Mechanical folding and unfolding of the redox protein Flavodoxin.

Author: Enric Sanmartí
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisors: Felix Ritort and Anna Alemany

Abstract: We investigate the unfolding and folding processes of redox protein Flavodoxin under the action of mechanical forces at the single molecule level using force spectroscopy methods with optical tweezers. Previous bulk studies have provided valuable information about the structure of the folded and unfolded states, and pointed out the existence intermediate states along the unfolding and folding pathways. However direct evidence about the existence of such kinetic intermediates and their lifetimes is still missing. Here we pull molecular constructs of the protein Flavodoxin flanked by double stranded DNA linkers tethered to micron-sized beads with optical tweezers to mechanically unfold the protein. The distributions of rupture forces have been characterized and intermediate states have been found along the unfolding and folding pathways. Moreover, the molecular free energy landscape of the protein has been characterized from the unfolding kinetics using the Bell-Evans model and an estimate of the free energy of formation of the protein has been extracted using the Crooks fluctuation theorem.

I. INTRODUCTION

The protein folding problem refers to the study of how does the amino-acid sequence of a protein establish its spatial structure, to the secondary and tertiary levels. Many questions arise: how can the structure of a protein be predicted from its amino-acid sequence? What mechanisms does a protein use in order to fold and unfold? The three-dimensional structure of a protein is intimately linked to its biological function. For instance, ATP-synthase [1] is an enzyme that catalyzes a condensation reaction of the adenosine diphosphate (ADP) molecule into an adenosine triphosphate (ATP) molecule, that is the most common energy source of cells. Folded ATP-synthases are able to fix themselves to the inner mitochondrial membrane and consist of two structures, one that is located inside (Fo) and the other located outside the membrane (F1). F1 contains a rotating subunit, which is responsible for the synthesis of ATP. Protons from the membrane enter the protein in Fo, driven by the charge gradient that exists between the membrane and the inner part of the mitochondrion, activating the rotation of the F1 subunit which is responsible for the synthesis of ATP. Here, the structure of the protein is a key element to develop its function: if an ATP-synthase was not correctly folded, it would not be able to fix itself to the membrane and accomplish its biological purpose. The example of ATP-synthases shows to what extent the structure of proteins determine their biological function, which is crucial for life.

The misfolding or the inability of a protein to reach its functional state can lead to misfunctional behaviours and even to lethal diseases. The most well-known example of a protein misfolding-related disease is given by Alzheimer's disease [2], related to the misfolding of the amyloid precursor protein (APP), which misfolds into a β-amyloid protein. This misfolded protein has aggregative propeties, that the correctly folded APP protein does not have, causing the formation of amyloid fibrils and plaques of high toxicity that impair the normal brain functions.

During the search for their native/functional state, proteins can explore billions of different configurations. If proteins would fold by randomly exploring their configurational space, it would take longer than the age of the universe until they reach their native configuration, but it has been observed that they are able to fold in very short time scales. This is known as Levinthal's paradox, formulated by American molecular biologist Cyrus Levinthal in 1969 [3]. Levinthal himself suggested that proteins may fold in a guided way, that is by forming partially folded states that act as nucleation points of the folding process. The protein folding process can be described by the free energy landscape (FEL) of the protein, which is the profile of the free energy of the molecule as function of a reaction coordinate such as the end-to-end distance or the number of native contacts of the protein. A protein's FEL allows to visualize the degree of stability of the different configurations, as well as the pathways that the protein follows in order to access the stable folded state. In Levinthal's view, this FEL has the shape of a golf course being essentially flat, with one low energy state or hole corresponding to the native structure (FIG. 1a). However, this scenario is not realistic because the search of the native state would take too long. The FEL must be rather curved, in a funnel-like shape, towards the folded state (FIG. 1b). Nevertheless, experimental evidence suggests that the FEL adopts the funnel-like shape but has a rough surface, which indicates the presence of one or many local minima, corresponding to states where the protein can be kinetically trapped in a partially folded configuration (FIG. 1c-d).

Protein folding has been extensively studied during the last half century, both experimentally and computationally. Bulk experiments allow to study the folding processes by modifying the protein's physical or chemical environment, that is by changing temperature and pressure or adding a concentrated solution of denaturant

(typically urea). It is important to emphasize that bulk experiments consist of measurements on an ensemble of proteins, implying that only the average properties over the ensemble can be studied. Bulk techniques such as Nuclear Magnetic Resonance Spectroscopy (Protein NMR), circular dichroism, X-ray crystallography and calorimetry are most commonly used [4–7].

Single molecule spectroscopy techniques have emerged in the last two decades as an extremely powerful tool to investigate molecular systems. They allow one to study an individual molecule, and have opened a brand new field in biomolecular research [8]. While bulk experiments may contain valuable information about the average properties of an ensemble of molecules, single molecule techniques allow to record the fluctuations of those properties and detect transient structures. Single molecule techniques can be divided in two groups: the ones that use fluorescence to visualize conformational fluctuations of the molecules [9]; and force spectroscopy techniques [10], which allow to apply mechanical forces to the ends of individual molecules. The most commonly used single molecule force spectroscopy techniques are Atomic Force Microscopy (AFM), Magnetic Tweezers (MT) and Optical Tweezers (OT), and have granted the manipulation of single molecules and the measurement of forces of the order of pico-newtons (pN), distances of the order of nm with changes in time scales of the order of μs. The use of each force spectroscopy technique mostly depends on the required range of forces: AFM ideally covers between 50-500 pN, while MT covers between 0.02-10 pN and OT covers between 1-100 pN.

In this project, we have studied the mechanical folding and unfolding mechanisms of protein Flavodoxin with single molecule experiments that were carried out using OT. Flavodoxins are small bacterial proteins with electron transfer function, participating in many redox reactions and containing between 140 and 180 amino-acids. Reduction-Oxidation (also termed "redox") reactions include all chemical reactions in which atoms have their oxidation state changed; in general, redox reactions involve the transfer of electrons between species. The Flavin Mononucleotide (FMN) cofactor tightly bounds to the native state of the protein, and is necessary to ensure its functionality. We study a Flavodoxin protein from bacterium *anabaena* that contains 168 amino-acids. The molecule was expressed in its *holo* form, i.e. in the presence of FMN. However, as can be seen in FIG. 2, the FMN cofactor binding only produces minor changes in the protein's secondary structure. Therefore, for simplicity, during force spectroscopy experiments, the binding of the cofactor is not monitored, hence transitions of the protein between its *holo* and *apo* forms are not controlled. Previous studies point out that the folding and unfolding processes and the binding of the cofactor are independent [11], given that well-folded *apoflavodoxins* are observed when the protein is synthesised at low concentrations of FMN.

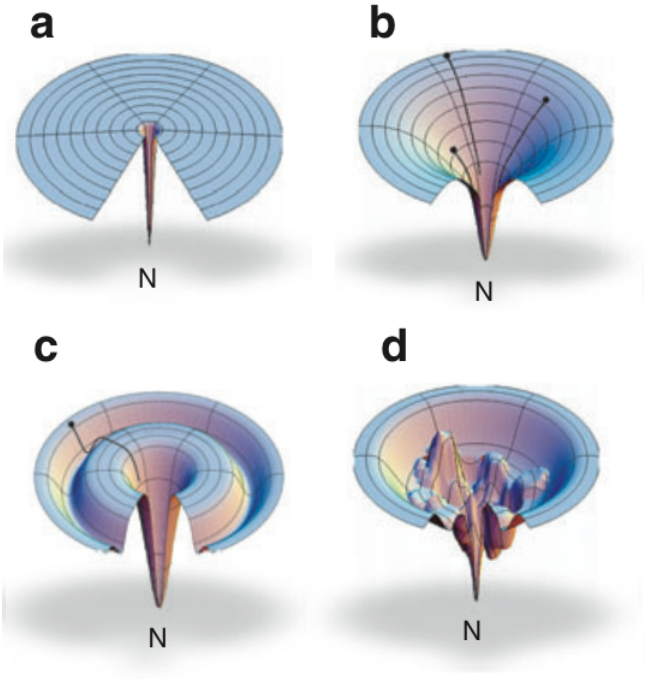

FIG. 1: **Protein free energy landscapes**. *a)* Flat FEL, proposed by Levinthal, where the protein explores many different configurations before randomly reaching the native state. *b)* Ideal funnel-shaped FEL, proposed by Levinthal as a solution to his paradox. *c)* Representation of a more realistic FEL where a mandatory intermediate is found in the folding pathway of the protein. *d)* Schematic FEL with several intermediate states on-pathway. Taken from [12].

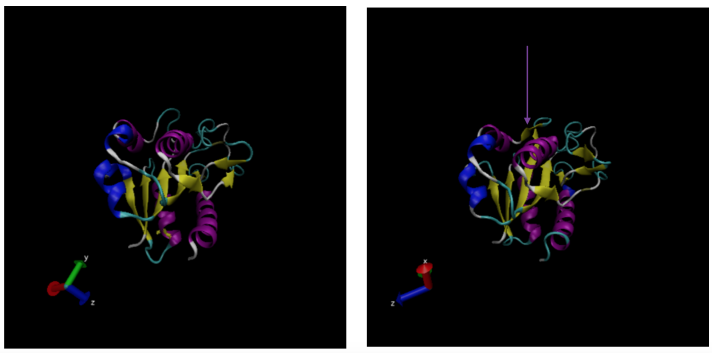

FIG. 2: **Secondary structure of apo- and holoflavodoxin.** The only remarkable difference in secondary structure between apoflavodoxin (*left*) and holoflavodoxin (*right*) is an extra β-sheet that appears in the second case, indicated with an arrow. These images have been produced with VMD [13]. with the molecular structures taken from [14, 15]

II. EXPERIMENTS & METHODS

A. Principles of optical tweezers

Arthur Ashkin opened the field of opical trapping in 1970, when he was able to show experimentally that elec-

tromagnetic radiation can be used to accelerate microscopic particles [17]. This is explained in FIG. 3a, where a microscopic dielectric polystyrene bead is irradiated with a laser beam. Each photon of a ray of the laser beam has a momentum h/λ in the direction of propagation of light, being λ the wavelength of the photons. When the ray enters the bead, a part of the radiation is lost by reflection while the other part is transmitted into the bead, changing its direction of propagation according to Snell's law. The same happens when the ray exits the bead. Due to transmissions at both interfaces, the momentum of the photons suffers a change in direction, accompanied by a change in its modulus due to the attenuation of light. To ensure momentum conservation in the whole system (light ray and bead), some momentum has to be transferred to the bead, which is then accelerated in the direction of propagation of light. This effect is called the *scattering force* f_{scatt}, and it is actually not the only force acting on the bead. Because of the typical Gaussian profile in the intensity of the laser beam, another effect, proportional to the difference in intensity, tends to accelerate the bead towards the center of the beam. This is called the *gradient force* f_{grad}. It is now clear that two counter-propagating laser beams of the same intensity and wavelength will generate opposed f_{scatt} that will compensate, and the resulting f_{grad} will be doubled (with respect to that of a single beam), thus being able to capture beads. This trap can be thought of as a potential well, which is, in good approximation, a harmonic well.

Sixteen years later, it was also Ashkin [18] who discovered that a strongly focused laser beam also creates a potential well able to trap small objects (FIG. 3b). This was the first optical tweezers setup and led to a turning point in the field of biomolecular research. Indeed, the forces exerted by the laser beam on the bead are of the order of pN, which is the typical order of magnitude at which many proteins mechanically unfold.

In this project, we use a setup consisting of two focalized counter-propagating laser beams to create an optical trap thus combining both previously described effects. We measure the force applied on the bead from the change in the momentum of light.

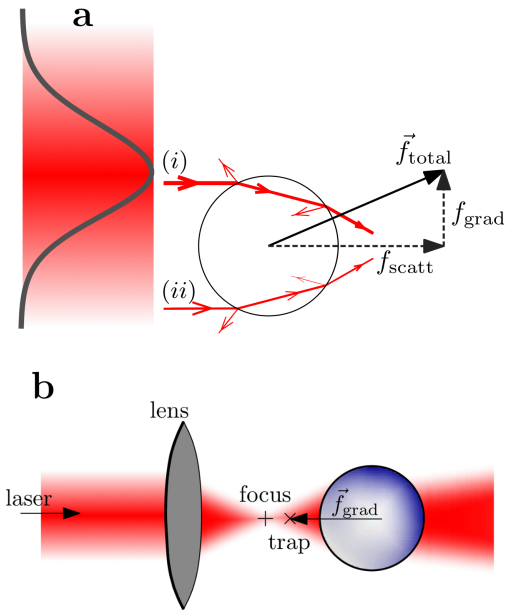

FIG. 3: **Principles of optical trapping.** *a)* Ray optics diagram of the optical trapping where scattering and gradient forces acting on the bead are generated by its interaction with the laser beam. *b)* Sketch of an OT setup generated with a focused laser beam.

B. Molecular setup

FIG. 4 shows the typical setup in our single molecule experiments. It consists of one single Flavodoxin protein intercalated between two identical double stranded DNA (dsDNA) handles of 558 base pairs each (details of the synthesis can be found in the Appendix). In each of the ends of Flavodoxin, a Ubiquitin protein has been placed as a connection between the end of the handle and that of the protein. Ubiquitins do not take part in the experiments (at least for not too high forces), but here they are required to prevent both ends of Flavodoxin from reacting. One of the handles is attached to a polystyrene bead by means of a streptavidin-biotin (SB) bond, and the bead is trapped in the tip of a glass micro-pipette by air suction, while the second handle is attached to a bead, which is confined in the optical trap in order to be able to manipulate it, by means of an antidigoxigenin-digoxigenin (AD) bond.

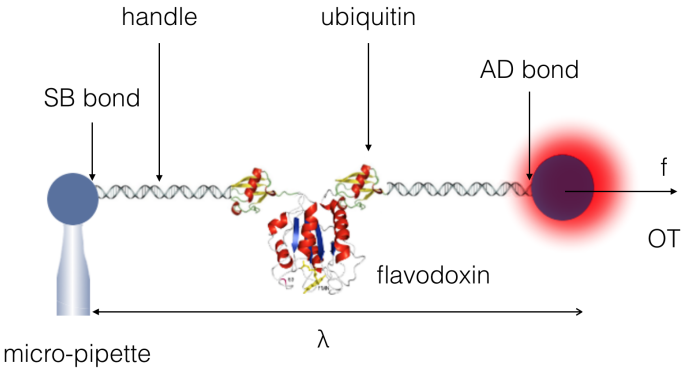

FIG. 4: **Setup of the OT instrument.** The Flavodoxin protein is tethered between two Ubiquitins and the resulting polyprotein is linked between two dsDNA handles. One end of the molecular construct is linked by means of an AD bond to a polystyrene bead captured in the optical trap, whereas the other is linked by a SB bond to a second bead immobilized in the tip of a micro-pipette.

C. Single molecule experiments

By moving the optical trap in FIG. 4 left or right, one is able to exert forces and manipulate the end-to-end distance of the molecular construct. Our control parameter is the distance λ between the optical trap and the tip of the micro-pipette, while the force is measured by the instrument. Applying a high-enough force to a folded protein causes its unfolding. Except where explicitly noted, experiments were carried out at 25℃ with 1 kHz temporal resolution and in a Tris 50 mM buffer at pH 8.0.

In pulling experiments, the protein is initially in its native/folded (N) state at zero force and is pulled up to ∼30 pN at a constant velocity $v = 60$ nm/s. The unfolding of the protein is observed as a sudden jump in force, as can be seen in the red curve in FIG. 5 (arrow). Remarkably, the unfolding forces (defined as the force where the jump takes place) change between 4 and 25 pN in different repetitions of the experiment. In the reverse process, the protein is initially in its unfolded (U) state and λ decreases at $-v$ until the force ∼0 pN is again reached. The folding of the protein is usually observed at forces below 4 pN (blue curve in FIG. 5, arrow). Remarkably, although a huge variability is observed in the unfolding pattern since unfolding forces cover a wide interval between 4 and 25 pN, all folding trajectories look almost identical. Pulling cycles can be repeated with one single molecule, reaching a maximum number of ∼60 unfolding/folding events until the molecular connection breaks. A particular behaviour (FIG. 6) has been observed when proteins are stretched up to ∼ 50 pN, which might be related to the unravelling of some structure in the molecular construct, most probably the unfolding of the Ubiquitins.

In passive-clamp experiments the protein is in a well known initial state (N or U). The distance λ is kept constant for ∼10 s, in order to allow the protein to fluctuate among different configurational states. These experiments have also been carried out at high temporal resolution (50 kHz), thus allowing us to better visualize the folding/unfolding process.

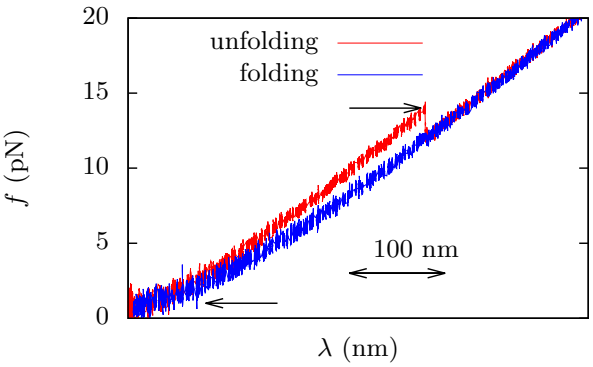

FIG. 5: **Pulling experiments.** Example of folding (blue) and unfolding (red) force-distance curves (FDC). Arrows indicate the folding and the unfolding of the protein respectively.

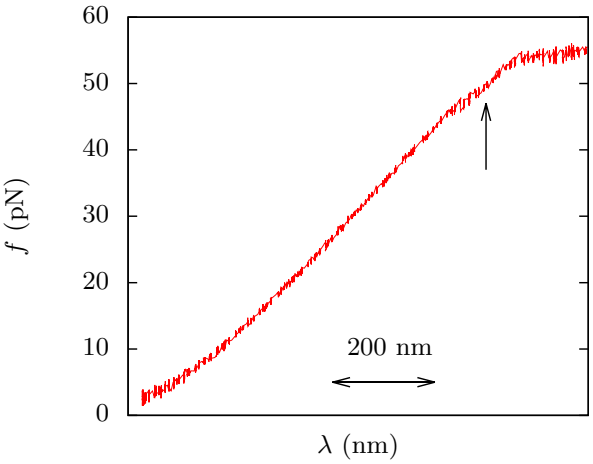

FIG. 6: **Pulling experiments up to very high forces.** A particular behaviour is observed at 48 pN (arrow), which is attributed to the unfolding of the Ubiquitins of the polyprotein construct. Next, above 50 pN, the overstrtching of the dsDNA handles is observed as a plateau in the extension.

III. RESULTS

A. Unfolding pathway

1. *Mechanical unfolding of Flavodoxin*

Pulling experiments suggest that the unfolding of the protein occurs within a wide range of values for the unfolding force, which varies between 4 and 25 pN in each trajectory. This can be observed in the collection of unfolding trajectories shown in FIG. 7 and in the histogram of unfolding forces (FIG. 8).

In order to assess whether the force jump corresponds to the unfolding of the protein Flavodoxin, the total number of amino-acids N_{aa} released in the transition is extracted as described in reference [19] and summarized in what follows. The force jump Δf is related to the difference between the end-to-end distances of the unfolded and the folded protein, x_p and x_d respectively, corresponding to the projection of the molecular extension along the force axis. For any unfolding force f where Δf is measured, it is found that:

$$\frac{\Delta f}{k_{\text{eff}}} = x_p(f) - x_d(f), \tag{1}$$

where k_{eff} is the effective stiffness of the whole construct, including the optical trap, measured as the slope of the FDC. The folded protein is modelled as a single bond of length $d = 2$ nm (equal to its diameter) that aligns with the external force as a magnetic dipole does in a

magnetic field. Then, the expression for x_d is:

$$x_d(f) = d\left[\coth\left(\frac{fd}{k_BT}\right) - \frac{k_BT}{fd}\right], \qquad (2)$$

where k_B is the Boltzmann constant and T is the temperature. x_p can now be obtained by substituting Δf, k_{eff}, and x_d into Eq. (1). We model the unfolded protein with the worm-like chain elastic model, where the protein is assumed to behave as a continuous strand of contour length L, with an energy penalty associated to its bending. x_p is related to the external force f by the following interpolating formula [20]:

$$f = \frac{k_BT}{P}\left[\frac{1}{4\left(1-\frac{x_p}{L}\right)^2} - \frac{1}{4} + \frac{x_p}{L}\right], \qquad (3)$$

where P is the persistence length, taken equal to 0.85 nm [21, 22] and L is the total length of the protein, equal to $N_{\text{aa}}d_{\text{aa}}$, where $d_{aa} = 0.36$ nm [21, 22] is the distance between two consecutive amino-acids. Therefore, by relating the measurement of Δf and the unfolding force with L, we can extract N_{aa}.

Results are shown in FIG. 9, where two major peaks are observed in the distribution of N_{aa}. We fit this results to a double Gaussian function. The main peak (right) appears at $N_{aa} = 165 \pm 4$, which is in good agreement with the total number of residues (168) of our protein. The secondary peak (left) has been found at $N_{aa} = 116 \pm 5$, and may correspond to events where the protein partially unfolds. This suggests the existence of an intermediate state in the unfolding pathway, as we discuss in the next section.

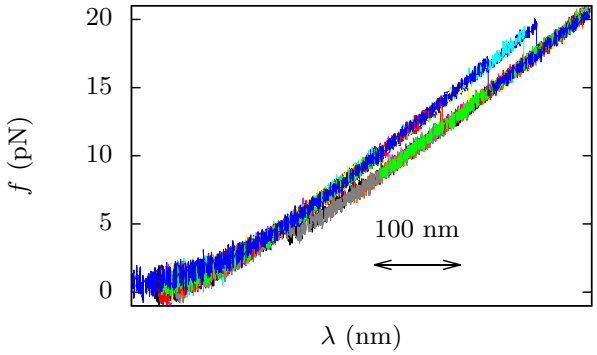

FIG. 7: **Collection of unfolding trajectories meaured in pulling experiments**. The unfolding of the protein is observed as a sudden force jump that changes in each repetition of the experiment between 4 and 25 pN.

2. *Intermediate states in the unfolding pathway*

Some of the unfolding trajectories suggested the possible existence of an intermediate state in the unfolding pathway of Flavodoxin, together with the result obtained

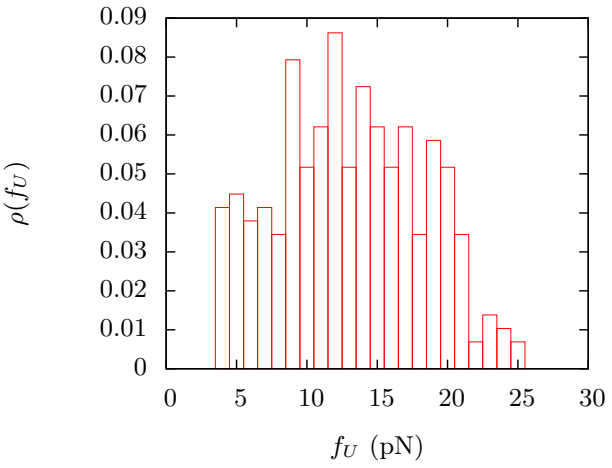

FIG. 8: **Histogram of the unfolding forces f_U.**

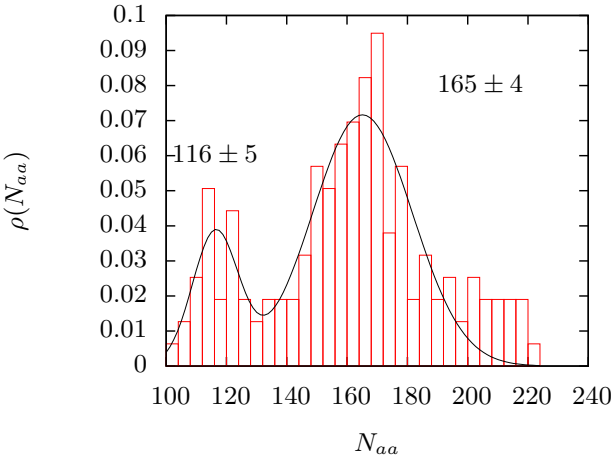

FIG. 9: **Amino-acids released in the unfolding transition.** Histogram of the number of released amino-acids N_{aa} in the unfolding transition (red bars) and corresponding fit to a double Gaussian function (black line).

in previous section (FIG. 9). In order to confirm its existence, passive-clamp experiments were carried out at high temporal resolution (50 kHz). A representation of such experiments is found in FIG. 10: we set the protein in state N at ∼14 pN, and then we keep λ constant for a few seconds, until Flavodoxin unfolds. A close-up (FIG. 11) of the transition events measured in passive-clamp conditions reveals the presence of an intermediate state I_1 close to N. Additionally, a second intermediate state I_2 with a shorter lifetime is observed, close to U. In order to characterize both intermediates, we plot the distribution of forces within the small time window corresponding to the close-up and we fit the data to a quadruple Gaussian function (FIG. 12). From the fit, we extract the fraction of amino-acids released in the transition between states I_1 and U ($N_{\text{aa}}^{I_1 \to U}$) and between

states N and I_2 ($N_{\text{aa}}^{I_2\to U}$). We repeat the analysis for 3 different trajectories. In Table 1, we present the results for 3 different trajectories, where it can be seen that, in average, $N_{\text{aa}}^{I_1\to U} = 118 \pm 5$, in good agreement with the number of amino-acids corresponding to the secondary peak in FIG. 9. These results, together with the ones obtained in pulling experiments, are a clear signature of the existence of intermediate states on-pathway. In pulling trajectories where the number of amino-acids released corresponds to the secondary peak in FIG. 9, we are measuring the partial unfolding of the protein starting from state I_1 and going to state U.

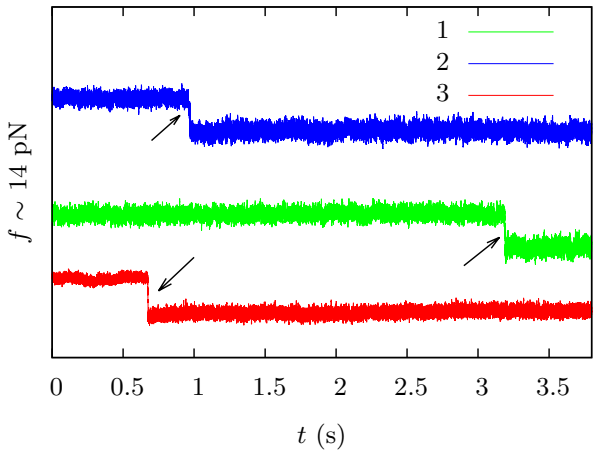

FIG. 10: **Collection of passive-clamp experiments.** Unfolding events measured (arrows) at 50 kHz in passive-clamp experiments. Different trajectories are labeled with numbers from 1 to 3. Trajectories have been shifted vertically for the purpose of visualization.

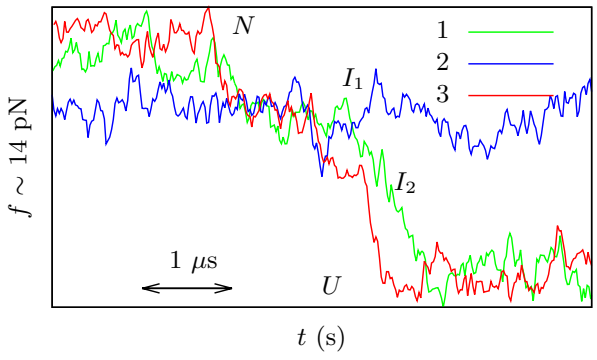

FIG. 11: **Close-up of the unfolding transition in passive-clamp experiments.** The molecule starts in the native (N) state and transiently visits two sequential intermediate states (I_1 and I_2) in its pathway to the unfolded state (U). Trajectories have been aligned for the purpose of visualization. Notice that trajectory 2 remains in I_1 for longer than the other trajectories.

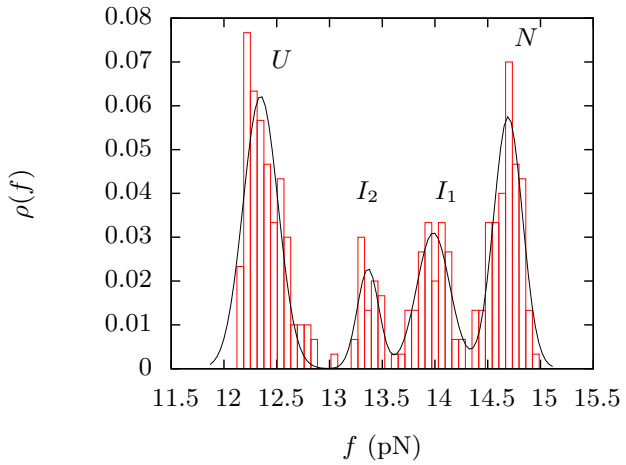

FIG. 12: **Intermediate states in the unfolding pathway.** Force histogram of the trace presented in FIG. 11 for trajectory 3 (red bars) and corresponding fit to a four-Gaussian function (black line).

Traj.	$N_{\text{aa}}^{I_1\to U}$	$N_{\text{aa}}^{I_2\to U}$
1	116 ± 8	60 ± 6
2	120 ± 8	-
3	117 ± 5	72 ± 8

TABLE I: **Analysis of unfolding passive-clamp experiments.** From the fit to multiple Gaussian functions of the force histograms obtained in FIG. 12 for 3 different trajectories, we extract the number of amino-acids released between I_1 and U ($N_{\text{aa}}^{I_1\to U}$) and between N and I_2 ($N_{\text{aa}}^{I_2\to U}$). Remarkably, for trajectory 2, I_2 was not observed.

3. Unfolding kinetics

The force-dependent kinetic rate $k_{N\to U}(f)$ is defined as the probability of going from state N to state U per unit time. We assume that the unfolding process is a first order Markov process and thus the survival probability, $P_N(f)$, of the protein to be in the N state along a pulling experiment satisfies the master equation:

$$\frac{dP_N}{df} = -\frac{k_{N\to U}}{r} P_F(f), \tag{4}$$

where r is the loading rate, defined as $r = k_{\text{eff}} v$, v being the pulling speed and k_{eff} the effective stiffness (Eq. 1). $P_N(f)$ and its derivative can be computed from the distribution of unfolding forces obtained in pulling experiments, since $P_N(f)$ is the ratio between the number of unfolding events taking place at a force greater than f and the total number of unfolding events measured. It is important to note that we have assumed that all the transitions go from N to U and thus we have not taken into account the presence of states I_1 and I_2 described in the former section. FIG. 13 shows the kinetic unfolding rate as a function of the force, obtained by inverting Eq. (4).

It is clear that the rate increases with force, which is consistent with the decreasing of the kinetic barrier that separates states N and U promoting the protein to hop to state U.

Different models exist in order to characterize kinetic rates in active transitions. The simplest and most renowned of those models is the Arrhenius-Kramers model [23, 24], which assumes that the kinetic rate increases exponentially as the barrier decreases, that is:

$$k_{N\to U}(f) = k_0 \exp\left(-\frac{B(f)}{k_B T}\right), \tag{5}$$

where k_0 is the attempt frequency and $B(f)$ is the force-dependent kinetic barrier. The Bell-Evans (BE) model [25] assumes that the barrier depends linearly on the force and thus $B(f) = B_0 - x^\ddagger f$, where $x^\ddagger$ is the position of the transition state with respect to the native state and B_0 is the value of the kinetic barrier at zero force. Accordingly, the kinetic rate can be written as:

$$k_{N\to U}(f) = k_m \exp\left(\frac{f x^\ddagger}{k_B T}\right), \tag{6}$$

where $k_m = k_0 \exp(-B_0/k_B T)$ is the unfolding kinetic rate at zero force. We fit our data to the BE model (FIG. 13) in order to extract these parameters, obtaining $x^\ddagger = 0.56 \pm 0.05$ nm and $k_m = 0.06 \pm 0.01$ s^{-1}. The resulting value for $x^\ddagger$ is of the same order of magnitude than previous results obtained for other proteins [12, 22]. The value obtained for k_m is relatively large, given that it implies that every protein spontaneously unfolds in an average time of 17 s at zero force. This might be related to the fact that we are actually observing transitions from N to I_1, and thus we are in fact measuring $k_{N\to I_1}(f)$ instead of $k_{N\to U}(f)$. Further experiments will clarify this.

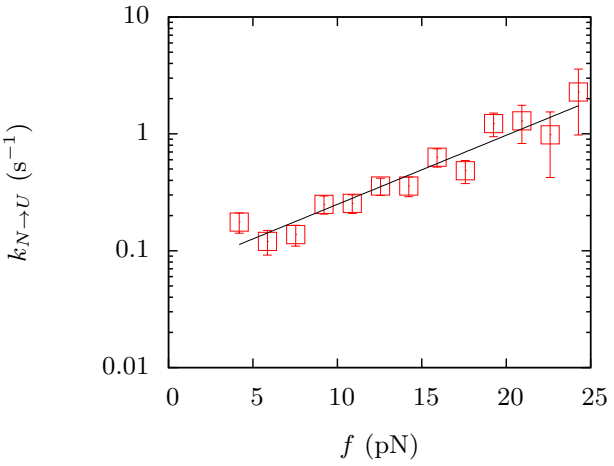

FIG. 13: **Unfolding kinetic rate as a function of the force.** Kinetic rate measured from the survival probability of the native state in pulling experiments (Eq. (4), red squares) and corresponding fit to the BE model (Eq. (6), straight line).

B. Folding pathway

Passive-clamp folding experiments at low forces reveal the presence of a jump in force, identified as the folding of the protein. FIG. 14 shows a collection of passive-clamp traces, where the protein was initially set at state U at $\sim$3 pN. Trajectories have been shifted vertically for the sake of visualisation. Folding events are indicated with arrows. In FIG. 15, a close-up of one of the trajectories is shown: the force jump is visibly smooth, uncovering the possible presence of an intermediate state in the folding pathway. High temporal resolution passive-clamp experiments were performed in order to confirm the existence of this fast intermediate. In this case, its presence is masked due to large fluctuations unavoidable at low forces. In future experiments, we will gain further understanding of the fast intermediates through detailed fluctuation spectroscopy measurements.

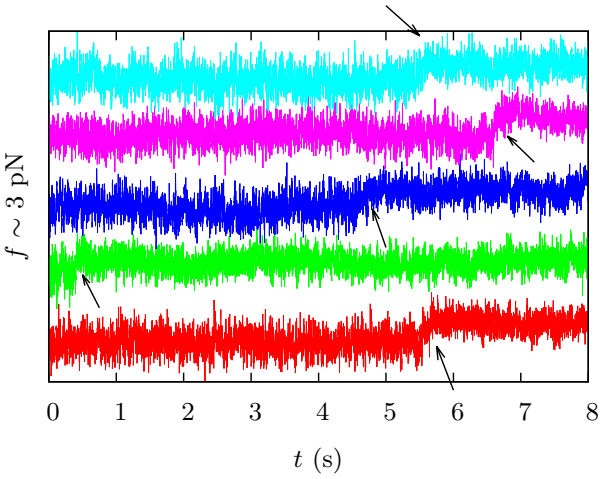

FIG. 14: **Passive clamp folding experiments.** Collection of folding trajectories, shifted vertically, where the molecule was originally set unfolded at $\sim$3 pN and the folding is visible as a smooth increase in force (arrows).

C. Free energy of formation

Fluctuation relations have been used in order to extract an estimate of the free energy of formation of the protein. These relations relations appeared two decades ago, in the context of a need for non-equilibrium statistical-mechanical theories to describe small systems (i.e. systems with a number of particles of the order of one). For macroscopic systems, independent realizations of an identical experimental protocol are performed by continuously modifying some control parameter under non-equilibrium conditions, an almost identical value is obtained for the work performed in all cases. This work is always larger than the difference in free energy ΔG between the initial and final states, according to the second

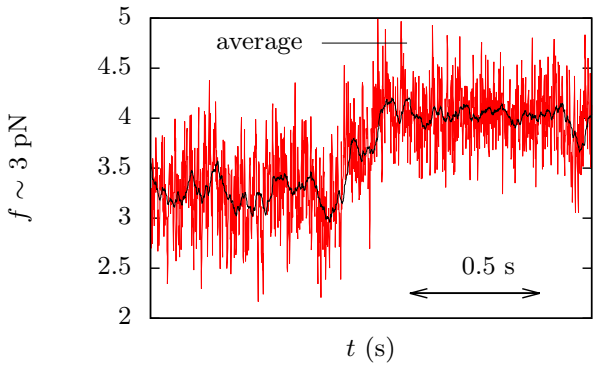

FIG. 15: **Intermediate states in the folding pathway.** Close-up of an individual folding event measured in a passive-clamp experiment. The presence of an intermediate state on-pathway is hinted.

law of thermodynamics. For small systems, fluctuations become important as they are of the same order as the typical energies involved in this scale, thus the value of the work done to drive the system between two states under non-equilibrium conditions is significantly different upon distinct realizations of the protocol. Interestingly enough, for some realizations of the protocol, the work will be lower than ΔG. Those are called *transient* violations of the second law, and represent a tiny fraction of the whole set of trajectories. However, the second law always holds, since the average work over all experimental realizations is always larger than ΔG.

Here, we use the Crooks Fluctuation Relation (CFR) [26–28] to extract the free energy of formation of Flavodoxin. The CFR establishes a relation between the work performed in a non-equilibrium protocol and the free energy difference between the initial and final states. Suppose that a system (protein) is initially in equilibrium at state N. The system is driven out of equilibrium by a *forward* protocol consisting in modifying some control parameter λ until a final state U is reached. The *reverse* protocol is the time-reversal image of the forward one, where U (N) is the initial (final) state. The CFR reads:

$$\frac{P_F(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta G_{NU}}{k_B T}\right), \qquad (7)$$

where F and R stand for forward and reverse, respectively and $\Delta G_{NU} = G_U - G_N$ is the free energy difference between states N and F. $P_F(W)$ is the probability of measuring a value W for the work performed in a forward protocol, while $P_R(-W)$ is the probability of measuring the opposite value in a reverse protocol. Notice that the value of W that is equiprobable for forward and reverse protocols actually corresponds exactly to ΔG_{NU}.

In pulling experiments, the forward (reverse) protocol is identified as the unfolding (folding) process. The control parameter λ is the relative distance between the center of the trap and the tip of the micro-pipette and the work is calculated as the area below the FDC between an initial and a final value of λ. Work histograms for both forward and reverse protocols are shown in FIG. 16. It is observed that $P_F(W)$ is very wide, while $P_R(-W)$ is sharp. This is a consequence of the large variability of unfolding forces (FIG. 8) and the smooth folding transition observed in folding transitions.

According to Eq. (7), and from the work distributions shown in FIG. 16, $\Delta G_{NU} = (856 \pm 2)\ k_B T$. This number contains not only the free energy of formation ΔG_0 of the molecule, but also the reversible elastic contributions due to stretching the dsDNA handles and the peptide chain from the initial to the final value of λ and the energetic contribution due to the displacement of the bead in the optical trap along the protocol. Each of this contributions can be theoretically subtracted from ΔG_{NU} as follows. The dsDNA handles are modelled using the extensible worm-like chain elastic model, with an inter-phosphate distance of 0.34 nm/base (and hence a contour length of 380 nm), a persistence length of 34 nm and a Young modulus equal to 850 pN [29]. Therefore we get that the reversible work required to stretch the dsDNA handles equals 66 $k_B T$. Next, the peptide chain is modelled as described in Section III.A.1, which gives that the reversible work needed to stretch the unfolded protein to the final value of λ is 51 $k_B T$. Finally, the displacement of the bead in the optical trap is assumed to be equal to the behaviour of a particle in an harmonic potential with stiffness equal to 0.05 pN/nm. The reversible work for this contribution is hence 727 $k_B T$. Taking into account all the different contributions, we obtain that the free energy of formation of protein Flavodoxin is $\Delta G_0 = 12 \pm 4\ k_B T$ (or 7±2 kcal/mol). This number is in the expected order of magnitude, since proteins usually have a small free energy of formation as a result of the balance between large enthalpic and entropic contributions.

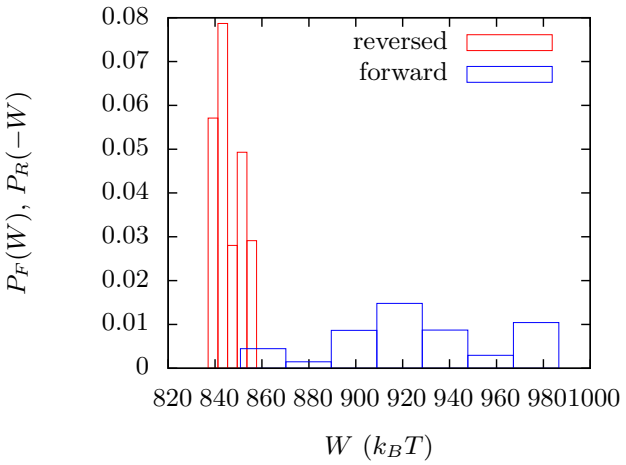

FIG. 16: **Work distributions for the forward and reversed protocols.** The intersection between both histograms corresponds to the free energy difference between the initial and final points of the protocol.

IV. CONCLUSION

In this work, we have investigated the mechanical folding and unfolding processes of protein Flavodoxin. Both the unfolding and folding pathways were characterized, as well as the main elements of the protein's FEL. First, pulling experiments have allowed us to extract the distribution of unfolding forces, which has been observed to be very wide. The distribution of the number of residues released in the unfolding event was computed, obtaining a double-peaked distribution. The main peak has been found to be centred around the total number of amino-acids of the protein (168), thus validating the single-molecule nature of our experiments, and the secondary peak has been related to the presence of an intermediate state. The unfolding pathway was studied with high-resolution passive-clamp experiments, unravelling the presence of one clear and relatively stable intermediate state close to the native state, and suggesting the presence of another, more unstable, intermediate state close to the unfolded state. The kinetic unfolding rate as a function of the applied force was found using the unfolding forces measured in pulling experiments. The fit to the BE model allowed us the find the position of the transition state with respect to the native state, as well as the kinetic unfolding rate at zero force. The unfolding pathway was characterized with passive-clamp experiments at forces of $\sim$3 pN, finding the possible presence of an intermediate state that could be related to those found for the unfolding pathway. The presence of such intermediate, as well as its characterization, should be confirmed in future passive-clamp experiments at high temporal resolution. At last, the free energy of formation of Flavodoxin was extracted using the Crooks fluctuation relation to be equal to $12 \pm 2\ k_BT$, which is a reasonable result in the same order of magnitude as those found for other proteins.

To further validate the results of our project, future control experiments should be performed studying the molecular construct without Flavodoxin (i.e. only handles and Ubiquitins), in order to find out if the construct is modifying our results. A deviation from the linear elastic behaviour is observed when the molecular construct is pulled up to forces of more than 50 pN. This atypical phenomenon might be related to the presence of the two Ubiquitin proteins which were necessarily included in the construct for its proper synthesis. Despite the alleged unfolding event takes place at high forces, compared to the typical rupture forces observed for Flavodoxin, the molecular construct might be playing a role in the phenomenology observed at low forces.

Even though previous studies suggest that the unfolding and folding processes are independent of the binding of the FMN cofactor, one can not exclude the possibility that the cofactor is playing some role. The experiments have been carried out using proteins expressed with the cofactor, but the cofactor may be expelled from the protein when it unfolds. Essentially, one cannot know if a given unfolding event has taken place with the cofactor attached to the protein. Experiments in a buffer with a concentrated solution of FMN could reveal if the cofactor has an influence on the folding and unfolding pathways of Flavodoxin.

V. APPENDIX: MOLECULAR SYNTHESIS

The synthesis consists of three main steps. The first is to transform cells, i.e. introduce the DNA sequence codyfing our polyprotein construct (Ubiquitin-Flavodoxin-Ubiquitin) inside the plasmid of the bacterial cells and induce protein expression with the use of a chemical product. The next step is lysing, which means eliminating the cells. Now, Flavodoxin has been synthesised together with other proteins which are not of interest, thus the sample is purified. The isolated polyprotein is then activated and lastly attached to dsDNA handles from both ends.

E. coli BL21(DE3) cells were transformed with the appropriate plasmid and grown at 37°C in LB (Luria Bertani). Upon reaching an optical density $\sim$0.8, protein expression was induced with 0.5 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) at 37°C for 5 hours. The cells were centrifuged and the pellet suspended in five volumes of lysis buffer (50 mM Tris-HCl pH 7.9, 2 mM EDTA, 100 mM NaCl). At this point, a small amount of FMN was added. Cells were lysed by passage through a French press and the soluble proteins were isolated by centrifugation at 16000 rpm for 30 minutes. The pure protein was pooled and subsequently 3 mM of DTT was added.

The 558 bp dsDNA handles were generated in large quantities by PCR using Taq DNA polymerase. Usually 400 mg of handles were generated at a time using 7 ml of PCR reaction. The two types of handles were generated using the primer 5' Thiol-CAGTTC TCCGCAAGAATTG together with either the primer 5' Bio-GGAATCTTGCACGCC CTCGC or the primer 5′ digoxigenin-GGAATCTTGCACGCCCTCGC. The PCR products were purified using HiSpeed Plasmid Maxi Kit, from QIAGEN.

The two types of handles are mixed in equal amounts to obtain digoxigenin/biotin handles. Then, the handle mixture was reduced with 30 mM DTT at room temperature for 1 h and concentrated down to 50-60 mL with a 30-kDa MWCO Microcon centrifuge tube. Reducing agents are removed from the handles by sequentially spinning them through three Micro Bio-Spin P6 columns equilibrated with the spin column buffer. The resulting DNA molecules were immediately reacted with a thiol-pyridine activated protein solution (protein molar ratio of 4:1; typically, $\sim$20 mM of DNA handles are reacted with $\sim$5 mM of activated protein). The reaction is allowed to proceed overnight at room temperature.

Acknowledgments

First and foremost, I want to thank Professor Felix Ritort for giving me the chance to work in his laboratory in such an interesting project. His help and encouragement have been essential to achieve this work. Secondly, special thanks go to Anna Alemany, who helped me through the whole project, especially during the first weeks, that were particularly hard because nothing worked as expected, and without whom it would not have been possible to complete this project. Thanks to Sílvia Frutos, who expressed the protein Flavodoxin sample with which I have worked during the project, and to all of my colleagues in the laboratory. Lastly, I thank my family and my friends for their support, in particular, my office mates Xavier, Víctor, Pere, Laia, Óscar, David and Berta, for fruitful discussion through the course of this work.

[1] Yoshida M., Muneyuki E., Hisabori T. "ATP synthase - a marvellous rotary engine of the cell". Nature Rev. Mol. Cell. Biol. **2**, 669-677 (2001).

[2] Salawu F.K., Umar J.T., Olokoba A.B. "Alzheimer's disease: a review of recent developments.". Ann. Afr. Med. **10** (2): 73-79 (2011).

[3] Levinthal, C. "Are there pathways for protein folding?". Journal de Chimie Physique et de Physico-Chimie Biologique **65**: 44-45 (1968).

[4] Wthrich K. "The way to NMR structures of proteins". Nat. Struct. Mol. Biol. **8** (12): 923-925 (2001).

[5] Kelly S.M., Jess T.J., Price N.C. "How to study proteins by circular dichroism". Biochim. Biophys. Acta **1751**: 119-139 (2005).

[6] Ilari A., Savino C. "Protein structure determination by x-ray crystallography". Methods Mol. Biol. **452**: 63-87 (2008).

[7] Johnson C.M. "Differential scanning calorimetry as a tool for protein folding and stability". Arch. Biochem. Biophys. **531** (1-2): 100-109 (2013).

[8] Ritort F. "Single-molecule experiments in biological physics: methods and applications". J. Phys. C **18**: 531-583 (2006).

[9] Joo C., Balci H., Ishitsuka Y., Buranachai C., Ha T. "Advances in Single-Molecule Fluorescence Methods for Molecular Biology". Annu. Rev. Biochem. **77**: 51-76 (2008).

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

[11] Sancho J. "Flavodoxins: sequence, folding, binding, function and beyond". Cell. Mol. Life Sci. **63** (7-8): 855-864 (2006).

[12] Dill K.A., Ozkan S.B., Shell M.S., Weikl T.R. "The Protein Folding Problem". Annu. Rev. Biophys. **37**: 289-316 (2008).

[13] Humphrey W., Dalke A., Schulten K. "VMD - Visual Molecular Dynamics". J. Molec. Graphics **14** (1): 33-38 (1996).

[14] Genzor C.G., Perales-Alcon A., Sancho J., Romero A. "Closure of a tyrosine/tryptophan aromatic gate leads to a compact fold in apo flavodoxin". Nat. Struct. Biol. **3**: 329-332 (1996).

[15] Rao S.T., Shaffie F., Yu C., Satyshur K.A., Stockman B.J., Markley J.L., Sundarlingam, M. "Structure of the oxidized long-chain flavodoxin from Anabaena 7120 at 2 A resolution". Protein Sci. **1**: 1413-1427 (1992).

[16] Alemany A. *Dynamic force spectroscopy and folding kinetics in molecular systems.* (Doctoral thesis, Barcelona, 2014).

[17] Ashkin A. "Acceleration and Trapping of Particles by Radiation Pressure". Phys. Rev. Lett. **24** (4), 156 (1970).

[18] Ashkin A. et al. "Observation of a single-beam gradient force optical trap for dielectric particles". Optics letters **11** (5): 288-290 (1986).

[19] Alemany A., Ritort F. "Determination of the elastic properties of short ssDNA molecules by mechanically folding and unfolding DNA hairpins". Biopolymers **101** (12): 1193-1199 (2014).

[20] Bustamante C., Marko J., Siggia E.D., Smith S.B. "Entropic elasticity of λ-phage DNA". Science **265** (5178): 1599-1600 (1994).

[21] Tskhovrebova L., Trinick J., Sleep J., Simmons R. "Elasticity and unfolding of single molecules of the molecules of the giant muscle titin". Nature **387** (6630): 308-312 (1997).

[22] Shank E.A., Cecconi C., Jesse W.D., Marqusee S., Bustamante C. "The folding cooperativity of a protein is controlled by its chin topology". Nature **465**: 637-641 (2010).

[23] Kramers H.A. "Brownian motion in a field of force and the diffusion model of chemical reactions". Physica **7** (4): 284-304 (1940).

[24] Bizarro C.V., Alemany A., Ritort F. "Non-specific binding of Na^+ and Mg^{2+} to RNA determined by force spectroscopy methods". Nucleic Acids Res. **40** (14): 6922-6935 (2012).

[25] Bell G.I. "Models for the specific adhesion of cells to cells". Science **200** (4342): 618-627 (1978).

[26] Crooks G.E. "Path ensemble averages in systems driven far from equilibrium". Phys. Rev. E **61**: 23-61 (2000).

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

[28] Ritort F., Alemany A., Ribezzi M. To be published in New Journal of Physics (2015).

[29] Ribezzi M., Ritort F. "Force Spectroscopy with Dual-Trap Optical Tweezers: Molecular Stiffness Measurements and Coupled Fluctuations Analysis". Biophys. Journal **103**: 1919-1928 (2012).