# Reconstructing Multiple Free Energy Pathways of DNA Stretching from Single Molecule Experiments

Eric W. Frey,[†] Jingqiang Li,[†] Sithara S. Wijeratne,[†] and Ching-Hwa Kiang*,[†,‡]

[†]Department of Physics and Astronomy and [‡]Department of Bioengineering, Rice University, Houston, Texas 77005, United States

ABSTRACT: Free energy landscapes provide information on the dynamics of proteins and nucleic acid folding. It has been demonstrated that such landscapes can be reconstructed from single molecule force measurement data using Jarzynski's equality, which requires only stretching data. However, when the process is reversible, the Crooks fluctuation theorem combines both stretch and relaxation force data for the analysis and can offer more rapid convergence of free energy estimates of different states. Here we demonstrate that, similar to Jarzynski's equality, the Crooks fluctuation theorem can be used to reconstruct the full free energy landscapes. In addition, when the free energy landscapes exhibit multiple folding pathways, one can use Jarzynski's equality to reconstruct individual free energy pathways if the experimental data show distinct work distributions. We applied the method to reconstruct the overstretching transition of poly(dA) to demonstrate that the nonequilibrium work theorem combined with single molecule force measurements provides a clear picture of the free energy landscapes.
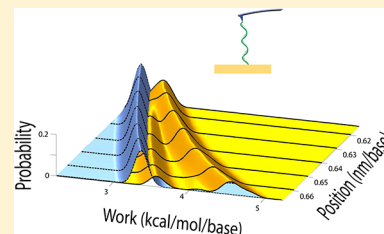
INTRODUCTION

Protein and nucleic acid folding free energy landscapes[1,2] can be reconstructed by using nonequilibrium work theorems, namely, Jarzynski's equality (JE)[3–8] and Crooks fluctuation theorem (CFT),[9–12] to analyze experimental single molecule force data and estimate the free energies. These theorems relate the work done on and/or by a system to the free energy difference between equilibrium states. These relations have been successfully used to reconstruct landscapes of the unfolding of RNA,[5] proteins,[6,13] and DNA hairpins,[7,12] using single molecule techniques such as atomic force microscopy (AFM)[14–18] and optical tweezers.[11,19,20]

It has been theoretically derived and experimentally verified that JE can be used to determine the free energy, G, as a function of molecular end-to-end extension, z, using single molecule force measurement data from pulling at a predetermined schedule, such as constant velocity.[4–6,21,22] On the contrary, the CFT requires measurement data from both forward and reverse directions, thus potentially providing a more accurate free energy estimate.[11] Similar to the JE, the CFT requires that the system starts at equilibrium for each realization. Measuring nonequilibrium work while switching a molecule from state A to state B and state B to state A allows for the determination of the free energy difference, ΔG_{AB}.[23] Similar to JE, the unique relationship between the control parameter λ (the distance that the cantilever lever traveled) and the desired parameter z (molecular end-to-end distance, Figure 1a), allows one to reconstruct the free energy curve using the constant velocity pulling single molecule force data without having to wait for the system to reach equilibrium at intermediate position M.[4,6,24] However, for CFT, each point on the free energy curve requires a pair of forward and reverse data starting at equilibrium; i.e., to determine the free energy G at any intermediate position M, one would need a pair of data A → M, where A starts at equilibrium and M → A, where M starts

at equilibrium. Here we show that one can reconstruct $G(z)$ using the CFT without waiting for the system to equilibrate at each position by deconvoluting the data M → A from A → B and B → M.

METHODS

Poly(dA) attached to AFM was prepared as described previously.[25] In brief, a silicon nitride AFM tip (Bruker) was used to pick up a single DNA molecule.[26–29] We monitored the cantilever displacement, D, from its equilibrium position to determine the cantilever restoring force, $F = -kD$, where k is the cantilever spring constant. The spring constant was determined using the equipartition theorem method[30,31] to be $k = 0.04$ N/m. We assumed the force on the molecule is balanced by the cantilever restoring force. Experiments were carried out in Tris−EDTA buffer (10 mM Tris−HCl, 1 mM EDTA, pH 8) with 150 mM NaCl. Pulling and relaxation experiments were taken at a speed of 40 nm/s, and 38 recordings were used (Figure 1). Force extension curves were aligned in the reversible regions excluding the overstretching transition. The x-axis was normalized to be displayed as nanometer per base by assuming poly(dA) is fully stretched at 600 pN with an interphosphate distance of 0.7 nm.[25,27] All recordings were averaged every two points before calculation.

Because deconvolution is sensitive to input noise, some of the deconvolved work distributions oscillated wildly or had large negative values. In these cases, the input noise prohibited an accurate estimate of the distributions; therefore, a threshold was used to exclude distributions with equal or lower values from further analysis. Results shown used a threshold of −0.1;
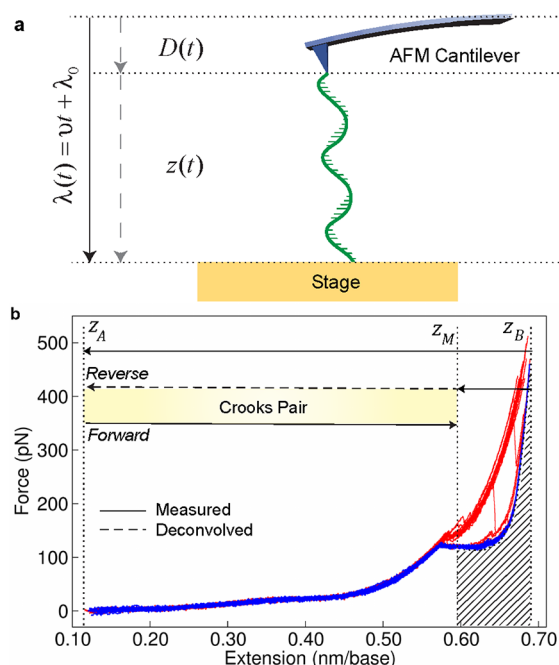
Figure 1. Single molecule AFM measurements. (a) The DNA molecule is attached to the AFM tip and the substrate. The distance between the cantilever lever and the substrate is increased at a constant velocity v. (b) Representative force extension curves of poly(dA) stretch/relaxation cycles. The stretching process started from an equilibrium state λ_A, and the relaxation process started from an equilibrium state λ_B. The stretch curves showed two pathways during the overstretching transitions, and the relaxation almost exclusively follows the low energy pathways. At 40 nm/s pulling speed, 60% of the curves go through the high energy pathway. The relative distribution between these two pathways depends on the pulling speed.[25] Data were averaged for every 60 points for display purpose.

lowering the threshold to include more distributions increased the noise in G but did not change its shape. Similarly, using a larger bin size in work histograms increased artifacts, and a smaller bin size reduced the number of acceptable deconvolved distributions and increased spurious noise spikes, but otherwise, the choice of bin size did not affect G. To determine where work distributions cross and thus estimate ΔG, in addition to the graphical method described in the main text, we also used the Bennett acceptance ratio method,[11] which gave similar results.

■ RESULTS AND DISCUSSION

We demonstrate our approach on poly(dA), a homopolymeric single-stranded DNA with unique thermodynamic properties.[32] At equilibrium, poly(dA) forms a single-stranded helix due to the strong stacking interactions among adenines.[33] The stretching force curve does not show any obvious peak, i.e., an abrupt force drop, and therefore, the process is reversible.[25,27] Hence we chose this system for free energy reconstruction using CFT.

We used the AFM tip to pick up a poly(dA) molecule, and repeatedly stretched (forward, F process) and relaxed (reverse, R process) the molecule. The position, λ, of the piezo-controlled substrate stage was changed at constant speed, v (Figure 1a). The $\lambda(t)$ and the cantilever deflection D are recorded and converted to the extension, $z = \lambda - D$. The resulting force extension curve, $f(z)$, is displayed in Figure 1b.

At the end of either stretch or relaxation processes, the λ was kept constant to allow the system to reach equilibrium before the direction was turned. We waited 70 s, which has been determined to be sufficient to allow the system to reach equilibrium, before turning around for the relaxation process.

The CFT stated that[9,11]

$$\frac{P_F(W)}{P_R(-W)} = \exp(\beta[W - \Delta G]) \tag{1}$$

where $P_F(W)$ is the probability of work distribution W during the forward process, $P_R(-W)$ is the probability of work distribution during the reverse process, $\beta = 1/k_BT$, k_B is Boltzmann constant and T is temperature, and ΔG is the free energy change for switching the molecule from A to B.

As mentioned above, we chose poly(dA), where the forward force curves are predominantly uphill, with no significant free energy barriers like that of protein unfolding, to do the stretch/relaxation cycles. Because the CFT requires the system to start at an equilibrium state but not necessarily end with an equilibrated state, at the end of the forward process the system does not need to be equilibrated, although we do need to allow the system to reach equilibrium before starting the reverse process. Thus, we have data for A → B and B → A. From the data, the stretching from the initial state A to any intermediate position M can be calculated. Similarly, for the reverse process, the data from the initial state B to any intermediate position M can be obtained. However, to apply CFT to obtain the G of any intermediate position M, we need a forward and reverse pair of data from A to M. Because the reverse data from M to A were not measured, we applied deconvolution using the reverse data from B to A and B to M,

$$P_F(W)_{A \to B} = P_F(W)_{A \to M} \star P_F(W)_{M \to B}$$

$$P_R(-W)_{B \to A} = P_R(-W)_{B \to M} \star P_R(-W)_{M \to A} \tag{2}$$

where $\star$ indicates convolution. $P_R(-W)_{B \to A}$ and $P_R(-W)_{B \to M}$ are the work distribution for system to go from B → A and B → M, respectively. In each equation, only two of the three distributions were measured. The third distribution were deconvolved using the discrete convolution theorem,[34]

$$\mathcal{F}\{P_F(-W)_{B \to A}\} = \mathcal{F}\{P_F(-W)_{B \to M}\} \cdot \mathcal{F}\{P_F(-W)_{M \to A}\} \tag{3}$$

where $\mathcal{F}$ represents the Fourier transform. The reverse data between A and M can be solved by

$$P_F(-W)_{M \to A} = \mathcal{F}^{-1}\left\{ \frac{\mathcal{F}\{P_F(-W)_{B \to A}\}}{\mathcal{F}\{P_F(-W)_{B \to M}\}} \right\} \tag{4}$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform.

Figure 2 shows the measured forward distribution, $P_F(W)_{A \to M}$, and the reverse distribution, $P_R(-W)_{M \to A}$ calculated using eq 2. As indicated in eq 1, when $P_F(W) = P_R(-W)$, $W = \Delta G$; i.e., the intersection of the forward and reversion work distributions allows us to determine $\Delta G = G_M - G_A$.[11] Note that the forward work values show bimodal distributions, indicating additional stretching pathways. We have also calculated the distributions using the measured R and F from deconvolution, and the averaged results from both methods, G_{CFT}, are shown in Figure 3a. To show the validity of the CFT process, we used JE to reconstruct $G(z)$ from the same force data.[4,6,35] The forward G_{JE}^F and reverse G_{JE}^R were reconstructed
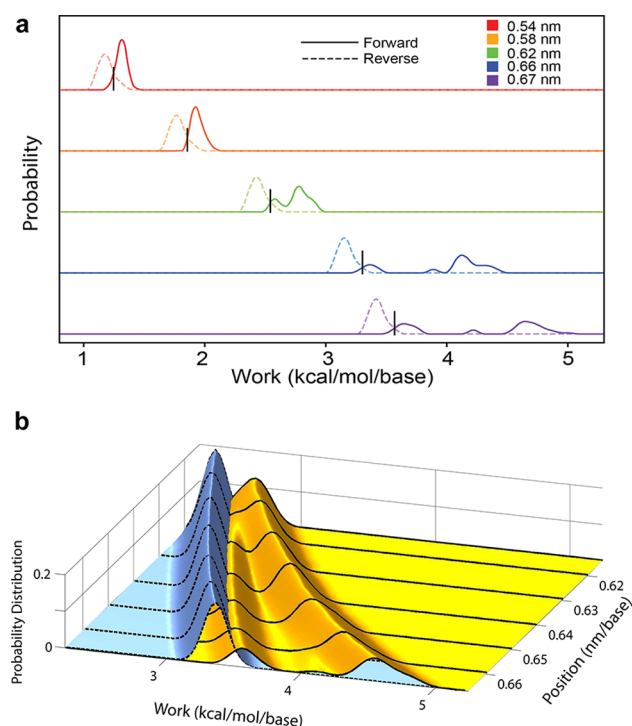
Figure 2. CFT free energy surface reconstruction. (a) Work distributions of poly(dA) at various positions. The black lines indicate the crossover between $P_F(W)$ and $P_R(-W)$ for a given position. (b) 3D representation showing the evolution of the crossover as a function of positions.
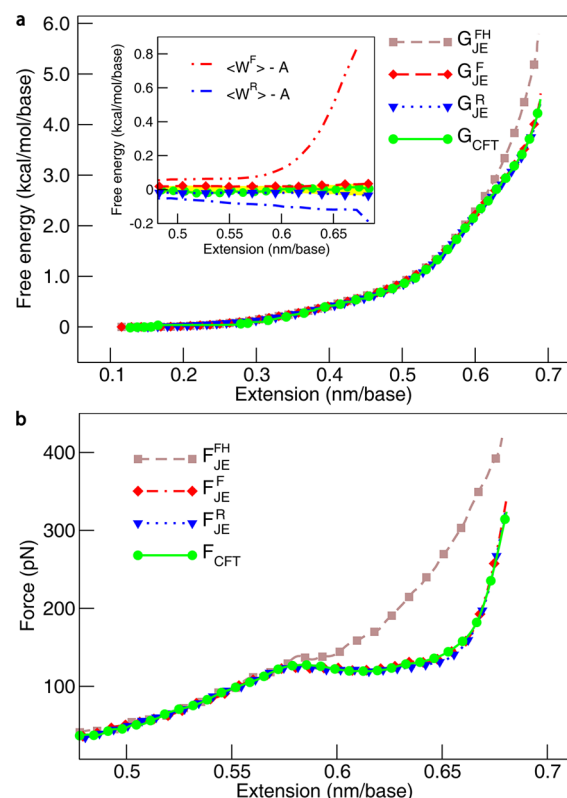

Figure 3. Reconstruction of free energy pathways. (a) Free energy curve reconstructions using the CFT and JE methods. Inset: agreement between CFT and JE methods. Results from forward, reverse, and average of the two works using JE, $A = (G_{JE}^F + G_{JE}^R)/2$, are shown for comparison. (b) Equilibrium force curves using JE and CFT recovered with $F = dG/dx$. The results from CFT, G_{CFT}, are consistent with those calculated from JE. The equilibrium force curve from the metastable pathways, G_{JE}^{FH}, was determined by averaging only the work in the high value peaks in the work distributions of the forward direction.

separately. Figure 3a shows our approach agrees with the JE-based method.

In principle, Jarzynski's equality is used to recover the equilibrium free energies, when metastable states that are higher in free energies exist and the higher free energies are averaged out due to their smaller weight. However, in the current studies, the metastable state work distribution is distinct from the ground state distribution. Therefore, we were able to manually separate these two distributions and reconstruct them individually to obtain the equilibrium free energy and metastable state energy separately. As shown in Figure 2, the resulting free energies from CFT were only affected by the position where the distributions cross and, therefore, were not biased by the additional work distributions at higher work values. When JE is used, the average is heavily overweighted by small work values, and the additional work that is far from equilibrium contributes little to the final average, and the results are satisfactory. When the weighted average is performed using Jarzynski's equality, only the equilibrium, i.e., the lowest free energy state, will be determined. However, in current studies, there were two distinct peaks in the work distributions, which allow us to manually separate the two distributions and obtain the average of each distribution using Jarzynski's equality. Figure 3a shows that, when JE is used to estimate free energies, the work from stretching overestimated G, whereas the work from relaxation underestimated G (Figure 3a). The derivative of the free energy G, dG/dz, shows two distinct averaged force−extension curves (Figure 3b). In addition to direct application in CFT, the reverse process data have also been used to accelerate the convergence of the free energies.[36] Both methods should increase the accuracy of the determined free energies. It will be interesting to compare quantitatively the

improvement on uncertainty by incorporating the reverse data using different methods.

■ CONCLUSIONS

In conclusion, in addition to obtaining the free energy difference between two states, CFT can be used to reconstruct the entire free energy curve, using the deconvolution method. JE can be used to determine both the equilibrium and the metastable free energy pathways, when the two energies are far apart so that the work distribution shows distinct bimodal distributions and is separable. The choice of free energy of overstretching transitions of DNA, which does not exhibit unfolding peaks that result in an abrupt force decrease, enabled the use of CFT, which requires the work distributions from reverse reactions. The overall free energy landscape of overstretching transitions of poly(dA) is depicted in Figure 4. The results from CFT and JE are consistent, suggesting the applicability using the nonequilibrium work theorem on single molecule force data to reconstruct the free energy landscapes. The unique energy landscapes of DNA overstretching transitions are likely due to different base-stacking and random coil states,[25] and the reconstruction of free energy pathways provides clues for mapping the dynamics of poly(dA) in conformational spaces.
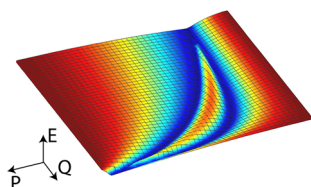
Figure 4. Free energy landscape of poly(dA). The overstretching transitions of poly(dA) showed two distinct pathways. The equilibrium pathway can be reconstructed using either JE or CFT, whereas the metastable, high energy pathway was reconstructed using JE. The axes, P, Q, and E are analogous to entropy, the fraction of native contacts, and energy.[2]

■ AUTHOR INFORMATION

Corresponding Author
*C.-H. Kiang. E-mail: chkiang@rice.edu. Phone: +1 713 3484130. Fax: +1 713 3484150 .

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.

(2) Dobson, C. M.; Sali, A.; Karplus, M. Protein folding: a perspective from theory and experiment. *Angew. Chem., Int. Ed.* **1998**, *37*, 868−893.

(3) Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.

(4) Hummer, G.; Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 3658−3661.

(5) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski'equality. *Science* **2002**, *296*, 1832−1835.

(6) Harris, N. C.; Song, Y.; Kiang, C.-H. Experimental Free Energy Surface Reconstruction From Single-Molecule Force Spectroscopy Using Jarzynski's Equality. *Phys. Rev. Lett.* **2007**, *99*, 068101.

(7) Gupta, A. N.; Vincent, A.; Neupane, K.; Yu, H.; Wang, F.; Woodside, M. T. Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nat. Phys.* **2011**, *7*, 631−634.

(8) Botello, E.; Harris, N. C.; Sargent, J.; Chen, W.-H.; Lin, K.-J.; Kiang, C.-H. Temperature and chemical denaturant dependence of forced unfolding of titin I27. *J. Phys. Chem. B* **2009**, *113*, 10845−10848.

(9) Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **1999**, *60*, 2721−2726.

(10) Paul Maragakis, M. S.; Karplus, M. A differential fluctuation theorem. *J. Phys. Chem. B* **2008**, *112*, 6168−6174.

(11) Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* **2005**, *437*, 231−234.

(12) Alemany, A.; Mossa, A.; Junier, I.; Ritort, F. Experimental free-energy measurements of kinetic molecular states using fluctuation theorems. *Nat. Phys.* **2012**, *8*, 688−694.

(13) Nome, R. A.; Zhao, J. M.; Hoff, W. D.; Scherer, N. F. Axis-dependent anisotropy in protein unfolding from integrated non-equilibrium single-molecule experiments, analysis, and simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20799−20804.

(14) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **1997**, *276*, 1109−1112.

(15) Rief, M.; Fernandez, J. M.; Gaub, H. E. Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys. Rev. Lett.* **1998**, *81*, 4764−4767.

(16) Evans, E. A.; Calderwood, D. A. Forces and bond dynamics in cell adhesion. *Science* **2007**, *316*, 1148−1153.

(17) Muller, D. J.; Dufrene, Y. F. Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. *Nat. Nanotechnol.* **2008**, *3*, 261−269.

(18) Lindsay, S. M. *Introduction to Nanoscience*; Oxford University Press: Oxford, U.K., 2009.

(19) Gore, J.; Brant, Z.; Nollmann, M.; Le, M. U.; Cozzarelli, N. R.; Bustamante, C. DNA overwinds when stretched. *Nature* **2006**, *442*, 1836−839.

(20) Woodside, M. T.; Anthony, P. C.; Behnke-Parks, W. M.; Larizade, K.; Herschlag, D.; Block, S. M. Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid. *Science* **2006**, *314*, 1001−1004.

(21) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **2003**, *119*, 3559−3566.

(22) West, D. K.; Olmsted, P. D.; Paci, E. Free energy for protein folding from nonequilibrium simulations using the Jarzynski equality. *J. Chem. Phys.* **2006**, *125*, 204910.

(23) Mossa, Alessandro; H, J. M.; Lorenzo, Sara de; Ritort, F. Measurement of work in single-molecule pulling experiments. *J. Chem. Phys.* **2009**, *130*, 234116−234125.

(24) Jarzynski, C. Work Fluctuation Theorems and Single-Molecule Biophysics. *Prog. Theor. Phys. Suppl.* **2006**, *165*, 1−17.

(25) Chen, W.-S.; Chen, W.-H.; Chen, Z.; Gooding, A. A.; Lin, K.-J.; Kiang, C.-H. Direct observation of multiple pathways of single-stranded DNA stretching. *Phys. Rev. Lett.* **2010**, *105*, 218104.

(26) Calderon, C. P.; Chen, W.-H.; Lin, K.-J.; Harris, N. C.; Kiang, C.-H. Quantifying DNA melting transitions using single-molecule force spectroscopy. *J. Phys.: Condens. Matter.* **2009**, *21*, 034114.

(27) Ke, C.; Humeniuk, M.; S-Gracz, H.; Marszalek, P. E. Direct measurements of base stacking interactions in DNA by single-molecule atomic-force spectroscopy. *Phys. Rev. Lett.* **2007**, *99*, 018302.

(28) Clausen-Schaumann, H.; Rief, M.; Tolksdorf, C.; Gaub, H. E. Mechanical stability of single DNA molecules. *Biophys. J.* **2000**, *78*, 1997−2007.

(29) Rief, M.; Clausen-Schaumann, H.; Gaub, H. E. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.* **1999**, *6*, 346−349.

(30) Hutter, J. L.; Bechhoefer, J. Calibration of atomic-force microscopy tips. *Rev. Sci. Instrum.* **1993**, *64*, 1868−1873.

(31) Butt, H.-J.; Jaschke, M. Calculation of thermal noise in atomic force microscopy. *Nanotechnology* **1995**, *6*, 1−7.

(32) Scholl, Z. N.; Rabbi, M.; Lee, D.; Manson, L.; S-Gracz, H.; Marszalek, P. E. Origin of Overstretching Transitions in Single-Stranded Nucleic Acids. *Phys. Rev. Lett.* **2013**, *111*, 188302.

(33) Chakraborty, S.; Sharma, S.; Maiti, P. K.; Krishnan, Y. The poly dA helix: a new structural motif for high performance DNA-based molecular switches. *Nucleic Acids Res.* **2009**, *9*, 2810−2817.

(34) Press, W. H., Teukolsky, S. A., Vettering, W. T., Flannery, B. P. *Numerical Recipes in Fortran 77*, 2nd ed.;Cambridge University Press: Cambridge, U.K., 2003; Chapter 13, pp 530−537.

(35) Harris, N. C.; Kiang, C.-H. Velocity convergence of free energy surfaces from single-molecule measurements using Jarzynski's equality. *Phys. Rev. E* **2009**, *79*, 041912.

(36) Minh, D. D. L.; Adib, A. B. Optimized Free Energies from Bidirectional Single-Molecule Force Spectroscopy. *Phys. Rev. Lett.* **2008**, *100*, 180602.