Entropy production of DNA-translocating molecular motors

Author: Víctor Rodríguez Franco
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Maria Mañosas
Felix Ritort Farrán

Abstract: Helicases are molecular motors that convert the chemical energy of ATP hydrolysis into mechanical work to move along one strand of DNA and unzip the double helix. Using magnetic tweezers (MT) we follow the activity of a single helicase unzipping a DNA hairpin. From these experiments we can characterize the enzyme motion (e.g. velocity and diffusion) but the ATP hydrolysis reaction is not directly measured. Here we investigate whether we can infer information about the helicase chemical cycle from our helicase displacement data by using non-equilibrium relations such as the thermodynamic uncertainty relation (TUR) and the fluctuation theorem (FT) for entropy production. To address this question, we use the random walk formalism to model the helicase motion and we analytically derive expressions for the TUR and the FT. The derived theoretical results are verified with simulations of the model and compared with experiments. We find qualitatively agreement between experiments and theory. However, some important differences are observed. In particular, the distributions of the helicase displacement deviates from the Gaussian distribution predicted by the theory and the experimental test of the FT fails. We conclude that a refined model is needed to better describe the real experimental system.

I. INTRODUCTION

DNA is the molecule carrying the genetic information in cells. It is made of two strands that wrap around each other forming a double helix. Each strand consists of a sequence of bases (adenine (A), guanine (G), cytosine (C), and thymine (T)) that encode the different genes. DNA is manipulated (unwound, rewound, cut and pasted, etc) by molecular motors, enzymatic proteins that convert chemical energy into mechanical work. Mutations in DNA-processing motors genes are responsible for severe diseases [1]. Understanding how these nanoscale machines work at the molecular level is crucial not only from a biophysical perspective but also for their potential applications in medicine. In this work we focus on the study of one of these motors, called helicase. Helicases are proteins that catalyse the separation of the two strands of the DNA double-helix, a common reaction in many cellular processes such as replication, transcription, or translation [2].

These enzymes translocate along one strand of DNA while promoting the unwinding of the DNA double helix using the energy of adenosine triphosphate (ATP) hydrolysis. In the helicase mechanochemical cycle the energy stored in the ATP bonds is released and converted into mechanical work used to reduce the activation barrier needed to unzip the DNA base pairs. If the mechanochemical coupling is tight, for each ATP consumed the motor takes one step forward, which is typically a step of one or few bases along the DNA chain. Since bases are separated by 3-5 Armstrong, this motor elemental motion, called physical step size, is on the scale of the nanometer (nm). On the other hand, the forces generated are on the scale of the PicoNewton (pN). Finally, the energies involved in this unwinding reaction, i.e the energy of breaking a DNA base-pair and the energy from ATP hydrolysis, are on the order of the thermal agitation energy k_BT (where k_B is the Boltzmann constant and T is the temperature) that at ambient temperature corresponds to $1k_BT \sim 4$ pNnm. Therefore helicases, and molecular motors in general, work in strong Brownian environment.

Traditionally, these molecular reactions have been studied using standard biochemical ensemble assays, that measure the average behaviour of a large ensemble of molecules (typically on the scale of the Avogadro's number). These average measurements provide limited information on molecular reactions occurring at nano-scale typically involving energies of few k_BT, where deviations from the average behaviour play a very important role [3]. In the last 30 years, the emergence of single-molecule techniques has completely changed the field of molecular biophysics, providing a way to measure properties of individual molecules (e.g. position, orientation, end-to-end extension) that can be used as reaction coordinates to follow the evolution of the molecule along a reaction pathway[4–6].

In particular, in single molecule force spectroscopy techniques, including atomic force microscope optical tweezers and magnetic tweezers (MT), mechanical manipulation is achieved by tethering a single molecule between a surface and a force probe (magnetically or optically trapped bead, or cantilever tip) [6, 7]. Using these techniques we can measure physical quantities (and their fluctuations) that are difficult to determine in bulk assays, such as mechanical force, torque or molecular extension. By designing a DNA molecular construct in such a way that the motor enzymatic activity leads to a changes in some of these magnitudes, we can use them as a reaction coordinates. Therefore by monitoring these reaction coordinates we can follow the enzymatic reaction of an

individual protein motor in real time. These techniques cover forces in the range of fN to nN and extensions in the range of nm to μm and have temporal resolution on the range of seconds to microseconds allowing us to perform accurate measurements of complex molecular reactions.

One typical example of single-molecule experiment is the mechanical unzipping of DNA, where the double helix is opened mechanically pulling each strand apart. Using optical tweezers we can capture a micro-sized bead into a moving optical trap. A DNA molecule can be tethered between the trapped bead and another bead immobilized via suction to a pipette. Moving the micro-pipette (or the optical trap) we can increase the applied force until breaking the intramolecular bonds between the base-pairs leading to DNA unzipping. The result is a force-extension saw-tooth signal that is characteristic of the specific DNA sequence [8]. This assay allows investigating internal DNA properties and measure the base-pair energies, enthalpies and entropies [9]. This example has similarities with the experiments performed in this work where we follow the opening of the DNA but, in this case, carried out by the helicase.

In this work we use MT to study the activity of a replicative helicase, gp41 of the T4 bacteriophage. An schematic representation of the MT set up is shown in Fig. 5. A magnetic field gradient is used to manipulate a micron-sized bead by applying both force and torque. Tethering a molecular construct (e.g. a DNA molecule) between the bead and a surface we can stretch and rotate the tethered molecule. By measuring the molecular extension and the applied force or torque we can perform single molecule experiments. In our experiments we work with a DNA hairpin (see Fig. 5) attached by one of its extremities to a glass surface of the microfluidic chamber (via digoxigenin anti-digoxigenin interaction, see methods) and by the other to a magnetic bead (via biotin-streptavidin interactions, see methods). The magnets placed above the chamber generate a magnetic force that pulls the bead up keeping the DNA stretched. By changing the height of the magnets we can control the applied force, the closer the magnets to the beads the greater the exerted force. The force typically ranges from 10^{-3} to 100 pN [10]. An important advantage of using MT, as compared to optical tweezers, is the parallelization, since we can track several beads at the same time obtaining a lot of statistics in a single experiment.

In the assays performed in this work we follow in real time the extension of the DNA molecule monitoring the changes in the bead position. The unzipping of the DNA hairpin molecule catalysed by the helicase lead to an increase on the DNA hairpin extension, allowing us to follow the unzipping reaction in real time. In other words, the DNA hairpin extension is an indirect measurement of the enzyme position along the DNA. From the analysis of experimental traces showing helicase activity we can investigate the enzyme motion (its velocity, diffusivity....). However, we cannot track the ATP hydrolysis reaction that is coupled to enzyme motion. In particular we do not know the number of ATPs that are hydrolysed or the mean number of base-pairs unzipped per ATP. And whether the mechano-chemical reaction is tightly or loosely coupled is also unknown. The question we try to address in this work is whether by using relations from non-equilibrium statistical physics, such as the fluctuation theorems (FTs) [11, 12] or the TUR [13–15] we can extract some information about the helicase chemical cycle from experimental traces of the motor motion.

FTs are a set of relations for the energy fluctuations in small systems under non-equilibrium conditions [11, 16]. They provide a deeper understanding of some physical laws that for small systems need to be revised, such as the second law of thermodynamics. While the second law predicts that the entropy of an isolated system should always increase, in small systems, the entropy can decrease in a finite time interval.

The Crook's FT [16] quantifies the probability of observing negative entropy production events relative to positive ones. Considering S the entropy produced in a non-equilibrium process during a time t we can write the simplest form of this FT as:

$$\frac{P_t(S)}{P_t(-S)} = e^{\frac{S}{k_B}}. \tag{1}$$

Where $P_t(S)$ is the probability distribution of entropy production S. Rewriting the expression as $P_t(-S) = P_t(S)e^{\frac{-S}{k_B}}$ and integrating both sites of the equality, we get: $1 = \langle e^{\frac{-S}{k_B}} \rangle$, which is the so-called Jarzynski equality [11]. Applying Jensen inequality ($\langle e^x \rangle > e^{\langle x \rangle}$) we deduce the second law of thermodynamics, $\langle S \rangle \geq 0$. The Jarzynski equality and the Crooks FT have been successfully applied to predict the free energy of molecular structures from non-equilibrium irreversible processes [16]. Different versions of FTs have been derived for different types of non-equilibrium processes [17].

Helicases work in non-equilibrium conditions generating thermodynamic costs known as dissipation, i.e, an average positive entropy production during the process. However, the helicase motion is highly affected by Brownian fluctuations, since the energies involved in the unzipping reaction are on the order of thermal energy k_BT. As a consequence, at short times, a particular helicase trajectory can have negative entropy production. The fluctuations and dissipation are related by the TUR through the Q_{TUR} factor [13, 15],

$$Q_{TUR} = \sigma \frac{2D}{v^2} \geq 2k_B, \tag{2}$$

where σ is the entropy production rate, v is the mean helicase unwinding rate and D is the helicase diffusion constant. σ/v is related to the dissipation and the ratio between the diffusion coefficient and the mean motor velocity ($\frac{2D}{v}$) is related to the fluctuations. If we assume a tight mechanochemical coupling of one base pair

unwound for each ATP consumed, the entropy production can be written as the product of the mean entropy production per unwound base-pair S_{bp} times v, and the Q_{TUR} factor reads as:

$$Q_{TUR} = S_{bp}\frac{2D}{v}. \tag{3}$$

In a previous work we analyze the TUR relation for the gp41 helicase unzipping reaction. As in experiments we did not measure the ATP hydrolysis chemical cycle we used Eq. 3 and assume the 1 base-pair tight coupling approximation to compute S_{bp}. From MT experimental traces we estimated D and v and computed the Q_{TUR} factor finding a large Q_{TUR} value (10^2-$10^3 k_B$). This suggest that the lower limit of $2k_B$ given by the TUR is very far from the Q_{TUR} values of real molecular systems and raises the question of whether the TUR inequality gives useful information. However, it could also be that the 1 base-pair tight coupling approximation is not valid for this motor and the value of Q_{TUR} factor we obtained is not correct.

In this work we investigate how to estimate the entropy production from our measurements using tools from non-equilibrium statistical mechanics. In order to address this question we study a simple model for helicase motion and compute the entropy production using FTs. As a first step, we model the helicase motion along a DNA chain as a biased random walk moving in a one dimensional lattice (which represents the DNA chain). Because the motor movement is a dynamic process that generates a steady state for long times (times longer than the characteristic step time of the motor), the entropy production should fulfill a FT. We then use this model to verify if the FT holds and how the Q_{TUR} factor behaves. In order to check our analytical results we also perform simulations of the biased random walk. Apart from validating the theoretical results, the simulations also allow us to compare the model with experimental results. The simplest simulation considers the random walk dynamics in an homogeneous chain without external fluctuations, where the helicase is a random walker with three constant probabilities P_+, P_- and P_0 (moving forward, backward or no move). With this simple model we can check the theoretical results and make a first approximation to the experimental ones. However, simulations can be easily extended to take into account multiple kinetic pathways, inhomogeneities in the DNA chain (such as the DNA sequence) or fluctuations of the media.

II. METHODS

A. MT experiments

In this work we study the hexameric helicase gp41 of the T4 bacteriophage [18] as a model helicase. We use MT to manipulate a DNA molecule and follow the movement of the enzyme through the DNA. The experiment has a temporal precision of few millisecond and a spatial precision of the nanometer. In the assays we used PicoTwist instrument (www.picotwist.com). We use a DNA hairpin with single-stranded tails that are labeled with digoxigenin and biotin to their extremities. The magnetic beads are coated with streptavidin so that can be attached to one of the DNA hairpin tails. On the other hand, we build a microfluidic chamber with a glass surface that is treated with anti-digoxigenin and passivated with a special buffer to prevent non-specific binding of biomolecules. By inserting the bead-hairpin construct into the chamber we obtain the configuration shown in Fig. 5A where the 3' end of the hairpin is attached to the surface and the 5' end to the magnetic bead. We illuminate the microfluidic chamber with a red LED and use and inverted microscope connected to a CCD camera (working at 300hz) to image the beads. The parallel and monochromatic illumination used generates diffraction patterns that allow performing 3D tracking of the beads in real time with nanometric resolution [19] .
The experiments were carried out using a 531 bp hairpin with a known sequence at 25$^{\text{o}}$C in a buffer containing 150 mM KAc, 10mM MgAc and 1 mM DTT. The helicase reaction is started adding 60 nM gp41 helicase at constant ATP concentration (4mM) and constant force (∼12pN). In Fig. 5 we show the experimental setup (panel A) and the gp41 unzipping traces (panels B-C). The experimental traces consist of a rising edge corresponding to the unzipping of the hairpin catalysed by the helicase until reaching the hairpin loop and a falling edge where the helicase follows translocating along the single-stranded DNA and the hairpin re-anneals in its back. Since the hairpin is always reformed after an helicase burst, many DNA unwinding events can be observed in a single DNA molecule (Panel B). Moreover, several beads (typically ∼ 30-50) are tracked simultaneously, which allows taking a lot of statistics. The measured extension (panel B) can be converted into number of unwound base pairs knowing that our hairpin has a length of 531 base pair (bp), as shown in panel C.

For each bead we have obtained several unwinding traces (∼200-400). From the analysis of these traces, we can extract kinetic information about the motor such as the mean velocity v or the diffusion coefficient D. To do so we take every trace and segment it into windows of variable size and compute the displacement. Then, we calculate the mean displacement and variance and plot it as a function of the time window (Fig. 6). Both the mean and variance present a lineal dependence with time. Therefore we perform lineal fits: $\langle x\rangle = vt$ and $\langle x^2\rangle = 2Dt$ to estimate v and D (see Fig. 6). The analysis has been done using a home-made Python program that allows us to detect the individual traces. Sometimes the helicase detaches before reaching the hairpin loop or show long pauses. For the analysis we only select full unwinding traces that reach the hairpin loop and do not show significant pausing.

B. Simulations

Experiments and theory are complemented with simulations of a biased random walk. We developed a Python program that produce trajectories $\{n_0, n_1, n_2...\}$, where n_i is the walker position at time M_i, starting at $n_0 = 0$, with the three parameters P_+, P_- and P_0 representing the probabilities of moving forward, backward or staying in the same point.
The algorithm works as follows:

1. Generate a random number between 0 and 1
2. Check if that number is bigger or lower than P_+ or $P_+ + P_-$ or $P_+ + P_- + P_0$
3. If it is lower than P_+ we add 1 to the position of the walker. If it is bigger than P_+, but lower than $P_+ + P_-$ we subtract 1 to the position of the walker. If it is bigger than $P_+ + P_-$ we add 0 to the position of the walker.
4. Repeat this process M times

With this algorithm we can simulate many trajectories. Due to stochasticity of the process each trajectory will end at different point N, generating an ensemble of trajectories. From these trajectories we will be able to compute the probability distribution of a given displacement N, for a fixed M, which we can compare with theoretical results and experiments described in the results section.

1. *Extended simulation*

In the simulations described above we consider the simplest model. However, in the the real experimental system, the movement of the helicase is affected by additional factors such as the specific sequence of the DNA. Moreover, the bead position which gives the experimental signal, has large Brownian fluctuations. If we want to consider a more realistic description of the bead motion we can consider that the position of the bead (x) follows a Langevin dynamics of an overdamped particle with evolution given be:

$$\gamma \frac{dx}{dt} = F - k_{DNA} x + \xi, \tag{4}$$

where we have modeled the DNA as an elastic spring with stiffness k_{DNA}, and F is the external applied force. The noise (ξ) is a white noise with zero mean and variance $2k_BT\gamma$, being γ the drag coefficient of the bead.

On the other hand, because different base-pairs have different energy of stabilization (going from 1 to 4 k_BT) [9], the probabilities of stepping forward (and unzipping one base-pair) might be affected by the DNA sequence. In overall, sequence and fluctuations modify the probabilities which are no longer constant, but depend on the force, position of motor and position of the bead, making the simulation more complex. For a more realistic description of the motor motion in the experimental conditions we define the probabilities of stepping one bp forward and backwards per unit time as the rates k_+ and k_- given by:

$$k^+(n, F) = k_0 \exp\left(\frac{-\Delta G_{bp}(n) + W_{net}(F)}{k_B T}\right), \tag{5}$$

$$k^-(n) = k_0 \exp\left(\frac{-\Delta\mu}{\delta k_B T}\right), \tag{6}$$

where k_0 is an attempt frequency, $\Delta G_{bp}(n)$ is the free energy of formation of the n base-pair, $\Delta\mu$ is the energy coming from the ATP hydrolysis, δ is the motor step size and $W_{net}(F)$ is the net work done by the force. These rates verify the detailed balance condition, since the total change in free energy between two consecutive positions is given by $\Delta G_T = \Delta\mu/\delta - \Delta G_{bp} + W_{net}$. To calculate the mechanical work W_{net}, we consider the DNA as a polymer and we used the worm-like chain model (WLC) which describes the behaviour of polymers with a force-extension relation [20] characterized by a persistence length (P) and a contour length (L). The persistence length characterizes the stiffness of a polymer (for lengths shorter than P the polymer acts as a rigid rod) whereas the contour length, is the length of the polymer when its completely elongated.

The ATP dependence that influences the progress of the helicase can also be considered. If the helicase mechano-chemical coupling is tight, for each ATP hydrolyzed the motor takes an step of δ bps forward and the mean velocity increase as the ATP concentration increases following a Michaelis-Menten like curve. Soft coupling can be introduced by adding an extra kinetic rate for the loose coupling. Another complexity that can be added in the simulations is the presence of different kinetic pathways that lead to pausing or enzyme dissociation. The presence of different competing kinetic pathways has been observed for different helicases.

C. FT in stochastic systems

We consider an arbitrary process in a system in contact with a thermal bath at temperature T. Initially at time t=0, the system is in equilibrium with $P^{eq}_{\lambda_i}(n) = \frac{e^{-\beta E_{\lambda_i}(n)}}{Z_{\lambda_i}}$, being $\beta = \frac{1}{k_B T}$. Then a perturbation parameterized by λ takes the system along a path towards a different state [17]. A trajectory or path Γ in configurational space is described by a discrete sequence of configurations in phase

space, $\Gamma \equiv \{n_0, n_1 ..., n_M\}$, where the system occupies configuration n_k at time $t_k = k\Delta t$, being Δt the discrete time step. We consider paths that start at $n_0 = 0$ at time $t = 0$ and end at $n_M = M$ at time $t = M\Delta t$. The continuous time limit is recovered by taking $M \to \infty$ and $\Delta t \to 0$. The control parameter(λ) is a value or set of values that characterise the external perturbations that drive the evolution of state of the system, for example, the magnetic field, the concentration of ATP or, in the case of single molecule experiments with optical tweezers, the distance between the trap and the bead.
We can compute the energy difference in the time interval k as,

$$\Delta E_k = E_{\lambda_{k+1}}(n_{k+1}) - E_{\lambda_k(n_k)} = W_k + Q_k. \tag{7}$$

Where we have defined work as the change in energy due to the change in the control parameter (λ) and heat as the change of energy due to a change in configuration maintaining λ constant:

$$\begin{aligned} Q &= \sum_{k=0}^{M-1} (E_{n_{k+1}}^{\lambda_{k+1}} - E_{n_k}^{\lambda_{k+1}}), \\ W &= \sum_{k=0}^{M-1} (E_{n_k}^{\lambda_{k+1}} - E_{n_k}^{\lambda_k}), \\ E &= Q + W = E_{n_M}^{\lambda_M} - E_{n_0}^{\lambda_0}. \end{aligned} \tag{8}$$

The path Γ can be written in terms of conditional probabilities as, $P(\Gamma) = \prod_{k=0}^{M-1} P_{\lambda_k}(n_{k+1}|n_k)$. Where the transition probabilities must satisfy detailed balance in order to guarantee that the equilibrium state ($P_\lambda^{eq}(x)$) is stationary,

$$\frac{P_{\lambda_k}(n_{k+1}|n_k)}{P_{\lambda_k}(n_k|n_{k+1})} = \frac{P_{\lambda_k}^{eq}(n_{k+1})}{P_{\lambda_k}^{eq}(n_k)} = e^{-\beta(E_{\lambda_k}(n_{k+1}) - E_{\lambda_k}(n_k))}. \tag{9}$$

From this relation we can write,

$$\begin{aligned} P(\Gamma) &= \prod_{k=0}^{M-1} P_{\lambda_k}(n_{k+1}|x_k) = \\ &= \prod_{k=0}^{M-1} P_{\lambda_k}(n_k|n_{k+1}) e^{-\beta(E_{\lambda_k}(n_{k+1}) - E_{\lambda_k}(n_k))} = \\ &= P(\Gamma^*) e^{-\beta Q_k}. \end{aligned} \tag{10}$$

where $\Gamma^* = \{n_M, n_{M-1}, \ldots, n_0\}$ is the reverse path of Γ. Finally, rearranging terms we obtain,

$$\frac{P(\Gamma)}{P(\Gamma^*)} = e^{-\beta Q(\Gamma)}. \tag{11}$$

Which has the form of a FT for the heat. If we want to include information about the work (W) we have to consider the probability of exerting some work into the system,

$$P(W) = \sum_\Gamma P(\Gamma) P_{\lambda_0}^{eq}(n_0) \delta(W - W(\Gamma)), \tag{12}$$

where $\delta(W - W(\Gamma)) = \delta_W$ only takes into account the trajectories in which some work has been exerted. By using Eq. (11) and remembering $P_{\lambda_0}^{eq}(n_0) = \frac{e^{-\beta E_{\lambda_0}(n_0)}}{Z_{\lambda_0}}$ with $Z_\lambda = e^{-\beta F_{\lambda_i}}$ we have,

$$\begin{aligned} P(W) &= \sum_\Gamma P(\Gamma^*) P_{\lambda_0}^{eq}(x_0) e^{-\beta Q} \delta_W = \\ &= \sum_\Gamma P(\Gamma^*) \frac{e^{-\beta E_{\lambda_0}(n_0)}}{Z_{\lambda_0}} e^{-\beta Q} \delta_W, \end{aligned} \tag{13}$$

multiplying and dividing by $e^{\beta E_{\lambda_M}(n_M)}$ we get,

$$P(W) = \sum_\Gamma P(\Gamma^*) \frac{e^{\beta(Q+W)}}{Z_{\lambda_0}} e^{-\beta Q} \delta_W e^{-\beta E_{\lambda_M}(n_M)}, \tag{14}$$

where $e^{\beta(E_{\lambda_M}(n_M) - E_{\lambda_0}(n_0))} = e^{\beta(Q+W)}$ and taking $e^{-\beta E_{\lambda_M}(n_M)} = Z_{\lambda_M} P_{\lambda_M}^{eq}(n_M)$

$$P(W) = \frac{Z_{\lambda_M}}{Z_{\lambda_0}} \sum_\Gamma P(\Gamma^*) P_{\lambda_M}^{eq}(n_M) e^{\beta W(\Gamma)} \delta_W. \tag{15}$$

Analogously, as in Eq. (12) the reverse probability is

$$P^*(W) = \sum_{\Gamma^*} P(\Gamma^*) P_{\lambda_M}^{eq}(x_M) \delta(W - W(\Gamma^*)), \tag{16}$$

using the relation $P_{\lambda_M}^{eq}(x_M) = e^{\beta W} \frac{Z_{\lambda_M}}{Z_{\lambda_0}} P^*(-W)$ we arrive at

$$\frac{P(W)}{P(-W)} = e^{\beta(W - \Delta F)}. \tag{17}$$

This is the Crooks FT for work, where $\Delta F = F_{\lambda_M} - F_{\lambda_0}$ is the difference of Helmholtz free energy ($F = -k_B T \log(Z)$) between the initial and final state and $W - \Delta F = W_d$ is the dissipated work in the forward process. $P^*(-W)$ is exponentially suppressed relative to $P(W)$.
According to Eq. (17), we can find a value for $W = \Delta F$ looking at the crossing point of the forward ($P(W)$) and reverse ($P(-W)$) work distribution, independly of how far the system is driven out of equilibrium.
Rewriting Eq. (17) as $P^*(-W) = e^{-\beta(W - \Delta F)} P(W)$ and integrating over W we get $1 = \int_{-\infty}^{\infty} dW P(W) e^{-\beta(W - \Delta F)} = \langle e^{-\beta(W - \Delta F)} \rangle$ obtaining:

$$\Delta F = -k_B T \log(\langle e^{-\beta W} \rangle). \tag{18}$$

This result, known as the Jarzynski equality [11], relates the free energy difference between states A and B with the irreversible work carried out when moving the system from A to B along a given protocol. In the case of an infinitely slow protocol, the work exerted on the system is the equilibrium one, obtaining: $\Delta F = W$. Jarzynski equality and FTs are useful tools from non-equilibrium physics that go beyond classical thermodynamics connecting equilibrium information with non-equilibrium measurements. These relations give the possibility of recovering free energy differences from measurments of out-of-equilibrium irreversible processes [21, 22].

III. RESULTS

A. Random walk

Here we model the helicase motion as a random walk. We consider the movement of the helicase through a DNA chain as a random walker moving in a one dimensional lattice taking steps forward and backwards with probabilities P_+ and P_-.

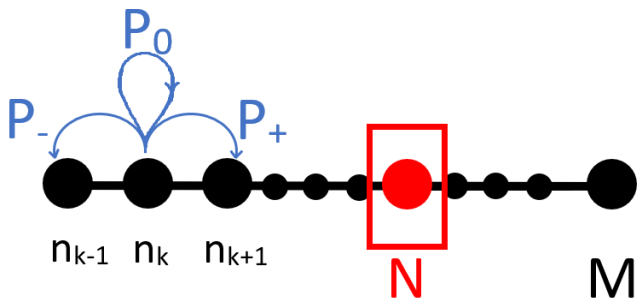

FIG. 1. Schematics of the one-dimensional random walk for the helicase motor. The motor moves forward and backwards with probabilities P_+ and P_- respectively. The probability of staying at the same place is defined by normalization: $P_0 = 1 - P_+ - P_-$. In red is marked the final position N of the motor after taking M steps.

At each time step ΔM the walker moves a distance Δn or stays in the same place. This latest possibility is accounted in the model with the probability P_0 that takes into account the waiting time between steps. After M steps the walker has done n_+ steps forward, n_- step backward and $\hat{n}$ null steps. For simplicity and without loss of generality we take $\Delta M = 1$ and $\Delta n = 1$. Note that taking $\Delta n = 1$ means that the motor displacement n would be in units of the motor step size δ, which is defined as the number of base-pairs the enzyme moves (and unzips) in an elementary step of the mechano-chemical reaction (e.g the hydrolysis of a single ATP molecule). Many helicases hydrolyze one ATP per bp unwound, $\delta = 1$ bp, but others have an step size of few base-pairs $\delta = 2 - 4$ bp [23]. The position n of the walker after $M + 1$ steps can be obtained in three ways from the M: Start at n and stay there, from $n - 1$ take a step forward or from $n + 1$ take a step backwards. Therefore the probability of being at n after $M + 1$ steps can be expressed as:

$$p(n, M+1) = P_0 p(n, M) + P_+ p(n-1, M) + P_- p(n+1, M). \tag{19}$$

For long times, this model describes the non-equilibrium steady state (NESS) of an ideal motor. In the following we will use this model to compute the entropy production during motor activity and test whether the FT for NESS holds.

B. FT for a random walk

Let us consider a stochastic system describing a trajectory $\Gamma \equiv \{n_0, n_1 ..., n_M\}$. We define $P_k(n_k)$ as the probability measured that the system is at time k in position n_k (see methods). We will focus on the trajectory described by a random walker going forward, backward or staying in the same location with probability P_+, P_- and P_0 respectively. We assume that the dynamics of the system is Markovian, i.e, the probability of the system being in a given configuration at any given time only depends on its previous configuration. Considering that our external control parameter, the ATP concentration (λ=[ATP]), is fixed, and setting the starting point at $n_0 = 0$ we can write:

$$P(\Gamma) = P_0(n_0) \prod_{k=0}^{M-1} P(n_{k+1}|n_k). \tag{20}$$

The expected value of any observable $A(\Gamma)$ is given by,

$$\langle A \rangle = \sum_{\Gamma} A(\Gamma) P(\Gamma), \tag{21}$$

where summing over Γ is equivalent of summing over every possible state n_k of the motor.

$$\langle A \rangle = \sum_{n_0 \dots n_M} A(\Gamma) P_0(n_0) \prod_{k=0}^{M-1} P(n_{k+1}|n_k), \tag{22}$$

by applying detailed balance (9) we have,

$$\langle A \rangle = \sum_{n_0 \dots n_M} A(\Gamma) P_0(n_0) \prod_{k=0}^{M-1} P(n_k|n_{k+1}) \frac{P^{eq}(n_{k+1})}{P^{eq}(n_k)}. \tag{23}$$

Lets consider $A = e^{-S_t^*(\Gamma)} = \frac{b(p_M)}{P_0(n_0)} \prod_{k=0}^{M-1} \frac{P^{eq}(n_k)}{P^{eq}(n_{k+1})}$. Where $b(p_M)$ is a positive normalized function ($\sum b(p_M) = 1$):

$$\langle e^{-S_t^*(\Gamma)} \rangle = \sum_{n_0 \dots n_M} b(p_M) \prod_{k=0}^{M-1} P(n_k|n_{k+1}) = 1. \quad (24)$$

Where we have applied a telescopic sum, we first summed for every $b(p_M)$ and then over the rest of variables using $\sum P(n_k|n_{k+1}) = 1$. We can define the total dissipation (S_t^*) as,

$$S_t^* = \sum_{k=0}^{M-1} \log(\frac{P^{eq}(n_{k+1})}{P^{eq}(n_k)}) + \log(P_0(n_0)) - \log(b(p_M)). \quad (25)$$

Applying Eq. (9) we get,

$$S_t^* = \sum_{k=0}^{M-1} \log(\frac{P(n_{k+1}|n_k)}{P(n_k|n_{k+1})}) + \log(P_0(n_0)) - \log(b(p_M)). \quad (26)$$

In the latest equation we have a boundary term ($\log(P_0(n_0)) - \log(b(p_M))$) and we can define the entropy production of the system as,

$$S = \sum_{k=0}^{M-1} \log(\frac{P(n_{k+1}|n_k)}{P(n_k|n_{k+1})}). \quad (27)$$

This expression is equivalent to the one obtained for heat using the detailed balance condition and the definitions in Eq. (8) divided by β,

$$Q = \sum_{k=0}^{M-1} (E_{n_{k+1}} - E_{n_k}) = -\frac{1}{\beta} \sum_{k=0}^{M-1} \log(\frac{P(n_{k+1}|n_k)}{P(n_k|n_{k+1})}). \quad (28)$$

Therefore the expected value of the exponential of the entropy production $A = e^{-S} = e^{\beta Q}$ can be computed as,

$$\begin{aligned} \langle A \rangle = \langle e^{\beta Q} \rangle &= \sum_{n_0, n_1 \dots n_M} e^{\beta Q} P_0(n_0) \prod_{k=0}^{M-1} P(n_{k+1}|n_k) = \\ &= \sum_{n_0=0, n_M=N} e^{\beta Q} \sum_{n_1 \dots n_{M-1}} P_0(n_0) \prod_{k=0}^{M-1} P(n_{k+1}|n_k) = \\ &= \sum_{n_0=0, n_M=N} e^{\beta Q} \phi_M(N), \end{aligned} \quad (29)$$

where we have defined $\phi_M(N) = \sum_{n_1 \dots n_{M-1}} P_0(n_0) \prod_{k=0}^{M-1} P(n_{k+1}|n_k)$. Taking into account that in the sum appearing in Eq. (29) N can take values in range of -M to M we can write the previous expression as:

$$\langle e^{\beta Q} \rangle = \sum_{N=-M}^{M} e^{\beta Q} \phi_M(N). \quad (30)$$

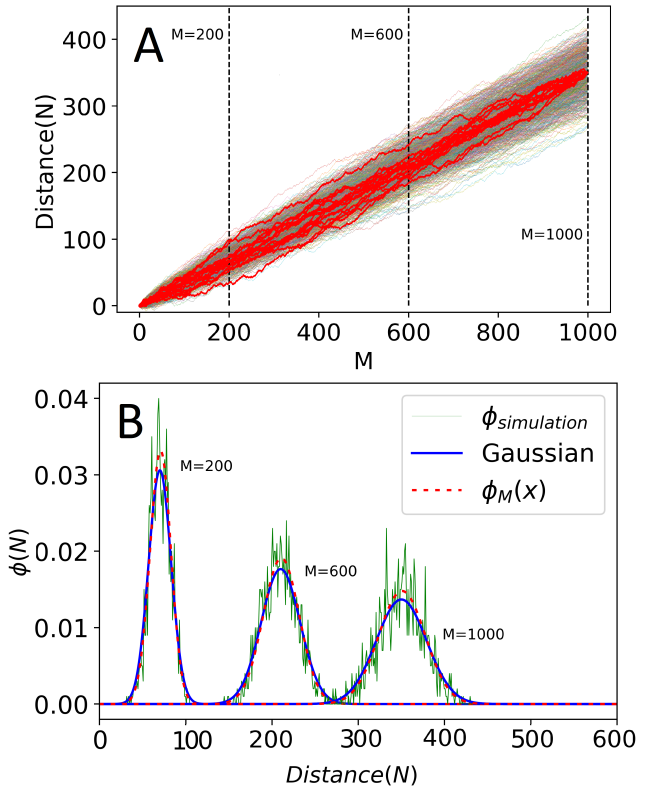

FIG. 2. **(A.)** A 1000 simulated trajectories of the random walk model with probabilities: $P_+ = 0.60, P_- = 0.25, P_0 = 0.15$. The trajectories that end at a particular position N=350 are shown in red. The probability ϕ at N=350 is computed as the ratio between the number of red trajectories and the total number of trajectories. In this case $\frac{\#red}{\#total} = 0.017$ that coincides with the theoretical ϕ value given by Eq. (31), $\phi = 0.016$. **(B.)**. Probability distribution of observing a displacement N in M steps, $\phi_M(N)$, computed theoretically (Eq. (31)) and from simulations for three different M. In red we plot the values obtained using Eq. (31), in blue the Gaussian approximation obtained in appendix B and in green the values obtained from simulations.

In order to proceed our calculation we need to compute $\phi_M(N)$, which is the probability of reaching the position N after M steps. We have performed this computation for a non-biased random walk using the Gaussian approximation for Ising-like model (see Appendix A) and for a biased one-dimensional random walk using the large fluctuations approach (see appendix B) obtaining:

$$\phi_M(N) = \frac{(\frac{P_0}{P_-})^N \frac{1}{\sqrt{a''(N)}} e^{Ma(N)}}{\sum_{N=-M}^{M} (\frac{P_0}{P_-})^N \frac{1}{\sqrt{a''(N)}} e^{Ma(N)}}, \tag{31}$$

where $a(N)$ is a continuous function defined in Eq. (58) and its second derivative $a''(N)$ is computed in Eq. (61). In order to check this result we performed simulations of the random walk model presented (Methods section). In Fig. 2 we show a set of simulated trajectories $n(M)$ (panel A) and the fraction $\phi_M(N)$ (panel B) computed from simulations. The results are compared to the theoretical expressions finding good agreement.

Now we need to compute $e^{\beta Q}$. For our model, using the definitions in Eq.8, we find $Q = E$ (fixed λ implies $W = 0$). Considering that energy at step k is proportional to the initial energy E_1 we can obtain a value for Q as, $Q = E = NE_1$. To find E_1 we apply the detailed balance condition, Eq. 9, obtaining, $E_1 = -\frac{1}{\beta}\log(\frac{P_+}{P_-})$ and therefore $e^{\beta Q} = (\frac{P_-}{P_+})^N$. With this result we can write,

$$\langle e^{\beta Q} \rangle = \frac{\sum_{N=-M}^{M} (\frac{P_0}{P_+})^N \sqrt{\frac{1}{a''(N)}} e^{Ma(N)}}{\sum_{N=-M}^{M} (\frac{P_0}{P_-})^N \sqrt{\frac{1}{a''(N)}} e^{Ma(N)}}. \tag{32}$$

This expression for $< e^{\beta Q} >$ can be greatly simplified considering the symmetry: $P_+ \leftrightarrow P_-$, $N \leftrightarrow -N$ where:

$$\begin{aligned} e^{Ma(-N)} &= e^{Ma(N)} \left(\frac{P_+ P_-}{P_0^2}\right)^{-N}, \\ a''(-N) &= a''(N). \end{aligned} \tag{33}$$

Imposing these relations we find $\phi_M^{P_+,P_-}(N) = \phi_M^{P_-,P_+}(-N)$ and also

$$\langle e^{\beta Q} \rangle = 1. \tag{34}$$

This result is equivalent to the expression (24) and it corresponds to the Jarzynski equality. Using Eq. (31-33) we find the following symmetry relation for $\phi_M(N)$:

$$\frac{\phi_M(N)}{\phi_M(-N)} = (\frac{P_+}{P_-})^N. \tag{35}$$

Using Eq. (28) it can be proved that $(\frac{P_+}{P_-})^N = e^{-\beta Q}$ and Eq. 35 can be written as:

$$\frac{\phi_M(N)}{\phi_M(-N)} = e^{-\beta Q} = e^S, \tag{36}$$

which is a form of Crook's FT for the entropy production Q. Eq. (31) has been obtained for large M using the saddle point approximation. In the following we will perform the calculation without the large M approximation obtaining the same results. Let's consider the probability of finding the walker at position N after M steps in one trajectory as,

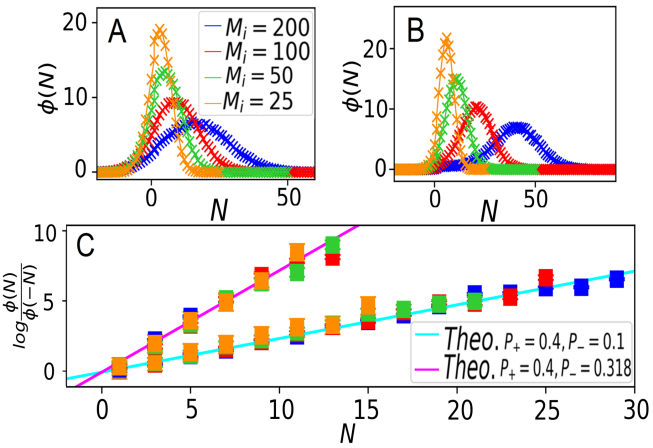

FIG. 3. **(A. and B.)** Displacement distributions $\phi_M(N)$ at four different M: $M_i = 200$ (blue), $M_i = 100$ (red), $M_i = 50$ (green) and $M_i = 25$ (yellow) for simulations of the random walk model using probabilities $P_+ = 0.4$, $P_- = 0.318$ and $P_0 = 0.282$ (panel A) and $P_+ = 0.4$, $P_- = 0.2$ and $P_0 = 0.5$ (panel B) **(C.)** Test of the FT for the random walk model. In magenta and cyan we show the theoretical predictions of the FT (Eq. 35) for the probabilities shown in panels A and B respectively, corresponding a linear behaviour with slope $\log\frac{P_+}{P_-}$ =0.23 and 1.38 respectively. The squares are values obtained from simulations for the ratio $\frac{\phi_{M_i}(N)}{\phi_{M_i}(-N)}$ for different M_i. Colors as in panel A.

$$\phi_M(N) = \frac{1}{\Omega_N} \sum_{\substack{M=n_+ + n_- + \hat{n} \\ N=n_+ - n_-}} P_+^{n_+} P_-^{n_-} P_0^{\hat{n}}, \tag{37}$$

where Ω_N is a normalization constant. Multiplying by $\frac{P_+^{n_-} P_-^{n_+}}{P_+^{n_-} P_-^{n_-}}$ and arranging terms we find:

$$\phi_M(N) = \frac{1}{\Omega_N} \sum_{\substack{M=n_+ + n_- + \hat{n} \\ N=n_+ - n_-}} P_+^{n_-} P_-^{n_+} P_0^{\hat{n}} (\frac{P_+}{P_-})^N. \tag{38}$$

Analogously we can write:

$$\phi_M(-N) = \frac{1}{\Omega_N} \sum_{\substack{M=n_+ + n_- + \hat{n} \\ N=n_+ - n_-}} P_+^{n_-} P_-^{n_+} P_0^{\hat{n}}. \tag{39}$$

These functions verify the relation found in Eq. (35). Eq. 35 tell us that if we plot the logarithm of $\frac{\phi_M(N)}{\phi_M(-N)}$ as

a function of N we should find a straight line with slope given by the logarithm of the ratio between probabilities. To check this result we use the simulations to compute $\phi_M(N)$ and verify the FT (Fig. 3).

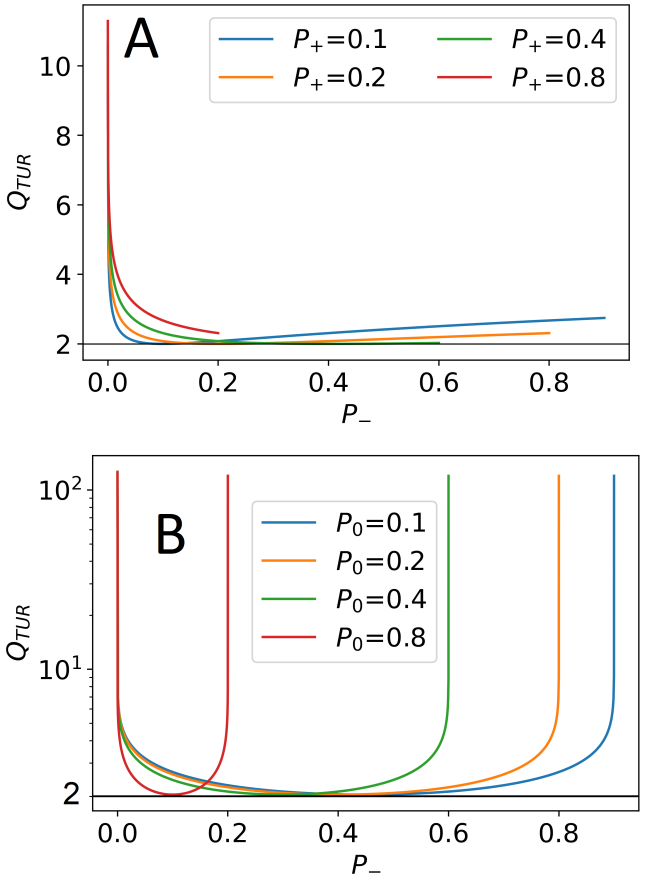

FIG. 4. **(A.)** Q_{TUR} factor for the random walk model, Eq. 41, as a function of P_- for different P_+ values. **(B.)** Q_{TUR} factor computed using Eq. 41 as a function of P_- for different P_0 values. In both representations the lower limit $Q \geq 2$, shown as a black line, is satisfied.

As previously discussed in our model $S = -\beta Q$. Therefore the entropy production rate can be expressed as,

$$\sigma = \frac{\langle S \rangle}{M} = \frac{\langle N \rangle}{M} \log \frac{P_+}{P_-}, \tag{40}$$

where $\langle N \rangle$ is the mean position given by $\langle N \rangle = vM = (P_+ - P_-)M$ (the relation $v = (P_+ - P_-)$ has been derived in Appendix B). Using the expression of the entropy production rate we can compute the Q_{TUR} factor (Eq. 2), as:

$$Q_{TUR} = \sigma \frac{2D}{v^2} = \frac{P_+ + P_-}{P_+ - P_-} \log \frac{P_+}{P_-}. \tag{41}$$

Note that Q_{TUR} only depends on the forward and backward probabilities. Taking into account the probability normalization: $P_+ + P_- + P_0 = 1$, we can rewrite the previous expression as,

$$Q_{TUR} = \frac{w}{2x - w} \log \frac{x}{w - x}, \tag{42}$$

where $x = P_+$ and $w = 1 - P_0$. Taking the derivative of Eq. 42 and equaling it to zero we find the extreme value at $x_{min} = \frac{w}{2}$. We find $Q_{TUR}(x_{min}) = 2$, coinciding with the lower limit found in [15]. In Fig. 4 we show the numerical results for Q_{TUR} for different values of the probabilities P_+, P_- and P_0. We can identify that Q_{TUR} increases fastly when we approach to the unidirectional case, where $P_+ = 0$ or $P_- = 0$.

C. Experimental results

We used MT to manipulate a 531 base-pair DNA hairpin and monitor the DNA unzipping activity of the helicase gp41. The hairpin is tethered between a magnetic bead and the glass surface of the microfluidic chamber where the experiments take place (Fig. 5). Changing the position of the magnets we can control the force applied. In the experiments presented the force is kept fixed (the magnets are always at the same height). By tracking the position of the bead we can indirectly track the position of the molecular motor along the hairpin. Initially the DNA hairpin is formed and the bead is close to the surface, at height h_1 (see Fig. 5A). When the helicase starts unzipping the hairpin, the DNA lengthens, gaining approximately 1 nm per base-pair unzipped, until reaching the maximum extension $h_3 > h_1$, when the helicase has fully unzipped the hairpin. This unzipping activity is observed in the experimental traces as bursts of extension increase (Fig. 5B). As observed in Fig. 5 C the experimental trace corresponds to a triangular signal, with a rising edge corresponding to the DNA hairpin unzipping catalysed by the helicase followed by a falling edge. The latter corresponds to the motion of the helicase translocating along the unwound single-stranded DNA after reaching the loop, whereas the hairpin reforms in its wake. Note that the slope of the rising edge (related to the unwinding rate of the helicase) is lower than that of the falling edge (related to the translocation rate of the helicase). This means that this helicase translocates slower when it has to unzip the DNA double-helix ahead than when it moves along the single-stranded DNA (without unzipping). This property has been associated with the passive character of this helicase [24]. Here we will focus our analysis on the unzipping part of the trace (rising edge).

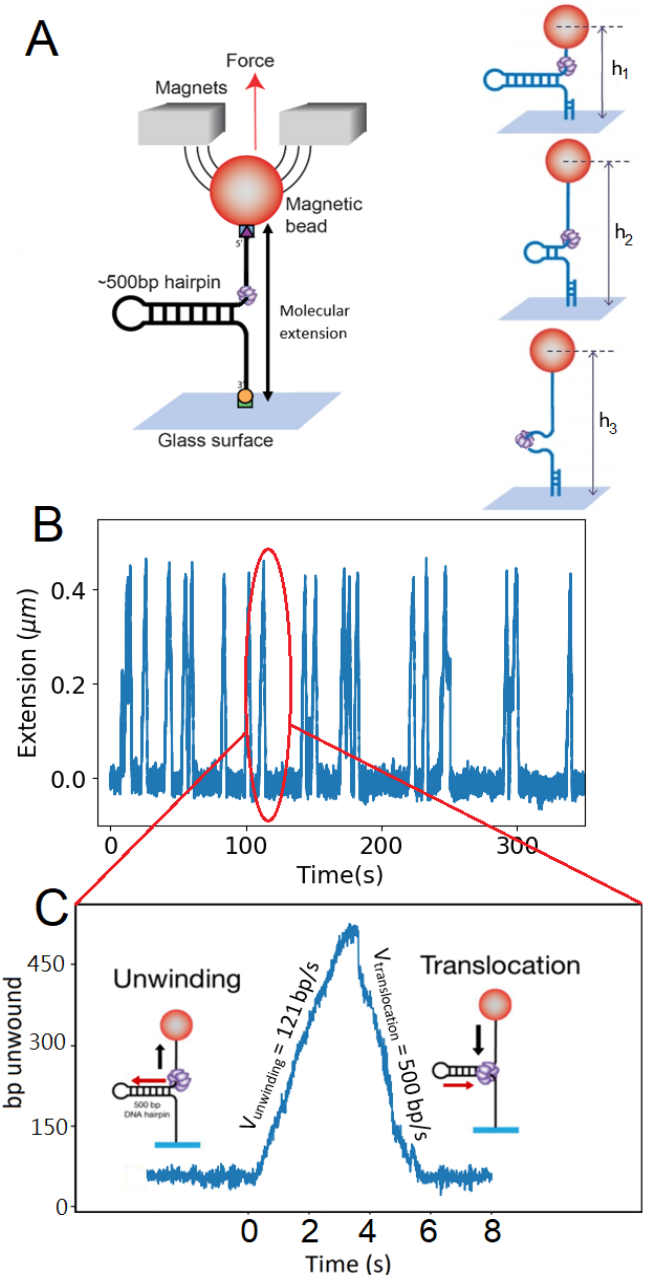

FIG. 5. **(A.)** Schematic representation of the magnetic tweezers experimental set up. The helicase (in purple) unwinds the hairpin going from the initial (closed) configuration at h_1 to the final (opened) configuration h_3. **(B.)** An example of DNA extension versus time data showing different helicase unwinding events. **(C.)** Experimental unwinding and rewinding trace. The full unwinding of the hairpin catalysed by the helicase is followed by the hairpin rezipping while the helicase translocation on the unwound DNA. Red arrows indicate helicase motion and black arrows indicate bead motion. The extension in μm (panel B) has been converted in number of unwound bps (panel C) by using that the maximum extension corresponds to the length of the hairpin (531bps).

With MT we can track several beads in the same experiment obtaining large statistics for the data analysis. However, the force exerted on beads may vary from bead to bead due to differences in the bead's magnetic moments and inhomogeneities in the magnetic field created by the magnets. Therefore, to avoid mixing traces at different forces, we will analyse the beads separately.
We developed a home-made python code to isolate the unwinding traces from the data. The traces obtained have a high degree of stochasticity (Fig. 6A), a typical signature of single-molecule data. Analyzing several unwinding traces we can compute the mean unwinding rate and the diffusion coefficient (see methods) for the helicase, as shown in Fig. 6B-C. For the data shown in Fig. 6 we obtain a mean unwinding of $v = 121 \pm 2$ bp/s and a diffusion coefficient of $D = 522 \pm 10\text{bp}^2/\text{s}$.

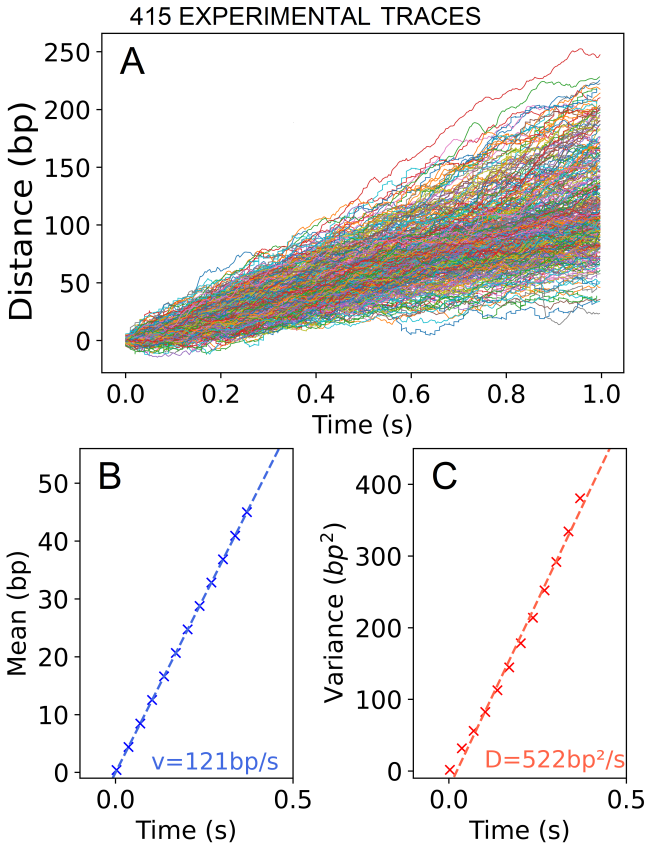

FIG. 6. **(A.)** Experimental helicase unwinding traces. **(B.)**. Mean unwinding rate obtained from the lineal fit to the mean displacement as a function of time computed from experimental traces in panel A. **(C.)** Diffusion coefficient obtained from the lineal fit to the variance of the displacement as a function of time computed from experimental traces in panel A.

Now we want to see if we can match the experimental results with the simple random walk model studied in previous section. In order to do it, we use the relation found in appendix B between the velocity and the diffusion coefficient and the forward and backward probabilities: $v = (k_+ - k_-)\Delta x$ and $D = \frac{k_+ + k_-}{2}(\Delta x)^2$ where k_+ and k_- are the forward and backward rates (proba-

bility per unit of time ($P_+/\Delta t$ and $P_-/\Delta t$) and Δx is the extension unit, in our case $\Delta x = 1$ bp. Using values for the probabilities that match the experimental v and D, we can try to reproduce the experimental results.

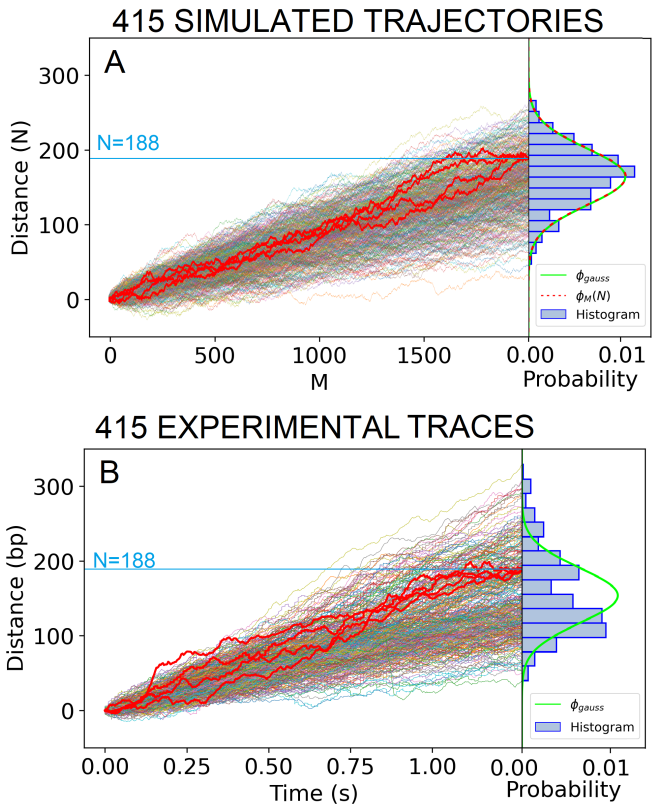

FIG. 7. **(A.)** Trajectories from simulations of the random model using P_+ and P_- values that fit the experimentally measured v and D (Fig. 6B-C): $P_+ = 0.4$ and $P_- = 0.318$. In red we plot the traces that ends at N=188. The ratio between number of red traces and the total number of traces is 0.009 which approximately coincides with the theoretical prediction $\phi = 0.01$, Eq. (31). On the right panel we show the histogram of the displacements N at the final M together with the theoretical $\phi_M(N)$ from Eq. (31) in red discontinuous line, and the Gaussian fit in a green line. **(B.)** Experimental unwinding traces, in red those that ends at x=188 bps. On the right panel we show the histogram of the helicase displacement after 1.25 seconds and the theoretical predicted Gaussian distribution (see appendix B) with the mean and variance measured in the experiments.

As shown in Fig. 7A we simulate motor trajectories using the random walk model (see methods) with P_+ and P_- values that reproduce the experimental v and D. In Fig. 7 we compare the simulated results with one set of experimental unzipping traces obtained for one bead. One can see that the traces are qualitatively similar. However, if we look at the distribution of the displacement probability, $\phi(x)$, some differences can be detected. Whereas $\phi(x)$ obtained from simulations follows a Gaussian distribution with mean vt and variance $2Dt$ as predicted by the theory (Eq. 67), the experimental $\phi(x)$ shows clear deviations from the expected Gaussian behavior. In particular, the experimental distribution present much larger tails, a typical characteristic of small systems. These deviations indicate that the model used is too simple to fully reproduce the experimental phenomenology. Introducing in the model different features of the real experimental system, such as the heterogeneity on the DNA sequence, the ATP chemical cycle with a single or multiple pathways, etc, we could investigate what features are responsible for the enlargement of the tails of the experimental distributions.

We next have investigated if we can apply the FT, Eq. 35, to our experimental data. To this end, we have computed the experimental distribution $\phi(x)$ for different times windows (Fig. 8A) and compute the logarithm of the ratio between $\phi(x)$ and $\phi(-x)$. We have found that the data collapse in a single line, as predicted by the FT (Eq. 35), Fig. 8B. The theory also predicts that the slope of the line should be related to the mean velocity v and the diffusion coefficient D as:

$$\log\left(\frac{\phi(x)}{\phi(-x)}\right) = x \log\left(\frac{\frac{2D}{\Delta x} + v}{\frac{2D}{\Delta x} - v}\right), \tag{43}$$

where $\Delta x = 1$ bp. This relation is not fulfilled in the experimental case, since the slope Eq. 43 using the values obtained experimentally for v and D leads to $\log(\frac{2D+v}{2D-v}) = 0.19$ (blue line in Fig. 8B) and the experimental slope using the FT is ~ 0.33 (black line in Fig. 8B). This is also an indication that the experimental system is more complex than a random walker moving in an homogeneous chain.
Since the probabilities P_+ and P_- should fulfil detailed balance, we have that the FT slope is related to the free energy difference $\Delta G = G(n) - G(n+1)$ between the states n and $n+1$, :

$$\frac{P_+}{P_-} = e^{\frac{\Delta G}{k_B}}. \tag{44}$$

Since n (or extension x) is in units of base-pairs the free energy ΔG corresponds to the free energy difference in 1 bp unzipping step given by: $\Delta G = W_F - \Delta G_{bp} + \Delta G_{ATP}$, where ΔG_{bp} is the energy of formation of a base-pair ($\sim 2k_BT$), W_F is the mechanical work done by the applied force when unzipping a single base-pairs ($\sim 1k_BT$ in our conditions) and ΔG_{ATP} is the energy associated to the ATP hydrolysis for a single base-pair movement. The energy associated with the hydrolysis of a ATP molecule, at the experimental conditions, is $\Delta\mu \sim 20k_BT$ [25]. If δ is the mechano-chemical step size, that is the number of bps unzipped per ATP hydrolyzed, then $\Delta G_{ATP} = \Delta\mu/\delta$. The logarithm of the detailed balance condition, Eq. 44, can be compared with the measured slope in Fig. 8B to extract ΔG.

As we can estimate the value of $W_F - \Delta G_{bp}$ at our experimental conditions ($\sim 1 k_B T$) we can use the measured slope to extract the value of ΔG_{ATP}, obtaining a value of $1.3 k_B T$, leading to an estimated mechano-chemical step size δ of 15 bps. This value is much larger than the typical measurements of helicase step sizes, which are in the range of 1-4 bps [23]. This results reflects again that we are probably lacking some important features in our modeling.

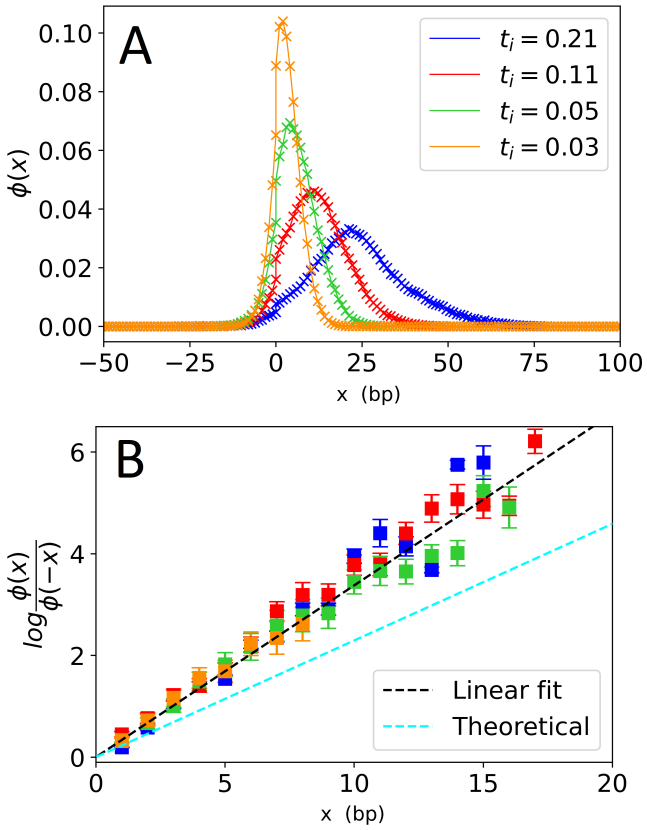

FIG. 8. **(A.)** Experimental distribution of displacement x computed from the experimental traces shown in Fig. (7B) at different times. **(B.)** The logarithm of the ratio of the probabilities $\frac{\phi_{t_i}(x)}{\phi_{t_i}(-x)}$ as a function of x for the data shown in panel A. The lineal fit to all data, shown in black, gives an slope of 0.33. The theoretical value obtained from $log(\frac{P_+}{P_-}) = 0.23$

Finally we compute the value of the Q_{TUR} factor using the values of D and v calculated previously. The expression for Q_{TUR} in terms of the velocity and diffusion coefficient reads as:

$$Q_{TUR} = \frac{2D}{v\Delta x} \log\left(\frac{\frac{2D}{\Delta x} + v}{\frac{2D}{\Delta x} - v}\right), \tag{45}$$

obtaining a value of $Q_{TUR} \approx 2$ (in k_B units) with $\Delta x = 1$ bp. This would indicate that the motor dissipates as minimum as possible. However note that the same experimental results for v and D also lead to the failure of the FT and to bad estimations of the motor step size. Indeed using an helicase step size of 1 bp, a value that might approach better to the real step size of the gp41 helicase, and using Eq. 3, one finds a $Q_{TUR} \sim 200$ (in k_B units). Therefore, in the future, we will need to revise the model to better reproduce the experimental results and obtain a reliable value for Q_{TUR}.

IV. CONCLUSIONS

Helicases are enzymes that participate in various metabolic processes. They convert the chemical energy from ATP hydrolysis into mechanical work to translocate along one strand of DNA unwinding the double helix. They work under non-equilibrium conditions performing a mechano-chemical cycle acting as molecular motors. The energies involved in this process are on the order of the thermal energy ($k_B T$) and consequently, the molecular motor work in strong Brownian environment, where fluctuations play an important role. In their out of equilibrium activity they incur thermodynamic costs (dissipation).

Using MT single-molecule experiments we followed the activity of an helicase while unzips a DNA hairpin. By tracking the extension of the DNA hairpin we follow the position of the helicase along the DNA in real time. Analysing the DNA unzipping traces we compute properties related to the motion of the enzyme, i.e, velocity and diffusion coefficient. But the progress of the ATP hydrolysis reaction is inaccessible in these type experiments. In this project we investigate whether we can obtain information about the mechano-chemical reaction using non-equilibrium relations such as the TUR or the FT for entropy production. In order to test these relations we use a simple model for the helicase motion along the DNA. We describe the DNA chain as a one-dimensional lattice, and model the helicase motion as a random walk taking forward and backward steps in the lattice with constant probabilities $(P+, P_-)$. We have been able to verify that our model satisfies the FT for the entropy production and we have computed an analytical expression for the Q_{TUR} factor that is a factor that relates the fluctuations and dissipation in the motor activity.

The theoretical results have been verified through simulations of the random walk model. The simulations act as a bridge between theory and experiments. Even that in this work we have only tested the simplest random walk model, the simulations can be easily extended to include several aspects of the real system (as described in Methods section).

Results from theory and simulations has been next compared to experiments, by using values for the probabilities P_+ and P_- that fit the best the experimental results. Simulations generate helicase trajectories that qualitatively reproduce the experimental ones. However,

the shape of the distribution of helicase displacement deviates from the Gaussian distribution predicted by theory, showing much larger tails. Interestingly, the experimental distributions also verify a kind of FT relation but with an slope that do not correspond to the predicted by the theory proposed. Moreover, from the experimental slope in the FT relation we can estimate the value of the helicase mechano-chemical step size (that is the number of bps the helicase unzips for ATP hydrolized) that disagrees with the expected one. In overall these results suggest that the simple random walk model lacks some important aspects of the experimental system, such as the DNA sequence, the Brownian fluctuations of the bead or the different steps of the mechano-chemical cycle (Methods section). In that direction we plan to include several experimental features in the simulations to investigate which ingredients might be responsible for the enlargement of the displacement distribution and the failure of the FT.

V. APPENDIX

A. Gaussian approximation to Ising-like model

Consider a one dimensional random walker moving on a lattice taking steps of size $a = 1$ to the left (backward) and the right (forward) with the same probability ($P_+ = P_- = \frac{1}{2}$). We can relate this structure with an Ising model of spins in a 1D lattice, where the spin up indicates one step forward and spin down indicates one step backward. The probability of finding the walker at distance N after M steps is $P_M(N)$ which satisfies the following recursive relation:

$$P_{M+1}(N) = \frac{1}{2D}\sum_{j}^{M} J_{ij} P_M(j) \tag{46}$$

where D is the dimension and J_{ij} is the interaction matrix. This J_{ij} can be computed in the one dimensional case considering steps of size 1 giving a contribution of $\delta_{n,\hat{a}}$ for the right step and $\delta_{n,-\hat{a}}$ for the left step. Moving to the Fourier space (wave vectors "k") we obtain a total contribution of $J_{ij}(k) = (\mathcal{F}(\delta_{n,\hat{a}}) + \mathcal{F}(\delta_{n,-\hat{a}})) = 2\sum_{\alpha=0}^{D} cos(k_\alpha a)$.

Introducing the generating function ($G_q(N)$) we can recover the probability function $P_M(N)$ using the definition:

$$P_M(N) = \frac{1}{M!}\frac{d^M G_q(N)}{dq^M}\Big|_{q=0} \tag{47}$$

Since the expression of the generating function is unknown we will work through the Fourier transform ($G_q(N) = \sum_{k=0}^{\infty}\hat{G}_q(k)e^{-ikN}$) which we know its result [26]:

$$P_M(N) = \frac{1}{M!}\frac{d^M \sum_{k=0}^{\infty}\hat{G}_q(k)e^{-ikN}}{dq^M}\Big|_{q=0} \tag{48}$$

where $\hat{G}_q(k) = \sum_{N=0}^{\infty} G_q(N)e^{-ikN} = \frac{1}{1-\frac{q}{2D}J_{ij}} = \frac{1}{1-\frac{q}{D}\sum_{\alpha=1}^{D} cos(k_\alpha a)}$. Substituting this expression in 48 and taking derivatives we find:

$$P_M(N) = \sum_{k=0}^{\infty} e^{ikN}\left(\frac{\sum_{\alpha=1}^{D} cos(k_\alpha a)}{D}\right)^M \tag{49}$$

which can be written as an integral using the relation $\sum_k = \frac{V}{(2\pi)^D}\int_k$ [27].

$$P_M(N) = \frac{1}{D^M}\frac{V}{(2\pi)^D}\int_k d^D k e^{ikN}\left(\frac{\sum_{\alpha=1}^{D} cos(k_\alpha a)}{D}\right)^M \tag{50}$$

As we have said, we are interested in the one dimensional case (D=1). The integration limits in Fourier space are $\frac{-\pi}{a} < k < \frac{\pi}{a}$:

$$\begin{aligned} P_M(N) &= \frac{V}{2\pi}\int_k dk\ e^{ikN} cos^M(ka) \\ P_M(N) &= \frac{V}{2\pi}\int_{\frac{-\pi}{a}}^{\frac{\pi}{a}} dk\ e^{ikNa} cos^M(ka) = \\ &= \frac{1}{\pi}\int_0^{\pi} dk\ cos(Nk)\ cos^M(k) \end{aligned} \tag{51}$$

where in the last equality of Eq. (51) we have changed the integration limits, we have taken only the real part of the exponential and we make explicit $a = 1$ giving $V = a^D = 1$. Integrating by parts we find a recursive integral and we obtain an expression for $P_M(N)$:

$$P_M(N) = \frac{M!}{2^M(\frac{M-N}{2})!(\frac{M+N}{2})!} \tag{52}$$

which can be physically understood: $M!$ refers to the number of total steps, $(\frac{M+N}{2})!$, $(\frac{M-N}{2})!$ are the number of forward and backwards steps respectively and 2^M is a normalization factor which takes into account all the possible combinations using these two parameters (forward and backward). We can see that the result can be expressed as combinatorics about how many ways we can arrange M steps in groups of $n_+ = \frac{M+N}{2}$ steps forward and $n_- = \frac{M-N}{2}$ steps backward. For large M this expression converges to a Gaussian distribution: $P_t(x) = \frac{1}{\sqrt{4\pi D t}}e^{-\frac{x^2}{4Dt}}$ with mean 0 and variance $2DM$, where the diffusion coefficient $D = 1/2$ for this particular case.

From the result found in Eq. (52) we can propose a combinatorics method that will allow us to understand a more general case. Let us consider the same one dimensional random walk with forward and backward probabilities given by P_+ and P_- respectively. The possible combinations of arriving at distance N after M steps is given by $C(n_+, n_-) = \frac{M!}{n_+! n_-!}$, then the probability should be written as:

$$P_M(N) = \sum_{n_+,n_-} C(n_+, n_-) P_+^{n_+} P_-^{n_-} \tag{53}$$

we can express Eq. (53) only in terms of n_+ using the relations $N = n_+ - n_-$, $M = n_+ + n_-$ as:

$$P_M(N) = \frac{M!}{\frac{M+N}{2}! \frac{M-N}{2}!} P_+^{\frac{M+N}{2}} P_-^{\frac{M-N}{2}} \tag{54}$$

As we can see if we take the symmetric case ($P_+ = P_- = \frac{1}{2}$) in Eq. (54) we obtain the same result of Eq. (52). If we take large M we also find a Gaussian distribution, in this case with a drift term ($\hat{x}$) due to the bias in the probability $P_+ \neq P_-$ (in general),

$$P_t(x) = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{(x-\hat{x})^2}{4Dt}} \tag{55}$$

with $D = 2P_+P_-$ and $\hat{x} = (P+ - P-)$.

B. Large deviation approach

Using this method we can compute the general case where the random walker can move forward (n_+), backward (n_-) and stay still ($\bar{n}$) with probabilities P_+, P_- and P_0 respectively. In this case we have changed the label for the probability from P to ϕ to differentiate it from the particular case. Then, the probability of finding the walker at distance N after M steps starting at zero is given by:

$$\phi(n_0 = 0, n_M = N) = \sum_{n_+,n_-,\bar{n}} \frac{M!}{n_+! n_-! \bar{n}!} P_+^{n_+} P_-^{n_-} P_0^{\bar{n}} \tag{56}$$

Taking into account the following restrictions: $n_+ + n_- + \bar{n} = M$ and $n_+ - n_- = N$.

$$\phi(n_+, N, M) = \frac{P_0^{M+N}}{P_-^N} \sum_{n_+=N}^{\frac{M+N}{2}} \frac{M!}{n_+!(n_+ - N)!(M + N - 2n_+)!} \alpha^{n_+}$$

$$\phi(n_+, N, M) = \frac{P_0^{M+N}}{P_-^N} \sum_{M_+=N}^{\frac{M+N}{2}} e^{A(n_+)} \tag{57}$$

where $A(n_+) = \log(\frac{M!}{n_+!(n_+-N)!(M+N-2n_+)!})\alpha^{n_+}$ and $\alpha = \frac{P_+P_-}{P_0^2}$. Using the Stirling approximation $\log(n!) \approx n\log(n) - n$ we obtain the following expression.

$$\begin{aligned} A(y,z) &= Ma(y,z) \\ a(y,z) &= z\log(\alpha) - z\log(z) - (z-y)\log(z-y) \\ &\quad - (1+y-2z)\log(z+y-2z) \end{aligned} \tag{58}$$

where $z = \frac{n_+}{M}$ and $y = \frac{N}{M}$. Multiplying Eq. (57) for the increment $\Delta n_+ = 1$ and dividing by the total number of steps M, we can write Eq. (57) as an integral considering the limit $M \to \infty$ ($\frac{\Delta n_+}{M} = dz$).

$$\begin{aligned} \phi(n_+, N, M) &= M \frac{P_0^{M+N}}{P_-^N} \sum_{n_+=N}^{\frac{M+N}{2}} e^{A(n_+)} \frac{\Delta n_+}{M} \\ \phi(y,z) &= M \frac{P_0^{M+N}}{P_-^N} \int_y^{\frac{1+y}{2}} e^{A(y,z)} dz \end{aligned} \tag{59}$$

Where the latest integral can be computed using saddle point approximation,

$$\phi(y, z^*(y)) = M \frac{P_0^{M+N}}{P_-^N} \sqrt{\frac{-2\pi}{a''(y, z^*(y)}} e^{Ma(y,z^*(y))} \tag{60}$$

with,

$$\begin{aligned} a''(y, z^*(y)) &= -\left(\frac{1}{z^*(y)} + \frac{1}{z^*(y) - y} + \frac{1}{1 + y - 2z^*(y)}\right) \\ z^*(y) &= \frac{y+1}{2} - \frac{1}{2(1-4\alpha)}\left(1 \pm \sqrt{4\alpha + y^2(1-4\alpha}\right) \end{aligned} \tag{61}$$

as we can see in the expression of the relative extrema ($z^*(y)$) we have two possible solutions: z_+^* and z_-^*. We use the one that guarantee that the second derivative of $a(y,z)$ is negative in order to have a maximum in the saddle point approximation, then $z^*(y) = \frac{y+1}{2} - \frac{1}{2(1-4\alpha)}\left(1 - \sqrt{4\alpha + y^2(1-4\alpha}\right)$.

Due to the number of approximations made (Stirling and saddle point) we need to re-normalize Eq. (60) obtaining the final result:

$$\phi(y, z^*(y)) = \frac{(\frac{P_0}{P_-})^N \sqrt{\frac{1}{a''(y,z^*(y))}} e^{Ma(y,z^*(y))}}{\sum_{N=-M}^{M} (\frac{P_0}{P_-})^N \sqrt{\frac{1}{a''(y,z^*(y))}} e^{Ma(y,z^*(y))}} \tag{62}$$

which in terms of our initial variables is,

$$\phi_M(N) = \frac{(\frac{P_0}{P_-})^N \frac{1}{\sqrt{a''(N)}} e^{Ma(N)}}{\sum_{N=-M}^{M} (\frac{P_0}{P_-})^N \frac{1}{\sqrt{a''(N)}} e^{Ma(N)}} \tag{63}$$

The time evolution of this model can be thought as a probabilistic combination of different states (p_i) at different times (M) described by the following master equation,

$$p_n(M) = P_+ p_{n-1}(M-1) + P_- p_{n+1}(M+1) + P_0 p_n(M) \tag{64}$$

expanding the latest expression as a Taylor series we obtain a partial differential equation, changing to continuous notation (x,t):

$$\Delta t \frac{\partial p_x(t)}{\partial t} = (\Delta x)^2 (\frac{P_+ + P_-}{2}) \frac{\partial^2 p_x(t)}{\partial x^2} - \Delta x (P_+ - P_-) \frac{\partial p_x(t)}{\partial x}, \tag{65}$$

If we identify $D = \frac{(\Delta x)^2}{\Delta t} \frac{P_+ + P_-}{2}$ and $u = \frac{(\Delta x)}{\Delta t}(P_+ - P_-)$ we get,

$$\frac{\partial p_x(t)}{\partial t} = D \frac{\partial^2 p_x(t)}{\partial x^2} - u \frac{\partial p_x(t)}{\partial x}, \tag{66}$$

which is the diffusion equation with an advection term. The solution for this equation is:

$$p_x(t) = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{(x-ut)^2}{4Dt}}, \tag{67}$$

with the corresponding values of D and u found before.

We have finally obtained an expression for the probability of finding the walker at position N after M steps in the more general case (Eq. (63)) and related with a diffusive process following a Gaussian distribution.

[1] Suhasini, A. N., Brosh, R. M., Jr (2013). Disease-causing missense mutations in human DNA helicase disorders. Mutation research, 752(2), 138–152.

[2] Hartman, T. R., Qian, S., Bolinger, C., Fernandez, S., Schoenberg, D. R., Boris-Lawrie, K. (2006). RNA helicase A is necessary for translation of selected messenger RNAs. Nature structural molecular biology, 13(6), 509-516.

[3] Ritort, F. (2006). Single-molecule experiments in biological physics: methods and applications. Journal of Physics: Condensed Matter, 18(32), R531.

[4] Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C., Ha, T. (2008). Advances in single-molecule fluorescence methods for molecular biology. Annual review of biochemistry, 77(1), 51-76.

[5] Greenleaf, W. J., Woodside, M. T., Block, S. M. (2007). High-resolution, single-molecule measurements of biomolecular motion. Annual review of biophysics and biomolecular structure, 36, 171.

[6] Neuman, K. C., Nagy, A. (2008). Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nature methods, 5(6), 491-505.

[7] Manosas, M., Meglio, A., Spiering, M. M., Ding, F., Benkovic, S. J., Barre, F. X., Croquette, V. (2010). Magnetic tweezers for the study of DNA tracking motors. In Methods in enzymology (Vol. 475, pp. 297-320). Academic Press.

[8] Huguet, J. M., Bizarro, C. V., Forns, N., Smith, S. B., Bustamante, C., Ritort, F. (2010). Single-molecule derivation of salt dependent base-pair free energies in DNA. Proceedings of the National Academy of Sciences, 107(35), 15431-15436.

[9] Huguet, J. M., Ribezzi-Crivellari, M., Bizarro, C. V., & Ritort, F. (2017). Derivation of nearest-neighbor DNA parameters in magnesium from single molecule experiments. Nucleic Acids Research, 45(22), 12921–12931.

[10] Lionnet T., Allemand J. F., Revyakin A., Strick T. R., Saleh O. A., Bensimon D. Croquette V. (2011). Single-Molecule Studies Using Magnetic Traps. Cold Spring Harbor Protocols

[11] C. Jarzynski. Non-equilibrium equality for free-energy differences. Physical Review Letters, 78:2690–2693, 1997.

[12] E. G. D. Cohen D. J. Evans and G. P. Morriss. Probability of second law violations in shearing steady states. Physical Review Letters, 71:2401–2404, 1993.

[13] Barato, A. C., & Seifert, U. (2015). Thermodynamic uncertainty relation for biomolecular processes. Physical review letters, 114(15), 158101.

[14] Pietzonka, P., Ritort, F., & Seifert, U. (2017). Finite-time generalization of the thermodynamic uncertainty relation. Physical Review E, 96(1), 012101.

[15] Song, Y., & Hyeon, C. (2021). Thermodynamic uncertainty relation to assess biological processes. The Journal of Chemical Physics, 154(13), 130901.

[16] G. E. Crooks. Entropy production fluctuation theorem and the non-equilibrium work relation for free-energy differences. Physical Review E, 60:2721–2726, 1999

[17] Ritort, F. (2008). Nonequilibrium fluctuations in small systems: From physics to biology. Advances in chemical physics, 137, 31.

[18] Lionnet, T., Spiering, M. M., Benkovic, S. J., Bensimon, D., Croquette, V. (2007). Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism. Proceedings of the National Academy of Sciences, 104(50), 19790-19795.

[19] Gosse, C., Croquette, V. (2002). Magnetic tweezers: micromanipulation and force measurement at the molecular level. Biophysical journal, 82(6), 3314-3329.

[20] Bouchiat, C., Wang, M. D., Allemand, J.-F., Strick, T., Block, S. M., & Croquette, V. (1999). Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements. Biophysical Jour-

nal, 76(1), 409–413.

[21] Liphardt, J., Dumont, S., Smith, S. B., Tinoco Jr, I., Bustamante, C. (2002). Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. Science, 296(5574), 1832-1835.

[22] Collin, D., Ritort, F., Jarzynski, C., Smith, S. B., Tinoco, I., Bustamante, C. (2005). Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. Nature, 437(7056), 231-234.

[23] Yodh, J. G., Schlierf, M., Ha, T. (2010). Insight into helicase mechanism and function revealed through single-molecule approaches. Quarterly reviews of biophysics, 43(2), 185-217.

[24] Manosas, M., Xi, X. G., Bensimon, D., Croquette, V. (2010). Active and passive mechanisms of helicases. Nucleic acids research, 38(16), 5518-5526.

[25] Astumian, R. D., Bier, M. (1996). Mechanochemical coupling of the motion of molecular motors to ATP hydrolysis. Biophysical journal, 70(2), 637-653.

[26] Parisi, G., 1987. Statistical field theory. Menlo Park, Calif.: Benjamin/Cummings Pub. Co.

[27] Peliti, L. and Epstein, M., n.d. Statistical mechanics in a nutshell.