

Folding energy and kinetics of mutually interacting DNA Hairpins

Author: David Schönenberger

Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Marc Rico / Felix Ritort

(Dated: June 18, 2020)

Abstract: The Szilard engine is a Maxwell demon thought experiment where information is converted into work. We propose a novel DNA molecule with two hairpins in serial that could act as a complex four-state Szilard engine using optical tweezers to apply a force. We synthesized the molecule with identical hairpins to perform two different experiments with optical tweezers, passive mode and pulling. Only 3 different states were observed as the intermediate state is degenerated. Their energy levels were obtained from kinetic rates measured in passive mode and work measured in pulling cycles. We conclude that the molecule follows well the fluctuation theorems and the Bell-Evans theory, with the exception of a Chevron type curvature in the kinetic rates due to the degeneration. The obtained data will help in the development and study of a Szilard engine with this molecule.

I. INTRODUCTION

The Szilard engine [1] is a single molecule thought experiment proposed in 1929 as a refinement of the famous Maxwell's Demon [2]. This imaginary engine demonstrates how information can be converted into energy. A cycle in the Szilard engine consists of two steps: observation of the state of the molecule and application of a work extraction process depending on the measurement outcome. The original version implements these steps using a single gas particle in a box. A observation is made when a wall is inserted that divides the box in two. Then, the particle pushes the wall to the opposite side, producing a work which depends on the previous position of the particle.

The realization of this experiment requires to manipulation and measurement of the state of a single molecule as a function of time. For this reason, thermal fluctuations play a role as they must be small enough, in the order of $k_B T$, to be able to measure the information. Novel implementations have been developed in recent years to overcome these difficulties. For example, in 2014 a single-electron box was created by trapping an excess electron between two connected metallic islands at 0.1 K [3]. The electron was manipulated by changing the potential difference between the islands, whereas the location was measured using an electrometer. Later, in 2019 a nanoscale Szilard engine was made with a single two-state DNA molecule tethered between two micro-beads at room temperature [4]. Optical tweezers were used to measure the state of the DNA and extract a work by varying the position of the optical trap. This method allows to create Szilard engines with an arbitrary number of DNA states.

Our objective is to study a DNA molecule made of two interacting hairpins serially connected in order to build a Szilard engine from it. We predict that each hairpin will influence the other by a continuous exchange of information and thus significantly alter the extracted energy in the possible Szilard engine. The behaviour of this

molecule will be investigated with optical tweezers, measuring its kinetic rates and calculating its energy levels using different methods.

First, we measured the kinetic rates in passive experiments where the state of the molecule is observed as a function of time. We subjected the molecule to different forces in order to study the force dependence of the kinetic rates and apply the Bell-Evans theory which allows us to estimate the folding energy at zero force, ΔG_0 , as well as to characterize the energy levels of the molecular Free Energy Landscape (FEL).

Second, we calculated the folding energy from pulling experiments using the Fluctuation theorem (FT) [5]. This theorem allows us to measure the equilibrium folding energy at a given force, ΔG , from non-equilibrium pulling experiments by relating the stretching and releasing work distributions. In order to convert ΔG to the folding free energy at zero force, we subtracted extra contributions by modeling the molecule using the Worm-Like Chain (WLC) elastic model. We validated our results of ΔG_0 by comparing the values obtained using the FT and the ones coming from the kinetic results. Moreover, we compared our values with the ones provided by the unified oligonucleotide dataset [6]. The implementation of the Szilard engine and his setup is left for future work.

II. MATERIALS AND METHODS

A. Synthesis of two serially connected DNA hairpins

A DNA hairpin is a simple structure that is formed by a single single-stranded DNA (ssDNA) that has two regions complementary with each other and bind to form a double helix ended with an unpaired loop. The DNA molecule we are studying consists of two hairpins linked by three equal length double-stranded DNA (dsDNA) handles joined in serial order: handle, hairpin, handle,

hairpin, handle (FIG 1a).

	Sequence
O1	AGTTAGTGGTGGAAACACAGTGCCAGCGC <u>GCGAGCCATAATCTCATCTG GAAA CAGA</u> TGAGATTATGGCTCGC AGTTAGTGGTG GAAACACAGTGCCAGCGC
O2	<u>GCGAGCCATAATCTCATCTG GAAA CAGA</u> TGAGATTATGGCTCGC AGTTAGTGGTG GAAACACAGTGCCAGCGC
O3	GCGCTGGCACTGTGTTTCCACCACTAATC

Table I: DNA sequence of the oligos used in the synthesis. The sequence that describes the hairpins is underlined.

To synthesize a DNA with two hairpins, we hybridized three different oligonucleotides (Table 1). The first oligonucleotide (O1 in Table 1) contains one strand of the desired dsDNA handle plus the first DNA hairpin and one strand of the handle between the two hairpins. The second oligonucleotide (O2 in Table 1) contains the second DNA hairpin plus a ssDNA handle. The third oligonucleotide (O3 in Table 1) has the complementary sequence of ssDNA handles and it is used to hybridize with them and form three dsDNA handles. The first oligonucleotide has a biotin at its 5' end and the second oligonucleotide was modified with a digoxigenin tail at its 3' end. These modified ends bind with beads coated with their complementary substance, streptavidin (SA) and anti-digoxigenin (AD) respectively.

B. Optical tweezers

Optical tweezers are instruments that generate highly focused beams to produce a force when an object and its surrounding medium have a different refractive index. They create force gradients called optical traps (OT) to manipulate micron sized dielectric objects. In this work we used a miniaturized optical tweezers [7] to manipulate the molecule by trapping the AD bead and modifying its position while the SA bead is fixed at the tip of a micro-pipette by air suction (FIG 1a).

This setup allows the molecule to be stretched and generates a mechanical tension to mechanically unzip the stem of the DNA hairpins. The control parameter λ is the distance between the tip of the micro-pipette and the center of optical trap.

III. RESULTS AND DISCUSSION

In this report we studied the simplest case where both hairpins are identical and each one has two possible states: folded and unfolded (FIG 1b). Therefore, the molecular system has 4 different states (FIG 1c). Nevertheless due to its symmetry we are not able to dis-

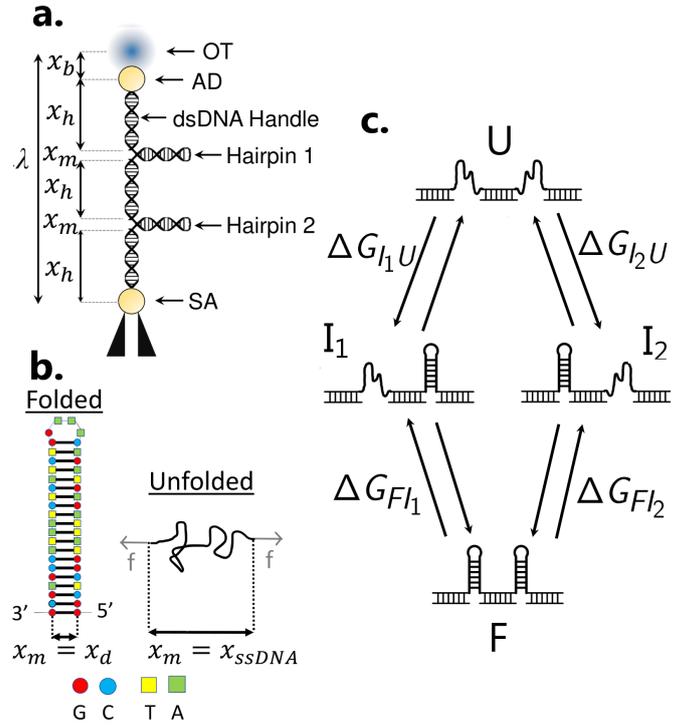


Figure 1: **a.** The experimental setup where x_b is the distance between the optical trap (OT) and the AD bead, x_h is the length of the dsDNA handles, x_m is the length of the hairpins and λ is the total distance between the SA bead and the OT. **b.** The two states of a hairpin where x_d is the length of the folded hairpin and x_{ssDNA} is the length of the unfolded hairpin. **c.** Representation of the 4 states of the molecule with all the possible observed transitions. U is the unfolded state (both hairpins unfolded), F is the folded state (both hairpins folded) and I_1, I_2 are the intermediary states (only one hairpin is folded).

tinguish the two intermediate states I_1 and I_2 . For this reason, we define the union of I_1 and I_2 as I .

A. Passive mode experiments

Passive mode hopping experiments allow us to determine the state of the DNA molecule in real time while the control parameter λ is kept constant. In this situation, we see sudden changes in the force (FIG 2a-left) associated with the different states. To associate one state at each force-time point we needed to determine with high accuracy the force corresponding to each state.

The force distribution at a given λ show three peaks corresponding to the different states, F , U and I (FIG 2a-right). As a folded hairpin has lower extension than the unfolded hairpin it implies that the highest force corresponds to the folded state F , the lowest force to the unfolded state U and the one between these two corresponds to the intermediate state I where only one of the two hairpins is open. We have made a Gaussian fit with

a probabilistic factor for each peak to classify the state of the molecule as F , I or U by choosing the one with the closest force at any time for a given value of λ .

Each stable state signals the presence of a local minimum of energy in the Free Energy Landscape (FEL). The FEL is the map of the different configurations of the molecule to their corresponding free energy. Since I_1 and I_2 are indistinguishable, our first approach is to simplify the FEL using a single reaction coordinate that characterizes each configuration: the extension of the DNA. FIG 2b shows a theoretical description of the FEL of our molecule. When a mechanical tension is applied to both ends of a hairpin, the state with more extension is energetically favoured. As the energy difference between two states is of the order of $k_B T$, the thermal fluctuations can cause the molecule to spontaneously change its state. These fluctuations are a random process that can be characterized by residence times and kinetic rates which are related with the FEL.

The identification was done by fitting the measured forces to a three Gaussian distribution where the mean of a Gaussian is the force of the state and the weight of a Gaussian is the occurrence of the corresponding state. To determine the kinetic rates, we calculated the average time that the molecule takes to change from one state to another, $k_{S \rightarrow S'} = 1/\langle \tau_{S \rightarrow S'} \rangle$. Note that direct transitions between F and U were extremely rare because they were only observed due to the time discretization of the measures. We can verify that our calculations are correct using the detailed balance condition,

$$P_S \cdot k_{S \rightarrow S'} = P_{S'} \cdot k_{S' \leftarrow S} \quad (1)$$

where $P_{S(S')}$ is the weight of state $S(S')$, $k_{S \rightarrow (\leftarrow) S'}$ is the kinetic rate from $S(S')$ to $S'(S)$. FIG 2c shows that Eq.(1) is fulfilled.

Several theoretical studies propose how to extract important features of the molecular FEL from the force dependent kinetic rates. According to the Bell-Evans model [8, 9] the kinetic force-dependent rates are:

$$\begin{aligned} k_{S \rightarrow S'} &= k_m e^{\beta f x_S^\ddagger} \\ k_{S' \leftarrow S} &= k_m e^{-\beta f x_{S'}^\ddagger + \beta \Delta G_{SS'}} \\ k_{I \rightarrow U} &= k_{I_1 \rightarrow U} + k_{I_2 \rightarrow U} \\ k_{F \leftarrow I} &= k_{F \leftarrow I_1} + k_{F \leftarrow I_2} \end{aligned} \quad (2)$$

$$k_{F \leftarrow I} = k_{F \leftarrow I_1} + k_{F \leftarrow I_2} \quad (3)$$

Where $\beta = 1/k_B T$, k_B is the Boltzmann constant, and T is the environmental temperature. x_S^\ddagger and $x_{S'}^\ddagger$ are the distances between the transition state TS and the state S or S' respectively (FIG 2b). k_m is the kinetic rate of $S \rightarrow S'$ at zero force. We determined the parameters that characterize the FEL by fitting the kinetic rates to these expressions.

We can see in FIG 2d that the kinetic rates from U to I and from F to I are linear on a logarithmic scale in relation to the force, as predicted by the Eq.(2). On the other hand, kinetics rates from I to U and from I to F have a curvature known as chevron plot shape. This

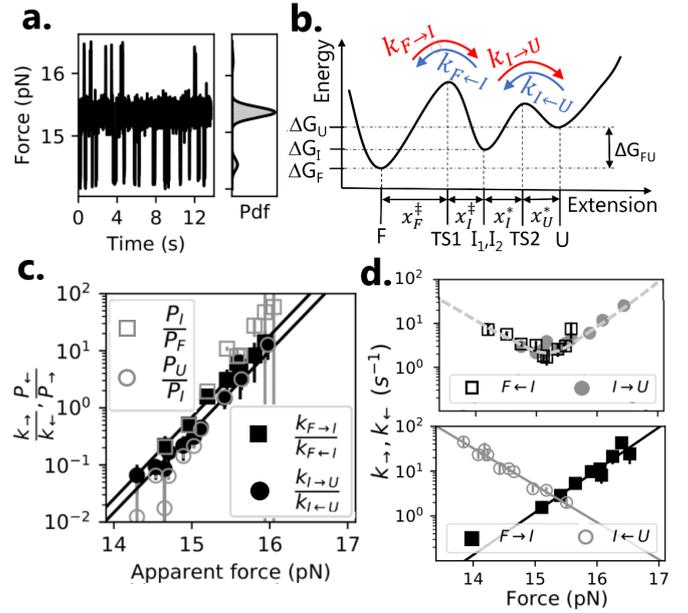


Figure 2: Results of the passive mode experiments. **a.** Sample of measured force vs time data and force histogram. **b.** Shape of the FEL of our molecule. It has 3 stable states at F , I and U . $TS1$ and $TS2$ are the transition states, located the local maximums. **c.** Representation of the detailed balance. The apparent force is the mean force of the two states involved. Black lines are the fitted to the kinetics rates for transitions between F and I (upper line) and between I and U (bottom line) **d.** Measured kinetic rates vs force of the initial state. The dashed lines in the upper plot did not include all the data points for the fit because of the visible curvature at the center of the upper plot.

shape originates from the degeneration of the I state and suggest that a single reaction coordinate is not enough to discriminate them. The proper application of the Bell-Evans theory requires to consider I_1 and I_2 separately.

To determine the energy levels of each state we computed the linear regression of the kinetic rates and determined the coexistence forces between the states $I-U$ (f_{IU}^c) and $I-F$ (f_{FI}^c) (Table I). The coexistence force is the force where both kinetics, $k_{S \rightarrow S'}$ and $k_{S' \leftarrow S}$, are equal. The difference between the energy levels at zero force can be calculated using,

$$\begin{aligned} \Delta G_{FI}^0 &= (x_F^\ddagger + x_I^\ddagger) f_{FI}^c \\ \Delta G_{IU}^0 &= (x_I^* + x_U^*) f_{IU}^c \end{aligned} \quad (4)$$

Due to the symmetry of the molecule, ΔG_{FI}^0 and ΔG_{IU}^0 should be equal or similar as in both cases only one hairpin unfolds. The results obtained in Table I are indeed compatible. We can compare these values with the prediction from MFold web server [6] for our hairpins. According to MFold the theoretical energy difference of a single hairpin is $\Delta G_{th} = 30.0$ kcal/mol. Pulling experiments offer another method to calculate energy differences which we can use to confirm these results.

x_F^\ddagger (nm)	x_I^\ddagger (nm)	f_{FI}^c (pN)	ΔG_{FI}^0 (kcal/mol)
8.9 ± 0.2	7.1 ± 0.2	15.18 ± 0.04	35 ± 1
x_I^* (nm)	x_U^* (nm)	f_{IU}^c (pN)	ΔG_{IU}^0 (kcal/mol)
8.3 ± 0.2	7.7 ± 0.2	15.34 ± 0.04	35 ± 1

Table II: Results obtained from the application of the Bell-Evans theory on the kinetic rates. x_F^\ddagger , x_I^\ddagger , x_I^* , x_U^* are the characteristic extensions of the DNA in the FEL (FIG 2b). F_{FI}^c and F_{IU}^c are the forces where $k_{F \rightarrow I} = k_{F \leftarrow I}$ and $k_{I \rightarrow U} = k_{I \leftarrow U}$ respectively.

B. Pulling experiments

An important phenomena caused by the fluctuating state of the molecule is that the work exerted on it does also fluctuate. In this work we used the Crooks fluctuation theorem [10] and its corollary the Jarzynski equality [11] to determine the folding energy difference from the work extracted from the pulling cycles. The Jarzynski equality brings a direct measurement of the folding free energy difference between the initial state and the final state from the work extracted from the unfolding trajectories,

$$\exp\left(-\frac{\Delta G}{k_B T}\right) = \langle \exp\left(-\frac{W}{k_B T}\right) \rangle \quad (5)$$

where ΔG is the free energy difference, W is the work required to go from an initial trap position to a final trap position and $\langle \dots \rangle$ is the average taken over an infinite number of pulls. In practical cases, the number of pulls is always finite and the Jarzynski equality is strongly biased. To overcome this, we used the Crooks fluctuation theorem that relates the work unfolding distribution, $P_F(W)$, and the refolding work distribution, $P_R(-W)$.

$$\frac{P_F(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta G}{k_B T}\right) \quad (6)$$

In Eq.(6) ΔG equals the reversible work, which is the value of W at which the unfolding and folding work distributions cross, $P_U(W) = P_F(-W)$. Our single molecule system allows us to validate the Jarzynski equality and Crooks fluctuation theorems while we apply them to calculate free energy differences.

We carried out pulling cycles by moving the optical trap from an initial position where the molecule was always in F to a final position where the molecule was always observed in U . The folding or forward process goes from F to U whereas the refolding or reverse process goes from U to F . The work required in each cycle was calculated by integrating the force - λ curve (FIG 3a) and the obtained histograms are plotted in FIG 3b.

First, we applied the Jarzynski equality Eq.(5) in both directions to calculate the energy difference between the folded state at f_{min} and the unfolded state at f_{max} . In the forward direction (from F to U) we obtained a value

of $\Delta G = 462 \pm 3 k_B T$ and in the reverse direction a value of $\Delta G = 459 \pm 3 k_B T$, both energies are compatible.

Then, we verified Crooks FT (Eq.(6)), calculating the ratio of the work histograms (FIG 3b). With the regression parameters we obtain a value of $\Delta G = 459.3 \pm 0.2 k_B T$ in accordance with Jarzynski's equality. We have found that both Jarzynski equality and Crooks FT are coherent with each other but we cannot directly compare the energy differences they give with the results from the passive mode experiments. To compare them we must transform the energies differences from pulling experiments to energies at zero force. The energy variation between F and U at zero force ΔG_{FU}^0 is related to the energy difference obtained in the pulling ΔG by the following equation:

$$\Delta G = \Delta G_{FU}^0 + \Delta G_{ssDNA}(f_{max}) - \Delta G_d(f_{min}) + \Delta G_{bh}(f_{min}, f_{max}) \quad (7)$$

where $\Delta G_{ssDNA}(f_{max})$ is the stretching energy of the single stranded DNA, $\Delta G_d(f_{min})$ is the energy to orient the folded molecule and $\Delta G_{bh}(f_{min}, f_{max})$ is the energy to stretch the handles plus the energy to move the bead inside the optical trap [13].

The ssDNA contributions was calculated using the Worm-Like Chain (WLC) model with the adequate parameters for DNA [12]. In addition, we have verified these parameters by estimating the single-stranded length released when the hairpins open at different forces (FIG 3c). In contrast with [12] where only one hairpin is opened, here we had to consider three different cases. First, the transition from F to I where only one hairpin was opened; second, the transition from I to U where only one hairpin was opened; third the transition from F to U where two hairpins were opened. We found that measured elastic properties considering the three different extensions agree between them and with the values reported in the literature. The others contributions were calculated as proposed in [12].

The mean folding free energy is $\Delta G_{FU}^0 = 65 \pm 3$ kcal/mol considering Jarzynski equality and Crooks FT. Moreover from Table II we can calculate the total folding free energy, ΔG_{FU}^0 , as the addition of ΔG_{FI}^0 and ΔG_{IU}^0 , $\Delta G_{FU}^0 = 70 \pm 1$ kcal/mol. Both values are compatible and henceforth the fluctuations theorems and the Bell-Evans theory yield the same result. We have verified that our analysis in the passive mode experiments is correct.

IV. CONCLUSION

We learned how to successfully synthesise a DNA molecule with two hairpins and how to manipulate it using optical tweezers. In this work we studied the case of a DNA molecule with identical hairpins and we easily identified three different states. A 3-state Szilard engine is possible with this molecule because the state was clearly identified at any time. On the other hand, it is not a

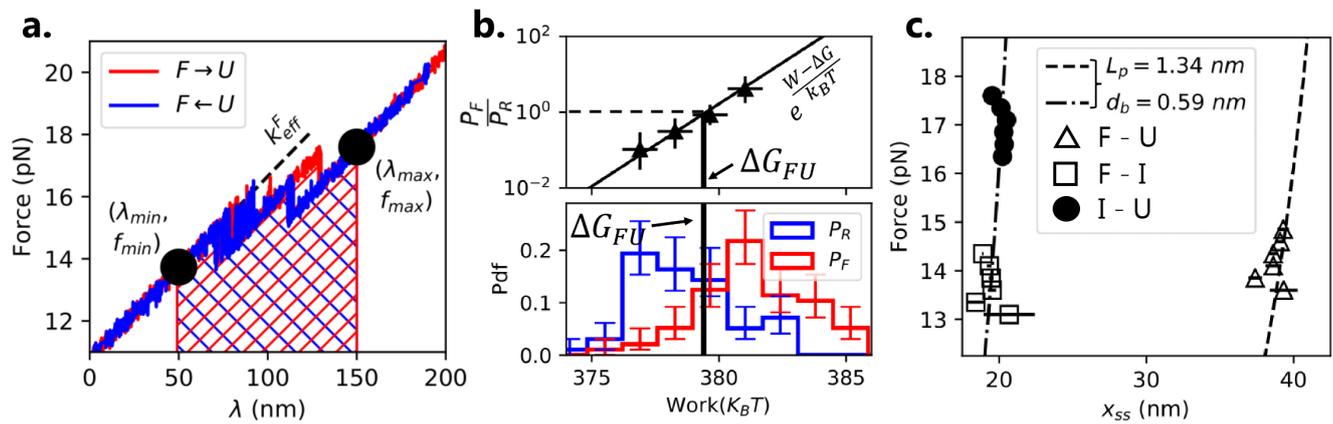


Figure 3: **a.** Pulling data from a single cycle. The black dots are the limits used to calculate the work (area dashed under the curve). The black dashed line is the best fit to the elastic constant k_{eff}^F used to calculate $\Delta G(f_{min}, f_{max})$ in EQ. 7. **b.** Work histograms (bottom plot) of the forward (red) and reverse (blue) process and their ratio (upper plot). ΔG_{FU} is the work value where $P_F = P_R$. **c.** Force vs released single-strand length between the different states. The dashed and dash-dotted lines are the WLC model using a standard [12] persistence length L_p and interphosphate distance d_b for DNA with two hairpins and one hairpin respectively.

normal 3-state molecule because the intermediate state is degenerated. Consequently we have seen that the kinetic rates have an atypical Chevron-shaped deviation. In spite of this, we have been able to calculate the free energy levels from the kinetic rates and the fluctuation theorems. The values obtained are consistent with the theoretical models. To continue and implement a Szilard engine, a software should be developed to act as a Maxwell daemon and identify the state with help from the results obtained in this work.

In future works, the degeneration of the intermediate state could be avoided by using hairpins of different number of base pairs or a different sequence. With such a

molecule, one could observe the two different intermediate state and study the transition between them. Moreover, with this four states configuration could be possible to study the information transfer between hairpins using a novel Fluctuation Theorem for feedback systems.

Acknowledgments

I would like to thank my advisors, Marc Rico and Felix Ritort, for their for all their dedication and invaluable help. I would also like to thank my family for their support and advice. I dedicate this TFG to my dear father, Hugo Schönenberger.

-
- [1] L. Szilard, “Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen,” *Zeitschrift für Physik*, vol. 53, no. 11-12, pp. 840–856, 1929.
- [2] C. H. Bennett, “Demons, engines and the second law,” *Scientific American*, vol. 257, no. 5, pp. 108–117, 1987.
- [3] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin, “Experimental realization of a szilard engine with a single electron,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 38, pp. 13786–13789, 2014.
- [4] M. Ribezzi-Crivellari and F. Ritort, “Large work extraction and the landauer limit in a continuous maxwell demon,” *Nature Physics*, vol. 15, no. 7, pp. 660–664, 2019.
- [5] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, “Verification of the crooks fluctuation theorem and recovery of rna folding free energies,” *Nature*, vol. 437, no. 7056, pp. 231–234, 2005.
- [6] M. Zuker, “Mfold web server for nucleic acid folding and hybridization prediction,” *Nucleic acids research*, vol. 31, no. 13, pp. 3406–3415, 2003.
- [7] S. B. Smith, Y. Cui, and C. Bustamante, “[7] optical-trap force transducer that operates by direct measurement of light momentum,” in *Methods in enzymology*, vol. 361, pp. 134–162, Elsevier, 2003.
- [8] G. I. Bell, “Models for the specific adhesion of cells to cells,” *Science*, vol. 200, no. 4342, pp. 618–627, 1978.
- [9] E. Evans and K. Ritchie, “Dynamic strength of molecular adhesion bonds,” *Biophysical journal*, vol. 72, no. 4, pp. 1541–1555, 1997.
- [10] C. Bustamante, J. Liphardt, and F. Ritort, “The nonequilibrium thermodynamics of small systems,” *arXiv preprint cond-mat/0511629*, 2005.
- [11] C. Jarzynski, “Nonequilibrium equality for free energy differences,” *Phys. Rev. Lett.*, vol. 78, pp. 2690–2693, Apr 1997.
- [12] A. Alemany i Arias, “Dynamic force spectroscopy and folding kinetics in molecular systems,” 2014.
- [13] A. Severino, A. M. Monge, P. Rissone, and F. Ritort, “Efficient methods for determining folding free energies in single-molecule pulling experiments,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2019, no. 12, p. 124001, 2019.