RICE UNIVERSITY

Experimental Free Energy Landscape Reconstruction of DNA Unstacking Using Crooks Fluctuation Theorem
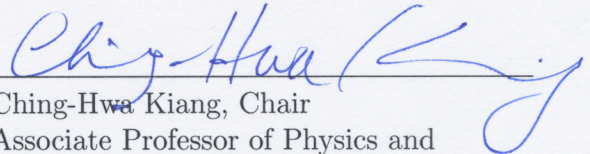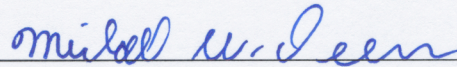
by

Eric W. Frey

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
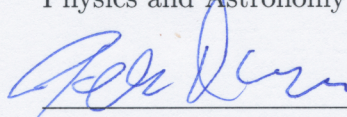
Master of Science

APPROVED, THESIS COMMITTEE:

Ching-Hwa Kiang, Chair
Associate Professor of Physics and
Astronomy and Bioengineering

Michael W. Deem
John W. Cox Professor in Biochemical
and Genetic Engineering and Professor of
Physics and Astronomy

Peter J. Nordlander
Professor of Physics and Astronomy and
Electrical and Computer Engineering

Houston, Texas

April, 2012

ABSTRACT

Experimental Free Energy Landscape Reconstruction of DNA Unstacking Using

Crooks Fluctuation Theorem

by

Eric W. Frey

Nonequilibrium work theorems, such as the Jarzynski equality and the Crooks fluctuation theorem, allow one to use nonequilibrium measurements to determine equilibrium free energies. For example, it has been demonstrated that the Crooks fluctuation theorem can be used to determine RNA folding energies. We used single-molecule manipulation with an atomic force microscope to measure the work done on poly(dA) as it was stretched and relaxed. This single-stranded nucleic acid exhibits unique base-stacking transitions in its force-extension curve due to the strong interactions among A bases, as well as multiple pathways. Here we showed that free energy curves can be determined by using the Crooks fluctuation theorem. The nonequilibrium work theorem can be used to determine free energy curves even when there are multiple pathways.

Acknowledgments

Contents

Figures

Chapter 1

Background and Motivation: Understanding the Physics of DNA Using Nanoscale Single-Molecule Manipulation

1.1 Introduction

DNA is the carrier of genetic information and is involved in biomolecular processes such as transcription and replication. Many of these processes are governed by the mechanics and thermodynamics of bending, stretching, twisting, and unzipping the double helix [8, 9, 10, 11, 12, 13]. Double-stranded DNA (dsDNA) is a semi-flexible polymer, with its base-stacking architecture and negative charges along its phosphate backbone. In physiological conditions, thermal fluctuations do not bend it significantly on length scales below 50 nm, which is equivalent to 150 base pairs (bp) [14]. The 10 μm-long DNA of a viral genome can be packed inside a capsid of 50 nm in diameter [4, 15, 16], and in eukaryotic cells, histones bend DNA into loops of 10 nm in diameter. The latter serves as the first step in the hierarchical packaging of the genome in eukaryotes (Fig. 1.1), and it regulates gene expression by obstructing access to base pairs [1]. Histones, helicases, topoisomerases, and RNA and DNA polymerases are examples of proteins that generate or relieve tension and torque in DNA to enable its biochemical functions [17, 12, 18, 19, 20]. With advances in single-molecule techniques, it has been possible to examine the physics of DNA directly. By providing control and measurement of force of a single molecule, these techniques

have revealed a variety of DNA conformations and much of DNA's complex behavior.

1.2   Single-Molecule Manipulation Experiments

Single-molecule manipulation techniques using atomic force microscopy (AFM), optical tweezers, and magnetic tweezers are illustrated in Fig. 1.2. These techniques have been used to manipulate a variety of biological molecules. In each of these methods, a single DNA molecule is attached between a substrate and a force probe, either an AFM tip or a micron-sized bead, in an aqueous solution. The change in molecular end-to-end extension is determined from the change in probe and substrate positions. The force on the molecule is determined from displacement of the probe relative to its equilibrium position. Nonspecific attachment, typically used in AFM, is achieved by adsorption of DNA to the substrate surface or the probe surface. Specific attachment, employed by optical and magnetic tweezers, is achieved by functionalization of probe and substrate surfaces. These modifications exploit the high affinity of binding among ligand-receptor, antibody-antigen pairs and DNA hybridization. Other techniques elongate DNA by confining the molecule within micro or nano-sized obstacles. Such techniques include driving DNA electrophoretically through microlithographic arrays [21], nanochannels [22, 23], and nanopores [24].

In AFM (Fig. 1.2a), the force probe is an AFM tip attached to a cantilever, and the solid substrate surface is mounted on a piezoelectric scanner. Moving the substrate toward the AFM tip allows nonspecific or specific molecular attachment between the substrate and cantilever. When the molecule is attached to the tip and the substrate, moving the substrate away from the cantilever produces force on the attached molecule, bending the cantilever. The force exerted on the molecule is determined
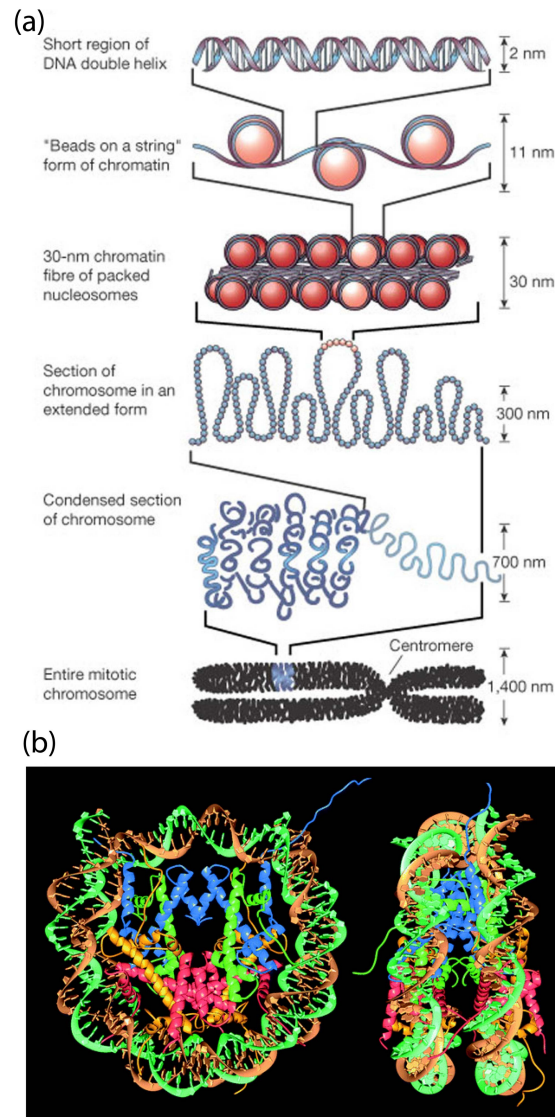
Figure 1.1 : (a) Hierarchical organization of DNA packaged into a chromosome. The nucleosomes are formed by histones which bend DNA into small loops. From [1]. (b) Structure of the nucleosome. A central octamer of histone proteins wraps two superhelical turns of the DNA double helix by hydrogen bonds and electrostatic interactions. From [2].
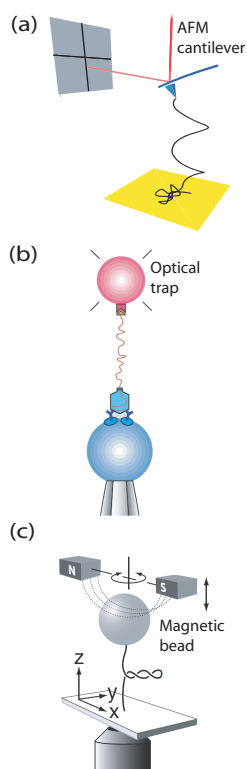
Figure 1.2 : Illustrations of single-molecule manipulation techniques. (a) AFM. The molecule is held by the tip and the substrate surface. The force on the attached molecule is determined based on the displacement of the cantilever. From [3]. (b) Optical tweezers. One end of a DNA molecule is attached to a bead trapped by a laser beam, while the other end is attached to a DNA-virus capsid complex on a second bead, held by a micropipette tip. From [4]. (c) Magnetic tweezers. Force is exerted on the molecule by an attached super-paramagnetic bead in a magnetic field. The molecule can be twisted as well as stretched by the applied field. Adapted from [5].

by Hooke's law, $F = kD$, where k is the cantilever's spring constant and D is the cantilever displacement from its equilibrium position. The displacement is detected by the deflection of the laser beam bouncing off the back of the cantilever. Using the equipartition theorem, the spring constant is determined using $\frac{1}{2}k_B T = \frac{1}{2}k\langle D^2 \rangle$, where k_B is Boltzmann's constant and T is temperature [25]. AFM cantilevers used for single-molecule manipulation typically have a spring constant $k = 10$ pN/nm or higher. This results in unloaded cantilever fluctuations of at least 5 pN at room temperature, which sets the limit of the noise level in the force on an attached molecule measurable by AFM. AFM is able to measure high forces up to a few nanoNewtons, the limit usually being set by the strength of the attachment [26].

In a typical optical tweezers setup (Fig. 1.2b), the force probe is a micrometer-sized dielectric bead captured in an optical trap. The substrate may be the side of a translatable fluid chamber, or a second bead held by micropipette or a second optical trap [17, 11, 19]. The optical trap consists of a tightly-focused laser, which exerts a three-dimensional restoring force on a dielectric bead trapped near the laser focus. To minimize photodamage to the trapped biomolecules, near-infrared wavelengths are used [27]. The displacement of the bead from the trap center can be measured by video tracking via an optical microscope. For small displacements of the bead, the force is determined using Hooke's law, and the trap stiffness using the equipartition theorem, as in the case of AFM. Optical traps typically have spring constants ranging from 0.005–1 pN/nm, which is softer than the AFM cantilevers. The low noise level allows measurement of forces on the molecule as low as 0.1 pN. Optical tweezers are generally used to probe forces less than 100 pN, where the ligand-receptor or antibody-antigen pairs used to attach the DNA unbind [28].

Magnetic tweezers (Fig. 1.2c) are similar to optical tweezers, except that the

force probe consists of a super-paramagnetic bead in an applied magnetic field. The force on the bead is proportional to the gradient of the square of the magnetic field. In addition, a torque is applied to the bead due to its small magnetic polarization anisotropy, which tends to align the bead with the applied magnetic field. Thus, by rotating the applied field, the attached molecule can be twisted as well as stretched. Magnetic tweezers have miniscule stiffness as low as 10^{-6} pN/nm, allowing them to probe forces as low as 10^{-3} pN. Like optical tweezers, they can probe up to 100 pN until the DNA handles break [28, 29, 30].

1.3   Polymer Physics Models of DNA

Single-molecule manipulation experiments measure the force-extension curve of DNA, and the data are fitted to polymer physics models to determine parameters that define its mechanical properties (Fig. 1.3). In solution, DNA adopts a random coil conformation which minimizes free energy. Extending the molecule imposes a constraint limiting the number of accessible conformations, thus the work done on the molecule is mainly used to offset the reduced entropy. For dsDNA, at forces less than 10 pN the force-extension curve is dominated by this entropic elasticity. At higher forces, dsDNA begins to exceed its contour length and, consequently, its double-helix structure is disrupted. The polymer elasticity models which best describe the force-extension curves of single-stranded DNA (ssDNA) and dsDNA are the freely-jointed chain (FJC) and worm-like chain (WLC) models, respectively.

In the FJC model, the polymer consists of a chain of freely rotating segments of characteristic Kuhn length. The extensible FJC assumes the polymer is stretchable,
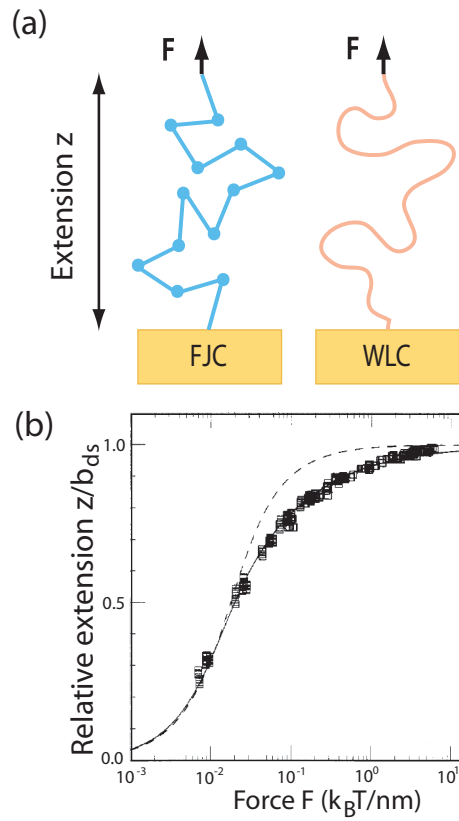
Figure 1.3 : The polymer physics models that describe DNA. (a) Illustration of the FJC and WLC models. (b) Force-extension behavior of a single dsDNA molecule. dsDNA can be described accurately by the WLC model (solid curve), but not the FJC model (dashed curve). Adapted from [6].

and the force is related to extension z by [28, 31]

$$z = b_{ss} \left[\coth(2\beta P_{ss} F) - \frac{1}{2\beta P_{ss} F} \right] \left[1 + \frac{F}{K_{ss}} \right] \tag{1.1}$$

where P_{ss}, b_{ss}, and K_{ss} are the persistence length, contour length, and stretch modulus of ssDNA, respectively, and $\beta = 1/k_B T$. The persistence length is a measure of bending stiffness. The K_{ss} accounts for the extensibility of the molecule. For ssDNA, $P_{ss} = 0.75$ nm and $K_{ss} = 800$ pN [28, 32, 33, 7].

The WLC models a polymer as a flexible rod characterized by a bending stiffness. In an extensible WLC model, force can be related to extension by [34, 14, 31]

$$z = b_{ds} \left[1 - \frac{1}{\sqrt{4\beta P_{ds} F}} + \frac{F}{K_{ds}} \right] \tag{1.2}$$

where P_{ds}, b_{ds}, and K_{ds} are the persistence length, contour length, and elastic stretch modulus of dsDNA, respectively. For dsDNA, $P_{ds} = 50$ nm and $K_{ds} = 1200$ pN [28, 32, 33, 7].

1.4 The Overstretching Transitions and Force-Induced Melting

Fig. 1.4a is a typical force-extension curve of dsDNA. At low forces, the curve can be fitted to the WLC model [34, 14, 31, 7]. When the force reaches 65 pN, the force-extension curve shows a plateau, indicating a cooperative transition of B-DNA to S-DNA. At forces around 150 pN, dsDNA melts into ssDNA [26, 11, 32, 35, 36, 7], where the force-extension curve is best described by the extensible FJC model (Eq. 1.1) with a persistence length and stretch modulus consistent with ssDNA [28, 32, 33]. Another example is poly(dA) (ssDNA composed only of adenine bases), where distinct plateaus and multiple force-extension pathways have been observed [37, 3]. One
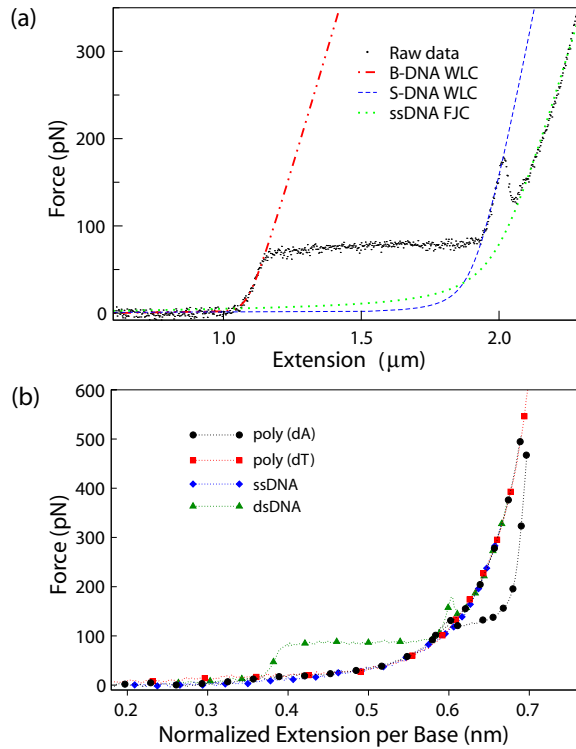
Figure 1.4 : Force-induced transitions of DNA. (a) Force-extension data showing stretching, melting, and overstretching of a λ-DNA. The data are fit to extensible WLC and FJC models. From [7]. (b) Force-extension pathways for poly(dA) compared to dsDNA and other ssDNA. From [3].

poly(dA) pathway is similar to that of random-sequence ssDNA, whereas the other pathway has an additional, energetically favored transition (Fig. 1.4b). The multiple pathways suggest that poly(dA) has two conformational states when stretched almost twice its contour length.

Pulling single DNA molecules has been found to unzip as well as stretch DNA. Unzipping occurs when the secondary structure, i.e. the double helix of dsDNA is disrupted, resulting in unpairing of the bases. The dynamics of unzipping are sequence-dependent, as evidenced by higher observed forces in GC-rich regions [38] and good reproducibility for unzipping/rezipping molecules of the same sequence

[39, 40].

1.5 Conclusions

The active processes of life, including the packaging, recombination, transcription, and replication of the information stored in DNA, typically involve the deformation of DNA from its equilibrium structures such as bending, stretching, twisting, and unzipping of the double helix. Single-molecule manipulation techniques have made it possible to control DNA conformation and simultaneously detect the induced changes, revealing a rich variety of mechanically-induced conformational changes and thermo-dynamic states. These single-molecule techniques helped us to reveal the physics of DNA and the processes involved in the passing on of the genetic code.

Chapter 2

Using the Crooks fluctuation Theorem to Determine the Free Energy Profile of Overstretching Single-Stranded DNA through Multiple Pathways

2.1   Introduction

Advances in statistical physics, namely the Jarzynski equality (JE) [41] and the Crooks fluctuation theorem (CFT) [42], have made it possible to obtain equilibrium information from nonequilibrium experiments. These equations relate the fluctuations in work done on a system repeatedly driven from equilibrium to the free energy difference between equilibrium states. Meanwhile, experimental techniques such as atomic force microscopy (AFM) and optical tweezers have enabled control and measurement of the force on a single molecule as it is stretched or unfolded, a process which typically drives the system out of equilibrium. The challenge of applying the JE and CFT to single experiments in order to recover free energies has become an area of great interest [43, 44].

One remarkable property of the JE is that it can be used to determine the equilibrium free energy profile of a molecule, $G(z)$ as a function of end-to-end extension z, without waiting for equilibration at each step along z [45, 44]. Instead, $G(z)$ is determined by repeated, nonequilibrium work measurements using a predetermined protocol, e.g. moving an AFM cantilever or optical trap from position $\lambda_A \rightarrow \lambda_B$

at constant velocity. In some cases, this can be done with a modest number of repetitions, and with fast protocols which drive the molecule far from equilibrium [46, 47, 44]. One might expect the same to be true of the CFT, given its close relationship to the JE [48].

For a repeated thermodynamic process carried out in forward (F) and reverse (R) directions, the CFT predicts a relationship between the work distributions $P(W)$ and the change in free energy:

$$\frac{P_F(W)}{P_R(-W)} = exp(\beta[W - \Delta G_\lambda]) \tag{2.1}$$

where $\beta = 1/k_B T$, k_B is Boltzmann's constant and T is temperature. Here, the system starts in equilibrium for each repetition of the process, which is allowed to drive the system arbitrarily far from equilibrium. The process is characterized by control parameter λ which is varied by $\lambda_A \to \lambda_B$ in the F direction, and the time-reversed process $\lambda_B \to \lambda_A$ in the R direction. ΔG_λ is the free energy difference between equilibrium states at λ_A and λ_B. In order to reconstruct the free energy *profile*, free energy differences must be determined across all intermediate positions λ_M, where $\lambda_A < \lambda_M < \lambda_B$. This ostensibly requires many sets of experiments in order to measure work distributions corresponding to $\lambda_M \to \lambda_B$ or $\lambda_M \to \lambda_A$, where the molecule must be allowed to equilibrate at every λ_M before pulling. Thus it seems that the CFT would be more cumbersome in practice than the JE.

Here we derive a deconvolution approach to construct $G(z)$ from nonequilibrium single-molecule measurements using the CFT, without waiting for equilibration at intermediate positions. The basis of our approach is to relate the measured work distributions for pulling from the end positions, to the unknown work distributions for pulling from intermediate positions, by convolution. The work distributions from all positions are then related to free energy differences using the CFT. We also show

how to determine free energy along z, even though Eq. (2.1) is written in terms of λ, and z is not directly controlled in single-molecule experiments [Fig. 2.1(a)]. Furthermore, we show that both the CFT and JE pick out the equilibrium pathway from nonequilibrium data involving multiple pathways.

2.2 Experimental Methods

We demonstrate our approach by studying the stretching and relaxation of poly(dA) using AFM. Sample preparation and data collection was as previously described [3]. A silicon nitride AFM tip (Bruker) with cantilever spring constant $k = 0.04$ N/m was used to pick up a single molecule [7, 37, 32, 26], which was repeatedly stretched (F process) or relaxed (R process) by changing the position λ of the piezo-controlled substrate stage at constant velocity ν [Fig. 2.1(a)]. We monitored cantilever displacement D from its equilibrium position to measure the spring-like restoring force on the molecule $F = kD$, and the extension $z = \lambda - D$, to obtain force-extension curves [FECs, Fig. 2.1(b,c)]. After stretching or relaxing, the system was kept at fixed λ to allow equilibration, which at high extensions was characterized by a sudden drop in force [3].

A total of 110 FECs were measured at two pulling rates ($\nu = 40$ and 250 nm/s). At forces above 114 pN, an overstretching transition is observed with multiple pathways [37, 3], which defines our region of interest. To remove instrument drift, FECs were aligned in the reversible regions of the curves. The z-axis was normalized assuming poly(dA) is fully stretched at 600 pN with an interphosphate distance of 0.7 nm [37, 3]. The FECs in Fig. 2.1(b,c) show the expected behavior, including a low-force, reversible pathway and a high-force, irreversible pathway, with occasional sudden transitions from the higher to lower pathway.
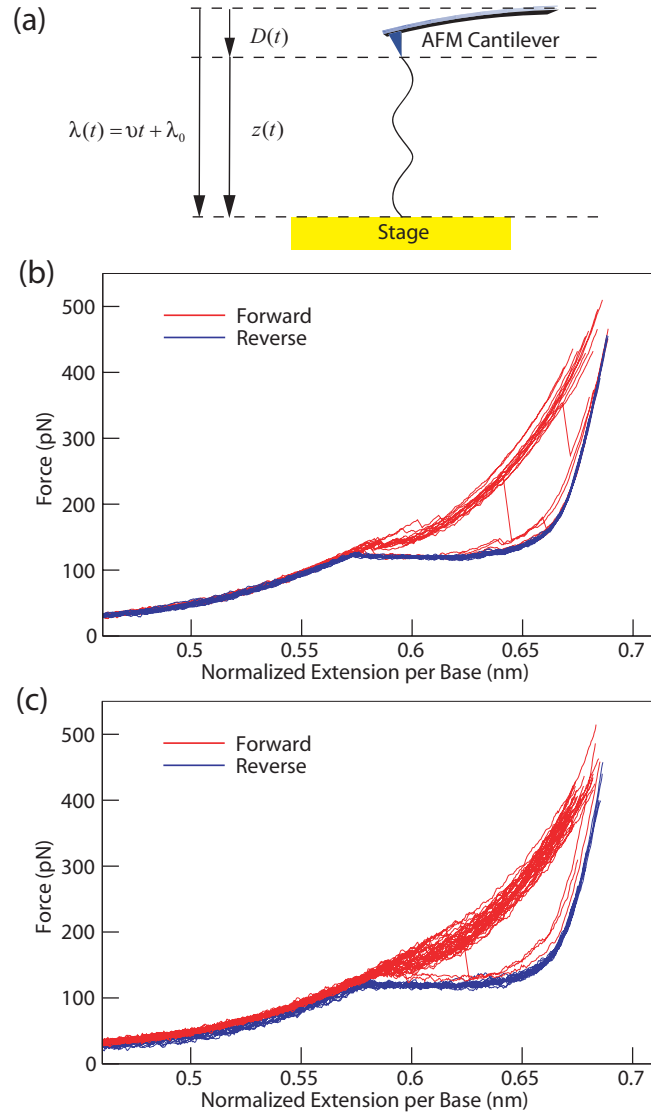
Figure 2.1 : AFM measurements. (a) Schematic representation of poly(dA) pulled by AFM. $\lambda(t)$ is the position of the substrate relative to the cantilever equilibrium position, and it is moved at speed ν. $D(t)$ is cantilever displacement, and $z(t)$ is molecular end-to-end extension. (b) FECs for pulling in the forward (red) and reverse (blue) directions at pulling speed $\nu = 40$ nm/s. (c) FECs for $\nu = 250$ nm/s. Curves shown were averaged for display purposes.

2.3 Results and Discussion

When applying the CFT to single-molecule experiments, it is important to realize that the work in Eq. (2.1) is $W = \int F d\lambda$, i.e. the work done on the entire *molecule + AFM cantilever* system [49]. Consequently, the free energy in Eq. (2.1) is the reversible work done on this combined system, in terms of λ. In order to derive a relation similar to Eq. (2.1) in terms of $G(z)$, we start with the Crooks path ensemble equation, from which both the JE and CFT can be derived [48]:

$$\langle \mathcal{F} \rangle_F = \langle \hat{\mathcal{F}} exp(-\beta[W + \Delta G_\lambda]) \rangle_R. \tag{2.2}$$

This equation relates work and free energy for the same repeated processes of λ described above, for the CFT. The brackets $\langle ... \rangle_F$ and $\langle ... \rangle_R$ represent averages of the paths taken by the system during the F process evaluated over forward paths $[x]$, and the time-reversed process R evaluated over reverse paths $[\hat{x}]$, respectively. $\mathcal{F}$ represents an arbitrary functional of the path, with time-reversal $\hat{\mathcal{F}}$, where $\mathcal{F}[x] = \hat{\mathcal{F}}[\hat{x}]$. Consistent with our above notation and following the convention of Ref. [48], the "delta" is defined in terms of the forward process, i.e. $\Delta G_\lambda \equiv G_{\lambda_B} - G_{\lambda_A}$.

Let us assume that at the beginning $(t = 0)$ and end $(t = \tau)$ of the process, the system has a well-defined z. That is, for the F process $z(0) = z_A$ and $z(\tau) = z_B$; for the R process $z(0) = z_B$ and $z(\tau) = z_A$. This assumption does not preclude nonequilibrium behavior, and it is satisfied by (i) waiting for the system to equilibrate after each pulling process, which is already required when using the JE or CFT to ensure the next measurement begins in equilibrium; and (ii) having a tight distribution of equilibrium z at the ends of the process, which is a good approximation in our experiments. It follows that the reversible work for switching the *molecule + cantilever* system from $\lambda_A \rightarrow \lambda_B$ is equal to the reversible work for switching the

molecule from $z_A \rightarrow z_B$, plus the reversible work for switching the *cantilever* from $(z_A, \lambda_A) \rightarrow (z_B, \lambda_B)$. That is,

$$\Delta G_\lambda = \Delta G_z + V(z_B, \lambda_B) - V(z_A, \lambda_A) \tag{2.3}$$

where $V(z, \lambda)$ is the energy stored in the cantilever, and ΔG_z is the reversible work for switching the bare molecule between equilibrium states at extensions z_A and z_B. The work can be written as $W = W_z + V[z(\tau), \lambda(\tau)] - V[z(0), \lambda(0)]$, where $W_z \equiv \int_0^\tau F dz$ is the "transferred" work done on the molecule [50, 49]. Plugging Eq. (2.3) and this expression for work into Eq. (2.2), and noting that inside the brackets $\langle ... \rangle_R$ we have $\lambda(0) = \lambda_B$, $z(0) = z_B$, $\lambda(\tau) = \lambda_A$, and $z(\tau) = z_A$, gives

$$\langle \mathcal{F} \rangle_F = \langle \hat{\mathcal{F}} exp(-\beta W_z[\hat{x}] + \Delta G_z) \rangle_R. \tag{2.4}$$

For clarity in the next step, we have written the work explicitly as a functional evaluated over time-reversed paths $W_z[\hat{x}]$, which is implied by its appearance within the time-reversed path average $\langle ... \rangle_R$. Choosing $\mathcal{F} = \delta(W_z - W_z[x])$, we have $\hat{\mathcal{F}} = \delta(W_z + W_z[\hat{x}])$, since work is odd under time-reversal. Plugging this choice into Eq. (2.4) gives

$$P_F(W_z) = P_R(-W_z)exp(\beta[W_z - \Delta G_z])$$

$$\frac{P_F(W_z)}{P_R(-W_z)} = exp(\beta[W_z - \Delta G_z]). \tag{2.5}$$

Here $P_F(W_z)$ represents the probability of measuring the amount of work W_z during the F process, and $P_R(-W_z)$ is the probability of measuring the negative of that amount of work during the R process. This is just another version of the CFT, which is applicable to single-molecule experiments in that it derives from averages of the *molecule + cantilever* system for a repeated process in λ, and relates those averages to a quantity of interest, ΔG_z. ΔG_z can be determined from the data using the CFT

prediction that the F and R distributions cross at work values equal to the free energy difference [51]. This can be seen in Eq. (2.5) by setting $W_z = \Delta G_z$, which implies the distributions on the left side are equal.

In order to determine free energy differences at intermediate positions and thereby reconstruct $G(z)$ using the CFT, we assume work distributions can be related by convolution:

$$P_F(W_z)_{z_A \to z_B} = P_F(W_z)_{z_A \to z_M} \star P_F(W_z)_{z_M \to z_B}$$

$$P_R(-W_z)_{z_B \to z_A} = P_R(-W_z)_{z_B \to z_M} \star P_R(-W_z)_{z_M \to z_A} \qquad (2.6)$$

where $\star$ indicates convolution, and the protocol from which each distribution derives is written in subscripts, which we write in terms of z to emphasize that the system starts and ends at well-defined z. Because the system was allowed to equilibrate at z_A and z_B before pulling, in each convolution relation above two of the three distributions were measured. The unknown distribution was deconvolved from Eq. (2.6) using the discrete convolution theorem [52]. Since deconvolution is sensitive to input noise, some of the deconvolved distributions oscillated wildly or had large negative values. In these cases the input noise prohibited an accurate estimate of the distributions, and they were excluded from further analysis.

Pairing the deconvolved R distributions $z_M \to z_A$ with measured F distributions $z_A \to z_M$ and using Eq. (2.5) gives $\Delta G_z = G_{z_M} - G_{zA}$. Fig. 2.2 illustrates this approach. As expected, at increasing extensions the distributions move to the right, and the F distributions separate into bumps due to multiple pathways. An analogous procedure of pairing deconvolved F with measured R distributions was also performed. The free energy differences estimated from both procedures were combined to obtain G_{CFT}, a reconstruction of $G(z)$ using the CFT [Fig. 2.3(a)].

To check the accuracy of our approach, we used the same data to reconstruct
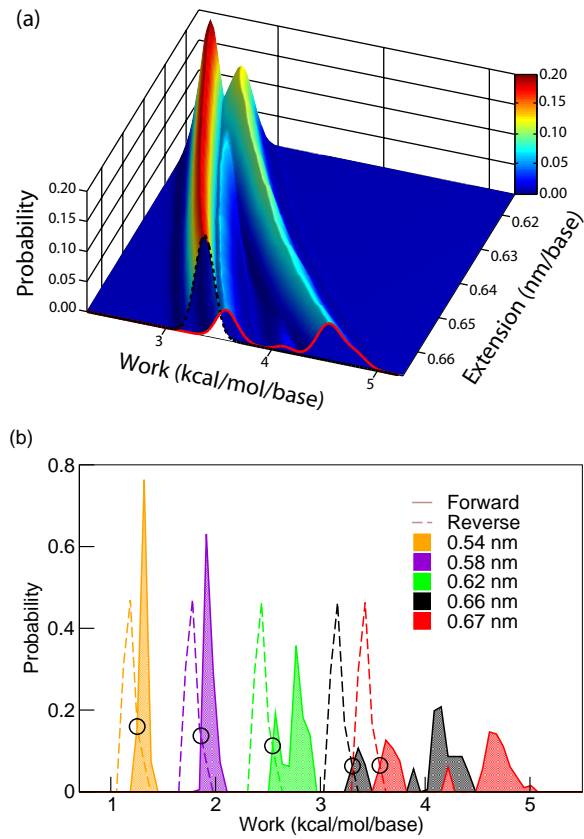
Figure 2.2 : Work distributions. (a) Distributions $P_F(W_z)_{z_A \to z_M}$ and $P_R(-W_z)_{z_M \to z_A}$ for intermediate extensions z_M spanning the overstretch transition, smoothed with a Gaussian kernel and visualized as two intersecting surfaces labeled at one end by a solid and dashed curve, respectively. R distributions were obtained from the data by deconvolution. (b) Representative unsmoothed F (solid) and R (dashed) distributions from (a). Circles indicate where F and R pairs cross, which determines our free energy estimate using the CFT.

$G(z)$ using the JE. Using a form of the JE applicable to single-molecule experiments [53, 50], the free energy profile was estimated as previously described [45, 47]. Because data was collected in both F and R directions, the JE affords two estimates: G_{JE}^F and G_{JE}^R. Curves were shifted to zero at low extension for comparison. The error determined by bootstrap analysis ($N_{boot} = 10^3$ iterations) was 0.5(1)% and 0.3(2)%, respectively, across the overstretching transition. In recent years, the JE has been applied to data from single-molecule experiments to construct free energy profiles for titin I27 domain unfolding [45, 47], membrane protein unfolding [54], the unfolding of DNA hairpins [44] and RNA hairpins to within $\pm \frac{1}{2} k_B T$ [55]. Figure 2.3(a) shows that our deconvolution approach agrees with the JE. At higher pulling velocity, however, the molecule rarely followed the reversible pathway, causing G_{JE}^F to overestimate the profile due to the finite number of measurements. As expected, the average work in the F direction $\langle W_z^F \rangle$ overestimates $G(z)$ due to the contribution of irreversible trajectories following the high-force pathway, while $\langle W_z^R \rangle$ tends to underestimate it.

Taking the derivative d/dz of free energy recovers the reversible FEC. Figure 2.3(b) shows derivatives of $G(z)$ estimates obtained by taking the analytical derivative of fitted smoothing splines. The average force measured in the F direction $\langle F_F \rangle$ deviates from the reversible curve, due to the presence of nonequilibrium pathways which shift the average force upward. The derivatives of our $G(z)$ estimates, on the other hand, exhibit the plateau characteristic of the reversible pathway. This demonstrates that both the CFT and JE picked out the equilibrium pathway from nonequilibrium data involving multiple pathways.

The utility of the CFT and JE requires a sufficient number N of repeated measurements. In the case of the JE estimator, the problem of its convergence is basically the problem of sampling the rare paths in the tail of the work distribution [55]. For
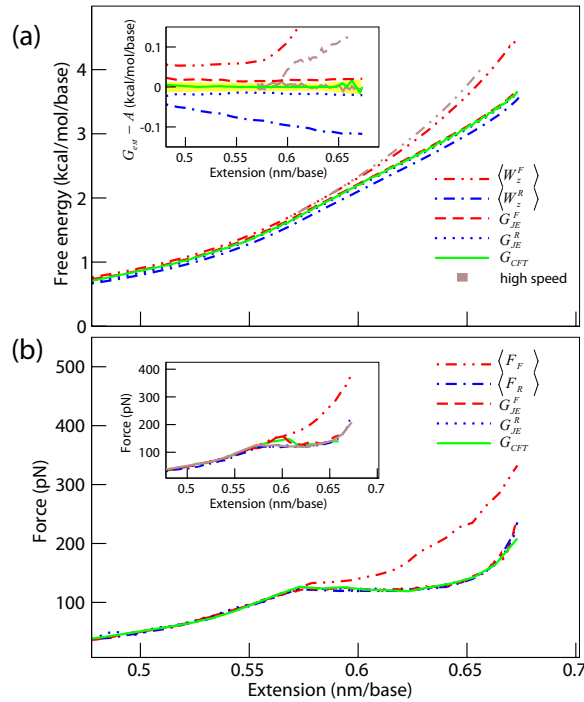
Figure 2.3 : Free energy profile reconstruction. (a) Profiles estimated using the CFT and our deconvolution method G_{CFT}, using the JE and F (R) direction data G_{JE}^F (G_{JE}^R), and by taking the average work in the F direction $\langle W_z^F \rangle$ and R direction $\langle W_z^R \rangle$. Inset: the difference between each estimate G_{est} and the average of low-speed pulling JE estimates, $A = (G_{JE}^F + G_{JE}^R)/2$. The yellow band represents $\pm$ the bootstrap error in A. Results shown are for low-speed ($\nu = 40$ nm/s), with selected high-speed ($\nu = 250$ nm/s) results in brown. (b) Derivative of free energy profile estimates, and the average force in each direction $\langle F_F \rangle$ and $\langle F_R \rangle$. Inset: results from high-speed pulling, compared to G_{CFT} from low-speed pulling (solid gray).

this reason, the JE estimator is generally expected to converge well for nano-sized systems and small dissipated work $\sim 1\ k_B T$ [41, 55]. However, here the work distribution had distinct peaks due to multiple pathways. The reversible work did not occur in the tail of a Gaussian, but at the lower of these peaks. This non-Gaussian nature of poly(dA) overstretching made the reversible pathway observable in an experimentally reachable N, for a system 1 μm long with dissipated work $\sim 1700\ k_B T$. The close agreement between G_{JE}^F and G_{JE}^R (Fig. 2.3) for $\nu = 40$ nm/s indicates good convergence after 19 stretching and 15 relaxing pulls. For comparison, for titin I27 domain unfolding it was found that at $\nu = 40$ nm/s, G_{JE}^F converged to within 10 percent in fewer than 30 pulls [47]. We also observed that the less-frequent, low-work FECs in the F direction look like typical R-direction FECs. According to a useful heuristic derived by Jarzynski [56], this is an indication that N was large enough to sample the equilibrium pathway in both directions.

2.4   Conclusions

Previous studies have derived relations from the Crooks path ensemble equation to determine the free energy profile of a composite system (*molecule + force probe*) [57]. Here we determined $G(z)$ by repeating the same pulling protocol which is sufficient for using the JE, without waiting for equilibration at intermediate positions. We also showed that when distinct pathways are present, as in the case of poly(dA) pulling, the CFT and JE pick out the equilibrium pathway.

Bibliography

[1] G. Felsenfeld and M. Groudine, "Controlling the double helix," *Nature*, vol. 421, pp. 448–453, 2003.

[2] K. Luger, A. W. Mder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 resolution," *Nature*, vol. 389, pp. 251–260, 1997.

[3] W.-S. Chen, W.-H. Chen, Z. Chen, A. A. Gooding, K.-J. Lin, and C.-H. Kiang, "Direct observation of multiple pathways of single-stranded DNA stretching," *Phys. Rev. Lett.*, vol. 105, p. 218104, 2010.

[4] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante, "The bacteriophage $\phi29$ portal motor can package DNA against a large internal force," *Nature*, vol. 413, pp. 748–752, 2001.

[5] T. Lionnet, S. Joubaud, R. Lavery, D. Bensimon, and V. Croquette, "Wringing out DNA," *Phys. Rev. Lett.*, vol. 96, p. 178102, 2006.

[6] J. F. Marko and E. D. Siggia, "Stretching DNA," *Macromolecules*, vol. 28, pp. 8759–8770, 1995.

[7] C. P. Calderon, W.-H. Chen, K.-J. Lin, N. C. Harris, and C.-H. Kiang, "Quantifying dna melting transitions using single-molecule force spectroscopy," *J. Phys.: Condens. Matter*, vol. 21, pp. 034114–1–7, 2009.

[8] M. Hogan and B. Austin, "Importance of dna stiffness in protein-dna binding specificity," *Nature*, vol. 329, pp. 263–266, 1987.

[9] S. D. Goodman and H. A. Nash, "Functional replacement of a protein-induced bend in a DNA recombination site," *Nature*, vol. 341, pp. 251–254, 1989.

[10] A. Stasiak and E. D. Capua, "The helicity of DNA in complexes with reca protein," *Nature*, vol. 299, pp. 185–186, 1982.

[11] M. Hegner, S. B. Smith, and C. Bustamante, "Polymerization and mechanical properties of single RecA-DNA filaments," *Proc. Natl. Acad. Sci. USA*, vol. 96, pp. 10109–10114, 1999.

[12] D. S. Johnson, L. Bai, B. Y. Smith, S. S. Patel, and M. D. Wang, "Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase," *Cell*, vol. 129, pp. 1299–1309, 2007.

[13] V. A. Bloomfield, D. M. Crothers, and I. Tinoco Jr., *Nucleic Acids: Structures, Properties, and Functions*. Sausalito, California: University Science Books, 2000.

[14] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, "Ionic effects on the elasticity of single DNA molecules," *Proc. Natl. Acad. Sci. USA*, vol. 94, pp. 6185–6190, 1997.

[15] D. N. Fuller, D. M. Raymer, V. I. Kottadiel, V. B. Rao, and D. E. Smith, "Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability," *Proc. Natl. Acad. Sci. USA*, vol. 104, pp. 16868–16873, 2007.

[16] D. N. Fuller, J. P. Rickgauer, P. J. Jardine, S. Grimes, D. L. Anderson, and D. E. Smith, "Ionic effects on viral DNA packaging and portal motor function in bacteriophage $\phi 29$," *Proc. Natl. Acad. Sci. USA*, vol. 104, pp. 11245–11250, 2007.

[17] B. D. Brower-Toland, C. L. Smith, R. C. Yeh, J. T. Lis, C. L. Peterson, and M. D. Wang, "Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA," *Proc. Natl. Acad. Sci. USA*, vol. 99, pp. 1960–1965, 2002.

[18] T. R. Strick, V. Croquette, and D. Bensimon, "Single-molecule analysis of DNA uncoiling by a type II topoisomerase," *Nature*, vol. 404, pp. 901–904, 2000.

[19] E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, and S. M. Block, "Direct observation of base-pair stepping by RNA polymerase," *Nature*, vol. 438, pp. 460–465, 2005.

[20] G. J. L. Wuite, S. B. Smith, M. Young, D. Keller, and C. Bustamante, "Single-molecule studies of the effect of template tension on T7 DNA polymerase activity," *Nature*, vol. 404, pp. 103–106, 2000.

[21] W. D. Volkmuth and R. H. Austin, "DNA electrophoresis in microlithographic arrays," *Nature*, vol. 358, pp. 600–602, 1992.

[22] J. O. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W. W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, P. Silberzan, and R. H. Austin, "The dynamics of genomic-length DNA molecules in 100-nm channels," *Proc. Natl. Acad. Sci. USA*, vol. 101, pp. 10979–10983, 2004.

[23] W. Reisner, K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, M. Rosen, J. Sturm, S. Chou, E. Frey, and R. Austin, "Statics and dynamics of single dna molecules confined in nanochannels," *Phys. Rev. Lett.*, vol. 94, p. 196101, 2005.

[24] A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, and D. Branton, "Unzipping kinetics of double-stranded DNA in a nanopore," *Phys. Rev. Lett.*, vol. 90, p. 238101, 2003.

[25] J. L. Hutter and J. Bechhoefer, "Calibration of atomic-force microscopy tips," *Rev. Sci. Instrum.*, vol. 64, pp. 1868–1873, 1993.

[26] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, "Sequence-dependent mechanics of single DNA molecules," *Nat. Struct. Biol.*, vol. 6, pp. 346–349, 1999.

[27] K. C. Neuman, E. H. Chadd, G. F. Liou, K. Bergman, and S. M. Block, "Characterization of photodamage to escherichia coli in optical traps," *Biophys. J.*, vol. 77, pp. 2856–2863, 1999.

[28] S. B. Smith, Y. J. Cui, and C. Bustamante, "Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules," *Science*, vol. 271, pp. 795–799, 1996.

[29] T. R. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette, "Behavior of supercoiled DNA," *Biophys. J.*, vol. 74, pp. 2016–2028, 1998.

[30] J. Yan, D. Skoko, and J. F. Marko, "Near-field-magnetic tweezer manipulation of single DNA molecules," *Phys. Rev. E*, vol. 70, p. 011905, 2004.

[31] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, "Stretching DNA with optical tweezers," *Biophys. J.*, vol. 72, pp. 1335–1346, 1997.

[32] H. Clausen-Schaumann, M. Rief, C. Tolksdorf, and H. E. Gaub, "Mechanical stability of single DNA molecules," *Biophys. J.*, vol. 78, pp. 1997–2007, 2000.

[33] S. Cocco, J. Yan, J.-F. Léger, D. Chatenay, and J. F. Marko, "Overstretching and force-driven strand separation of double-helix DNA," *Phys. Rev. E*, vol. 70, p. 011910, 2004.

[34] T. Odijk, "Stiff chains and filaments under tension," *Macromolecules*, vol. 28, pp. 7016–7018, 1995.

[35] S. Whitelam, S. Pronk, and P. L. Geissler, "There and (slowly) back again: Entropy-driven hysteresis in a model of dna overstretching," *Biophys. J.*, vol. 94, pp. 2452–2469, 2008.

[36] H. Fu, H. Chen, J. F. Marko, and J. Yan, "Two distinct overstretched DNA states," *Nucleic Acids Res.*, vol. 38, pp. 5594–5600, 2010.

[37] C. Ke, M. Humeniuk, H. S-Gracz, and P. E. Marszalek, "Direct measurements of base stacking interactions in DNA by single-molecule atomic-force spectroscopy," *Phys. Rev. Lett.*, vol. 99, p. 018302, 2007.

[38] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, "Mechanical separation of the complementary strands of DNA," *Proc. Natl. Acad. Sci. USA*, vol. 94, pp. 11935–11940, 1997.

[39] C. Danilowicz, V. W. Coljee, C. Bouzigues, D. K. Lubensky, D. R. Nelson, and M. Prentiss, "DNA unzipped under a constant force exhibits multiple metastable intermediates," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 1694–1699, 2003.

[40] K. Hatch, C. Danilowicz, V. Coljee, and M. Prentiss, "Measurements of the hysteresis in unzipping and rezipping double-stranded DNA," *Phys. Rev. E*, vol. 75, p. 051908, 2007.

[41] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, vol. 78, pp. 2690–2693, 1997.

[42] G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," *Phys. Rev. E*, vol. 60, pp. 2721–726, 1999.

[43] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr., and C. Bustamante, "Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies," *Nature*, vol. 437, pp. 231–234, 2005.

[44] A. N. Gupta, A. Vincent, K. Neupane, H. Yu, F. Wang, and M. T. Woodside, "Experimental validation of free-energy-landscape reconstruction from nonequilibrium single-molecule force spectroscopy measurements," *Nature Physics*, vol. 7, pp. 631–634, 2011.

[45] N. C. Harris, Y. Song, and C.-H. Kiang, "Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality," *Phys. Rev. Lett.*, vol. 99, p. 068101, 2007.

[46] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., and C. Bustamante, "Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality," *Science*, vol. 296, pp. 1832–1835, 2002.

[47] N. C. Harris and C.-H. Kiang, "Velocity convergence of free energy surfaces from

single-molecule measurements using Jarzynski's equality," *Phys. Rev. E*, vol. 79, p. 041912, 2009.

[48] G. E. Crooks, "Path-ensemble averages in systems driven far from equilibrium," *Physical Review E*, vol. 61, pp. 2361–2366, 2000.

[49] C. Jarzynski, "Work fluctuation theorems and single-molecule biophysics," *Prog. Theor. Phys. Suppl.*, vol. 165, pp. 1–17, 2006.

[50] G. Hummer and A. Szabo, "Free energy surfaces from single-molecule force spectroscopy," *Acc. Chem. Res.*, vol. 38, pp. 504–513, 2005.

[51] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr., and C. Bustamante, "Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies," *Nature*, vol. 437, pp. 231–234, 2005.

[52] W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery, *Numerical Recipes in Fortran 77*, ch. 13, pp. 530–537. second ed., 2003.

[53] G. Hummer and A. Szabo, "Free energy reconstruction from nonequilibrium single-molecule pulling experiments," *Proc. Natl. Acad. Sci. USA*, vol. 98, pp. 3658–3661, 2001.

[54] J. Preiner, H. Janovjak, C. Rankl, H. Knaus, D. A. Cisneros, A. Kedrov, F. Kienberger, D. J. Müller, and P. Hinterdorfer, "Free energy of membrane protein unfolding derived from single-molecule force measurements," *Biophys. J.*, vol. 93, pp. 930–937, 2007.

[55] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., and C. Bustamante, "Equilibrium information from nonequilibrium measurements in an experimental test

of Jarzynski's equality," *Science*, vol. 296, pp. 1832–1835, 2002.

[56] C. Jarzynski, "Rare events and the convergence of exponentially averaged work values," *Phys. Rev. E*, vol. 73, p. 046105, 2006.

[57] I. Junier, A. Mossa, M. Manosas, and F. Ritort, "Recovery of free energy branches in single molecule experiments," *Phys. Rev. Lett.*, vol. 102, p. 070602, 2009.