kcal/(mol \cdot Å)$ and the maximum force change was less than 1 $kcal/(mol \cdot Å)$. For simulations that were continued, an iterative approach was chosen in which the window was simulated for 5 ns more and the forces reexamined, which was continued until the above criteria were satisfied. For regions that were still poorly sampled, and for regions with large FCA 4 values in the one ADP calculations, additional windows were inserted to

enhance sampling. In Fig. S7, one-dimensional PMF calculations are presented, which are computed by Boltzmann weighting the free energy values over the FCA mode 4 dimension for coordinates along FCA mode 1, with the full level of sampling used in addition to 5, 10, and 15 ns less sampling per window. The favorable agreement between the full sampling PMF and the 5 ns per window less sampling PMF (which amounts to 235–300 ns less total sampling over the entire PMF) indicates good convergence of these calculations. Error estimates of the ABF calculations were computed by a bootstrap-like analysis in which 100 new PMFs were generated with each force being chosen from a Gaussian distribution with a mean centered around the calculated ABF value, and a distribution resulting from the standard error of that force (using a decorrelation time of 2.5 ps). Values in Table 2 were computed by a Boltzmann average and a standard deviation of the free energies in each region.

1. Mateja A, et al. (2009) The structural basis of tail-anchored membrane protein recognition by Get3. *Nature* 461:361–366.
2. Hu J, Li J, Qian X, Denic V, Sha B (2009) The crystal structures of yeast Get3 suggest a mechanism for tail-anchored protein membrane insertion. *PLoS One* 4:e8061.
3. Jacobson MP, et al. (2004) A hierarchical approach to all-atom protein loop prediction. *Proteins* 55:351–367.
4. Hornak V, et al. (2006) Comparison of multiple Amber force fields and development of improved protein back-bone parameters. *Proteins* 65:712–725.
5. Meagher KL, Redman LT, Carlson HA (2003) Development of polyphosphate parameters for use with the AMBER force field. *J Comput Chem* 24:1016–1025.
6. Peters MB, et al. (2010) Structural survey of zinc containing proteins and the development of the zinc AMBER force field (ZAFF). *J Chem Theory Comput* 6:2935–2947.
7. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
8. Case DA, et al. (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688.
9. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802.
10. Wang Y, Harrison CB, Schulten K, McCammon JA (2011) Implementation of accelerated molecular dynamics in NAMD. *Comput Sci Discovery* 4:015002.
11. Darden T, York D, Pedersen L (1993) Particle mesh Ewald—an $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
12. Martyna GJ, Tobias DJ, Klein ML (1994) Constant-pressure molecular-dynamics algorithms. *J Chem Phys* 101:4177–4189.
13. Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) Constant-pressure molecular-dynamics simulation—the langevin piston method. *J Chem Phys* 103:4613–4621.
14. Wereszczynski J, McCammon JA (2012) *Computational Drug Discovery and Design*, ed. Baron R (Humana Press, Clifton, NJ), pp 515–524.
15. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16:673–683.
16. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graphics* 14:33–38.
17. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447.
18. Oliphant T (2007) Python for scientific computing. *Comp Sci Engin* 9:10–20.
19. Jones E, et al. (2001) *SciPy: Open source scientific tools for Python*, http://www.scipy.org.
20. Hunter JD (2007) Matplotlib: A 2D graphics environment. *Comput Sci Eng* 9:90–95.
21. Darve E, Rodríguez-Gómez D, Pohorille A (2008) Adaptive biasing force method for scalar and vector free energy calculations. *J Chem Phys* 128:144120.
22. Hénin J, Fiorin G, Chipot C, Klein M (2010) Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J Chem Theory Comput* 6:35–47.
23. Lange O, Grubmüeller H (2008) Full correlation analysis of conformational protein dynamics. *Proteins Struct Funct Bioinf* 70:1294–1312.
24. Young MA, Gonfloni S, Superti-Furga G, Roux B, Kuriyan J (2001) Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* 105:115–126.
25. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55:383–394.

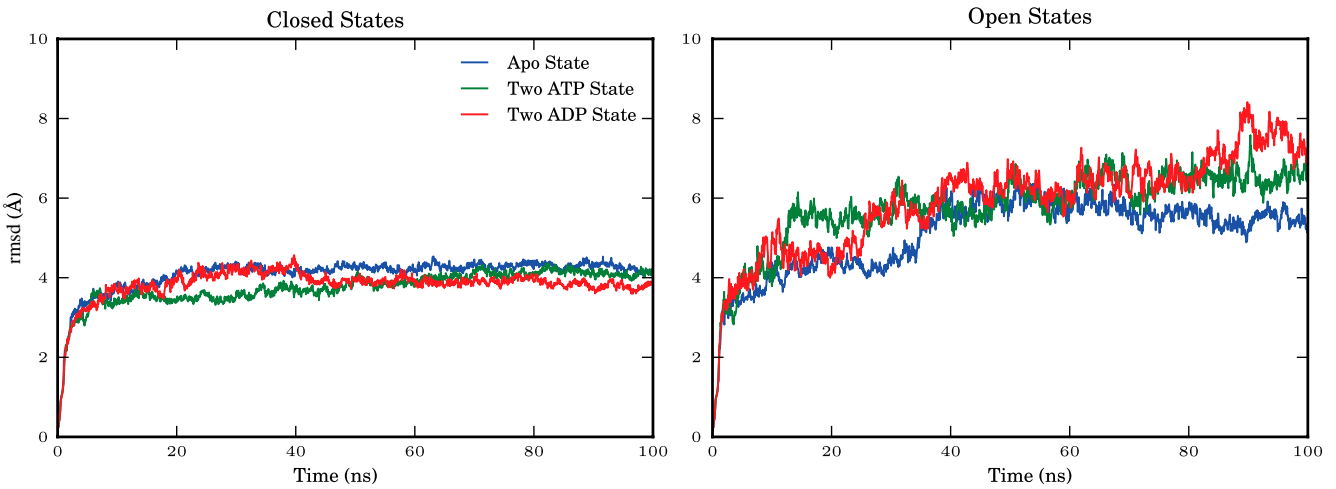


Fig. S1. Root mean square deviations of the cMD simulations.

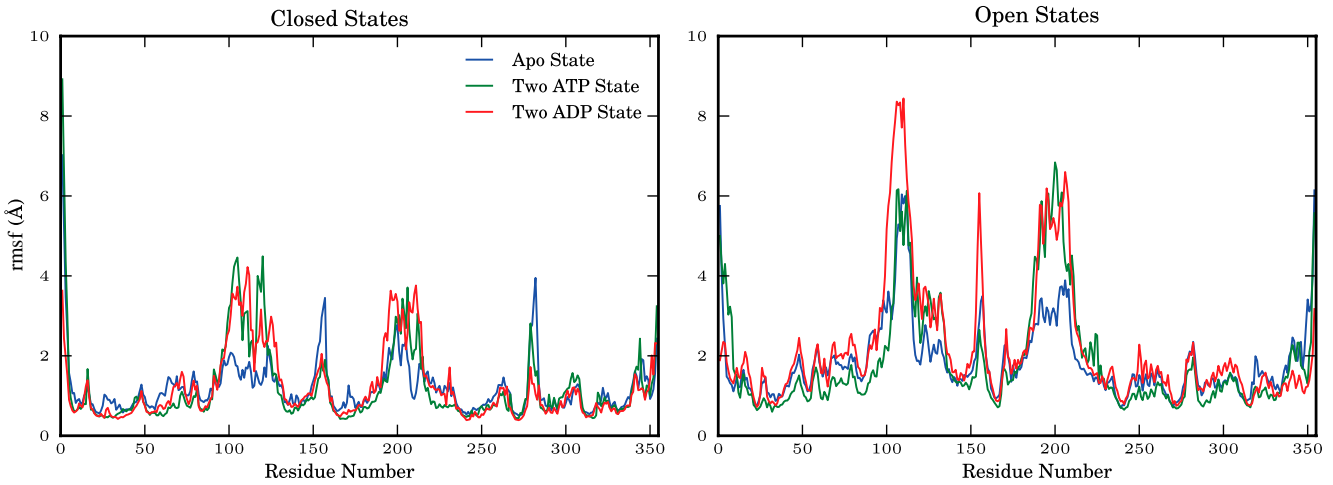


Fig. S2. Root mean square fluctuations of the cMD simulations.

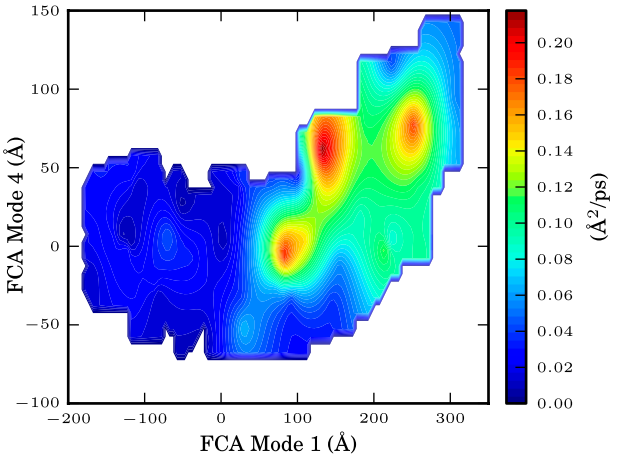


Fig. S3. The position-dependent diffusion coefficient estimated from the no-nucleotide ABF simulations.

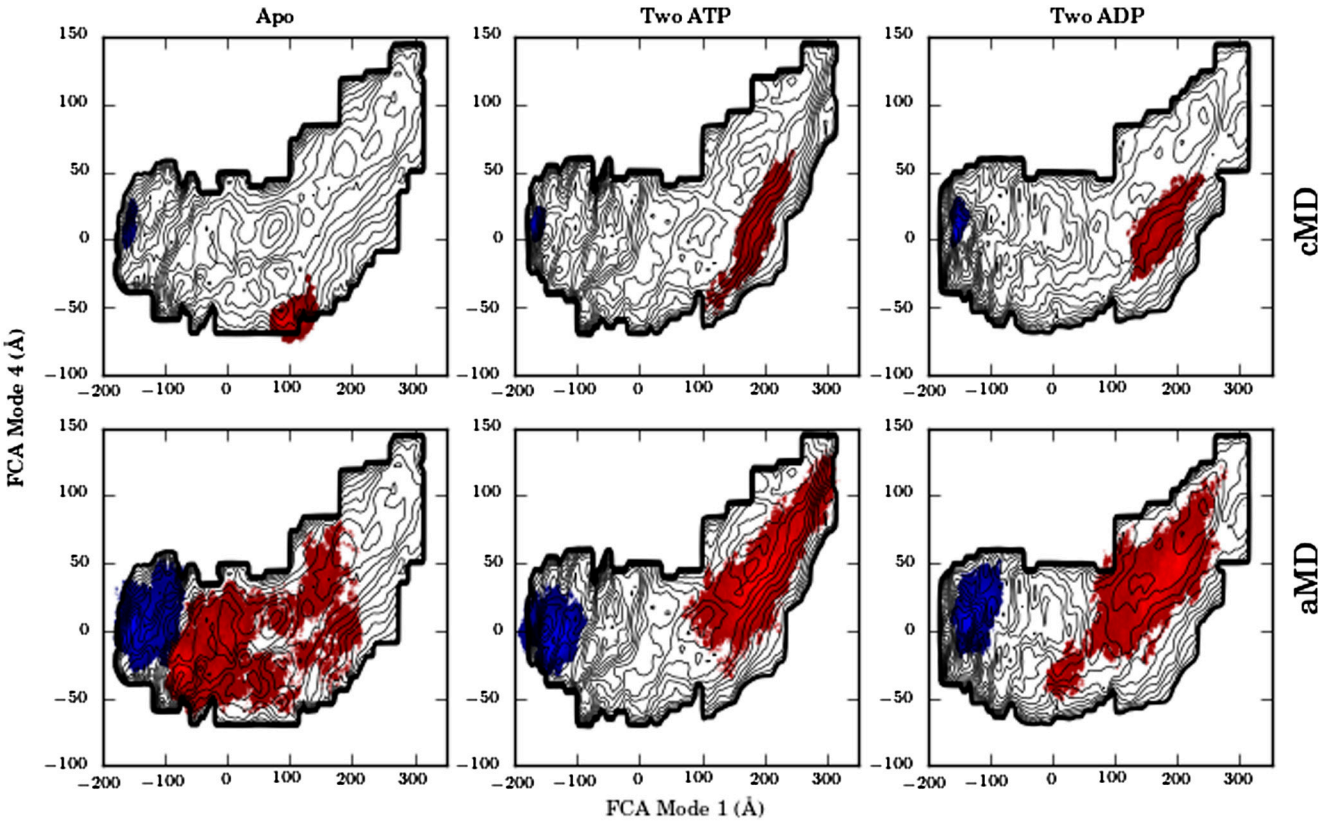


Fig. S4. Sampling of cMD and aMD simulations (Fig. 3) overlayed onto the computed free energy profiles (shown in Fig. 4).

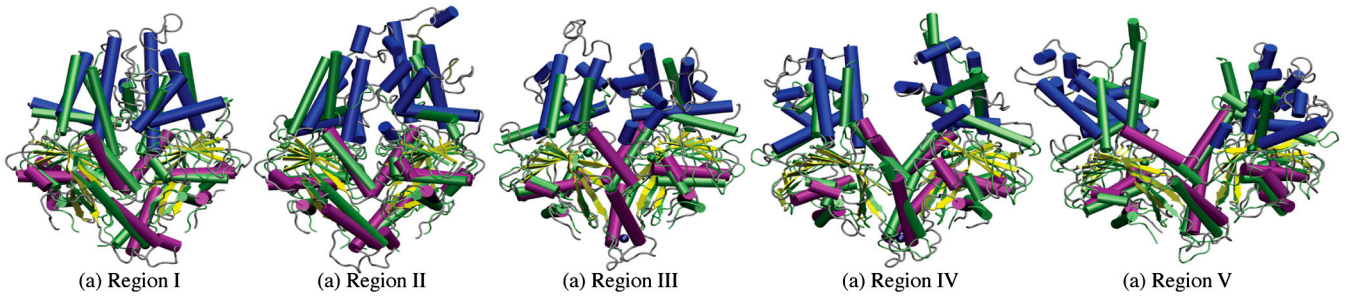


Fig. S5. Overlay of representative structures from each of the PMF regions to their representative crystal structures. Note that the wide-open state has not been experimentally observed and is therefore overlayed with the open state.

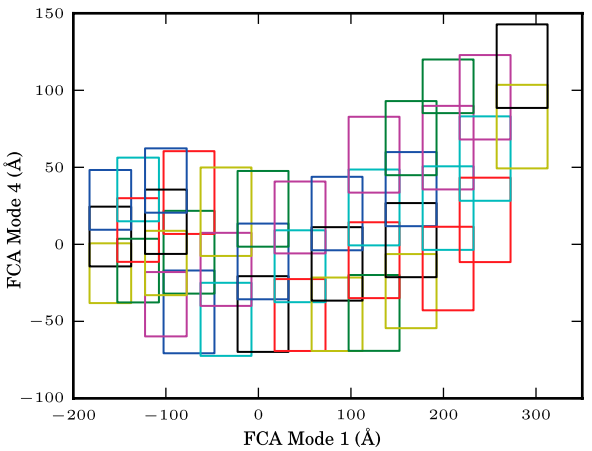


Fig. S6. The 42 windows FCA mode 1/4 space was divided into to improve the efficiency of ABF calculations.

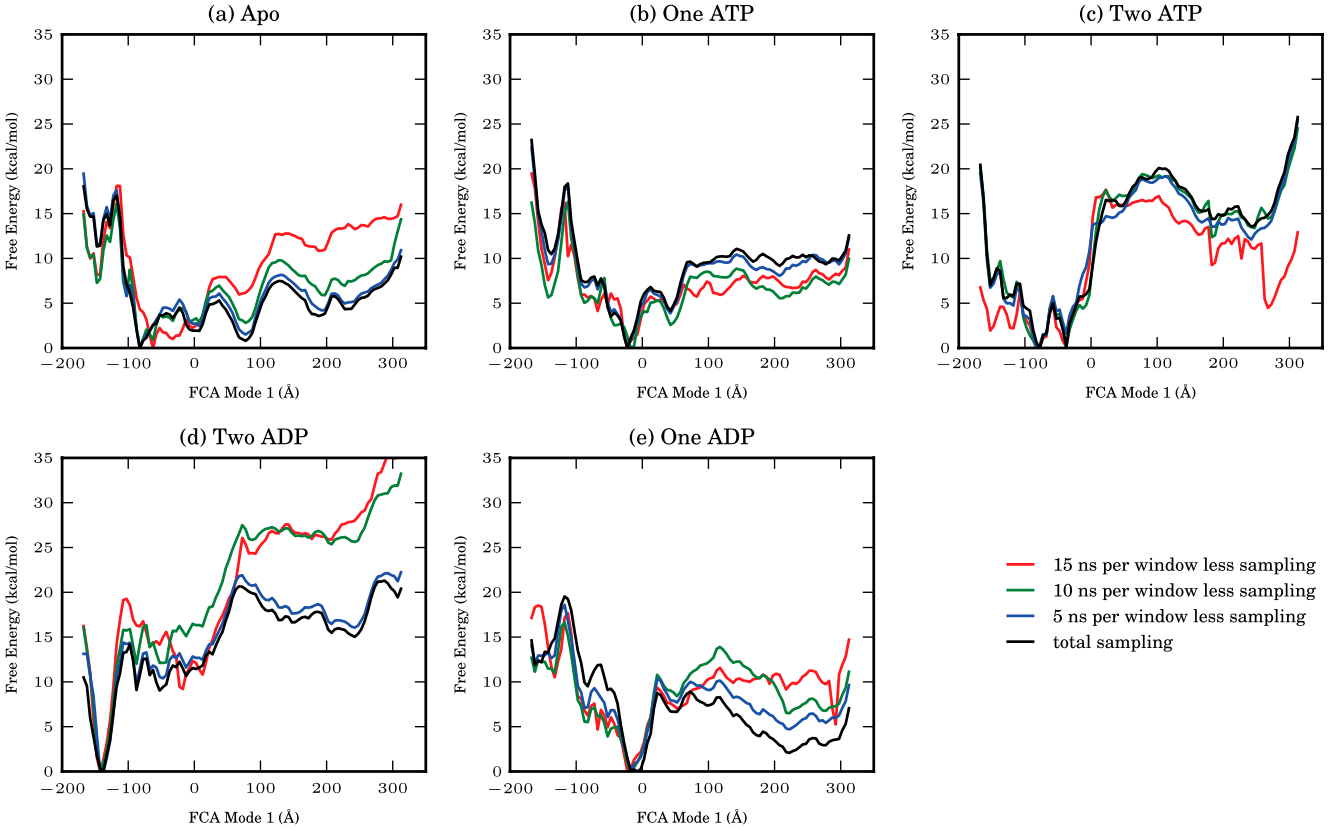
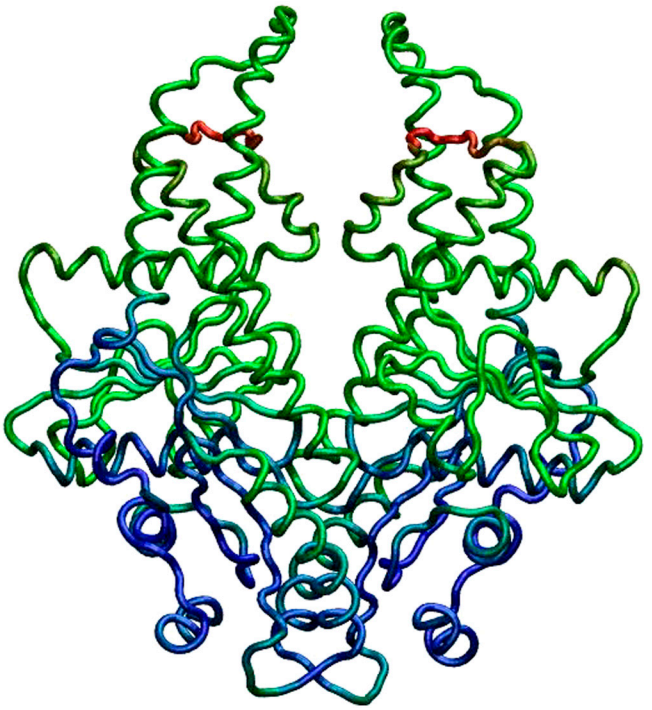
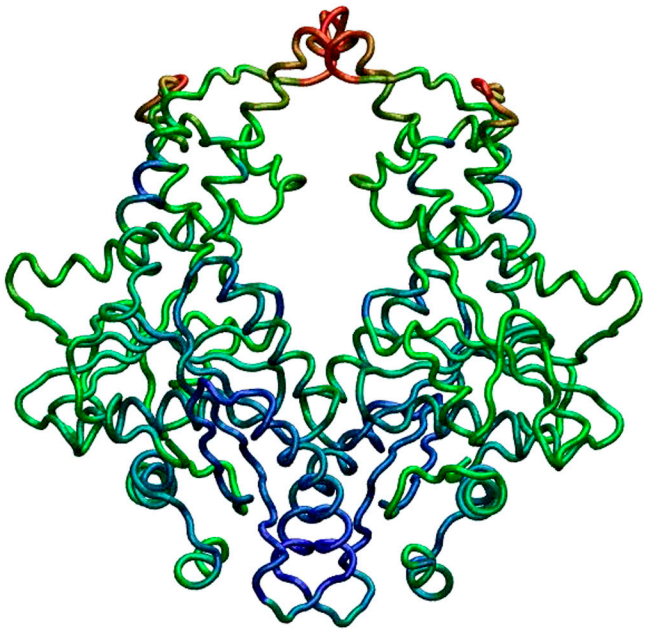


Fig. S7. One-dimensional PMF calculations for each of the five nucleotide states with the full amount of sampling employed in black, and lesser levels of sampling (5, 10, and 15 ns per window) shown in blue, green, and red. The high degree of similarity between the full sampling per window and 5 ns less per window (black and blue, respectively) suggest there is good convergence in the ABF results. Note that for each 5 ns per window removed, a total of 235–300 ns of sampling is removed from each PMF.



Movie S1. The motions along FCA mode 1.
Movie S1 (MPG)



Movie S2. The motions along FCA mode 4.
Movie S2 (MPG)

Table S1. Structural properties of conformations in energy wells for each of the five nucleotide state in each region of FCA mode 1/4 space and crystal structures

Nucleotide state	Property	Region 1	Region 2	Region 3	Region 4	Region 5
Apo	helix separation	27.0 ± 0.40	26.3 ± 0.63	32.0 ± 2.47	42.4 ± 4.52	49.2 ± 1.49
	sheet separation	36.9 ± 0.47	33.0 ± 0.46	35.6 ± 1.49	39.7 ± 1.60	41.2 ± 0.82
	rmsd	3.20 (fully closed)	3.52 (closed)	1.86 (semiopen)	3.60 (open)	2.42 (open)
	state	fully closed	closed	semiopen	open	Wide Open
One ATP	helix separation	29.0 ± 2.01	26.4 ± 0.46	29.4 ± 0.63	35.5 ± 2.03	51.6 ± 1.74
	sheet separation	34.4 ± 1.76	32.8 ± 0.48	34.5 ± 0.75	37.0 ± 1.17	43.2 ± 0.94
	rmsd	2.90 (closed)	3.40 (closed)	2.07 (semiopen)	2.15 (semiopen)	2.94 (open)
	state	closed	closed	semiopen	semiopen	wide open
Two ATP	helix separation	29.3 ± 0.61	27.6 ± 3.05	28.6 ± 0.88	37.2 ± 1.40	38.6 ± 2.22
	sheet separation	30.8 ± 0.50	32.5 ± 0.78	33.0 ± 0.92	37.4 ± 0.70	40.1 ± 0.59
	rmsd	1.30 (fully closed)	3.22 (closed)	2.25 (semiopen)	2.97 (semiopen)	2.32 (open)
	state	fully closed	closed	closed/semiopen	semiopen	open
Two ADP	helix separation	27.8 ± 1.53	27.1 ± 0.82	28.4 ± 0.94	39.4 ± 1.36	53.6 ± 3.38
	sheet separation	34.1 ± 1.66	31.9 ± 0.71	32.3 ± 1.14	38.0 ± 1.01	42.1 ± 1.62
	rmsd	1.12 (fully closed)	2.11 (fully closed)	2.35 (semiopen)	3.04 (open)	2.72 (open)
	state	fully closed	fully closed	semiopen	open	wide open
One ADP	helix separation	28.0 ± 1.00	26.1 ± 0.46	31.4 ± 0.75	42.5 ± 2.40	55.0 ± 1.76
	sheet separation	32.3 ± 0.66	32.7 ± 0.51	33.1 ± 0.33	37.7 ± 0.76	41.5 ± 1.11
	rmsd	1.48 (fully closed)	3.47 (closed)	2.36 (semiopen)	2.63 (semiopen)	3.35 (open)
	state	fully closed	closed	semiopen	open	wide open
Crystal structures	helix separation	28.6	28.8	31.5	33.6	43.5
	sheet separation	30.1	32.1	31.7	34.7	40.3
	state	fully closed	closed	closed	semiopen	open
	PDB ID	2WOJ	3IQW	3SJD	3SJC	3H84

PDB, Protein Data Bank.

Table S2. Average and range of C^{α} rmsd values for each energy well, relative to several Get3 crystal structures

Nucleotide state	Region	Fully closed (2WOJ)	Closed (3SJD)	Semiopen (3SJC)	Open (2H84)
Apo state	1	4.43 (3.20–5.06)	5.11 (3.86–5.61)	5.67 (4.77–6.27)	9.13 (8.48–9.54)
	2	3.71 (2.99–4.08)	4.40 (3.52–4.85)	3.82 (2.88–4.70)	7.89 (7.22–8.42)
	3	4.49 (3.67–5.52)	5.38 (4.20–6.67)	2.56 (1.86–3.43)	5.66 (4.85–6.55)
	4	7.16 (6.05–8.65)	7.49 (6.46–8.38)	4.96 (3.27–7.26)	4.81 (3.60–6.37)
	5	9.66 (8.51–11.21)	10.1 (9.11–11.78)	6.64 (5.29–8.46)	3.48 (2.43–5.43)
One ATP	1	3.03 (1.82–5.19)	4.10 (2.90–6.05)	4.56 (3.36–5.96)	8.37 (7.27–9.37)
	2	3.43 (2.78–3.87)	3.95 (3.39–4.63)	4.31 (3.18–5.25)	8.44 (7.18–9.38)
	3	3.86 (3.20–4.80)	5.03 (4.23–6.03)	2.42 (2.07–3.38)	6.32 (5.51–6.90)
	4	5.99 (4.41–6.92)	6.59 (4.73–7.69)	4.06 (2.15–5.09)	5.55 (4.30–6.51)
	5	10.77 (9.64–11.98)	11.15 (10.14–12.78)	7.76 (6.57–8.95)	4.34 (2.94–5.88)
Two ATP	1	1.61 (1.30–3.67)	3.45 (2.96–4.80)	4.16 (3.85–5.22)	8.44 (8.16–8.94)
	2	3.13 (2.50–3.55)	3.80 (3.31–4.47)	4.06 (3.31–4.58)	8.13 (6.54–8.63)
	3	3.62 (2.87–4.41)	4.96 (4.03–5.74)	2.78 (2.24–3.91)	6.71 (6.02–8.22)
	4	6.61 (5.50–7.50)	7.42 (6.13–8.48)	3.57 (2.97–4.21)	3.68 (2.89–5.18)
	5	9.75 (8.62–10.56)	10.36 (9.35–11.30)	6.56 (5.43–7.71)	3.40 (2.32–4.97)
Two ADP	1	2.92 (1.12–4.54)	4.28 (2.81–5.55)	4.26 (3.69–5.33)	8.15 (7.41–8.86)
	2	3.00 (2.11–4.03)	3.69 (2.98–4.90)	3.98 (2.99–5.26)	8.24 (7.22–9.62)
	3	3.27 (2.68–3.95)	4.46 (3.40–5.04)	2.89 (2.36–4.36)	6.98 (6.03–8.47)
	4	7.38 (5.06–8.12)	8.15 (5.82–8.97)	4.42 (2.51–5.09)	3.65 (3.05–4.89)
	5	10.84 (9.68–12.55)	11.35 (10.20–12.76)	7.79 (6.40–10.08)	3.90 (2.72–6.90)
One ADP	1	2.31 (1.48–2.94)	3.60 (3.01–4.15)	4.82 (3.85–5.44)	9.03 (7.95–9.64)
	2	3.75 (2.99–4.36)	4.34 (3.47–5.01)	4.15 (2.84–5.08)	8.21 (6.94–9.04)
	3	4.71 (4.00–5.70)	5.96 (5.16–7.03)	2.91 (2.36–3.62)	6.02 (5.63–6.45)
	4	7.58 (5.36–9.43)	8.22 (6.35–9.56)	5.09 (2.63–7.69)	4.49 (3.00–6.07)
	5	10.13 (9.07–11.30)	10.70 (9.77–11.81)	7.53 (6.43–8.70)	4.30 (3.35–6.02)