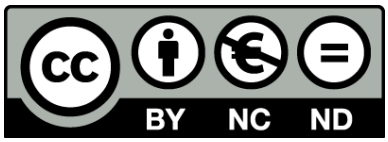
UNIVERSITAT DE BARCELONA

Thermodynamics and Kinetics of Nucleic Acids Folding

Paolo Rissone



Doctoral Thesis

Thermodynamics and Kinetics of Nucleic Acids Folding

Paolo Rissone

Universitat de Barcelona

UNIVERSITAT DE BARCELONA

Universitat de Barcelona

Facultat de Física

Thermodynamics and Kinetics of Nucleic Acids Folding

Memòria presentada per optar al grau de doctor per la Universitat de Barcelona

Paolo Rissone

Director: Dr. Felix Ritort Farran
Tutor: Dr. Giancarlo Franzese
Gener de 2023

CONTENTS

LIST OF ABBREVIATIONS

A	adenine
AD	anti-digoxigenin
AFM	atomic force microscopy
BAR	Bennett acceptance ratio
BE	Bell-Evans
BEL	barrier energy landscape
bp	base pair
C	cytosine
CEBA	continuous effective barrier approach
CFT	Crooks fluctuation theorem
DNA	deoxyribonucleic acid
ds	double-stranded
EBAR	extended Bennett acceptance ratio
EFR	extended fluctuation relations
EJE	extended Jarzynski equality
FDC	force-distance curve
FEC	force-extension curve
FEL	free energy landscape
FR	fluctuation relations
G	guanine
KD	kinetic diffusion
LOT	laser optical tweezers
MT	magnetic tweezers
NA	nucleic acid

NN nearest neighbor
NNBP nearest neighbor base pair
PBS polarizing beam-splitter
PSD position sensitive detector
RNA ribonucleic acid
SA streptavidin
SGD stochastic gradient descent
SM single molecule
ss single-stranded
T thymine
TJ temperature-jump
TS transition state
U uracil
UO unified oligonucleotide

1

INTRODUCTION

Aerodynamically, the bumblebee shouldn't be able to fly, but the bumblebee doesn't know that, so it goes on flying anyway.

Mary Kay Ash

The biological world is subject to the laws of physics: from the mesoscopic scale to entire ecosystems, biophysics aims at explaining natural phenomena with the tools and methods of physics [1]. However, the differences between the living systems studied by biologists, and the inorganic matter that is traditionally the domain of the so-called hard sciences, are discouraging. On the one hand, biology tends to be descriptive and complex, while physics always tries to explain the essential features underlying an ensemble of observations by proposing synthetic, unifying theories. Indeed, life is complex, and understanding this complexity requires very different approaches with respect to those used for studying inert matter. Since the second half of the XX century, physics has increasingly rejected a strictly reductionist approach. The idea that phenomena may emerge at each level of observation instead of being bottom-up dictated by "laws" has been gaining ground. This led to the discovery that stochastic processes operate everywhere in the universe,

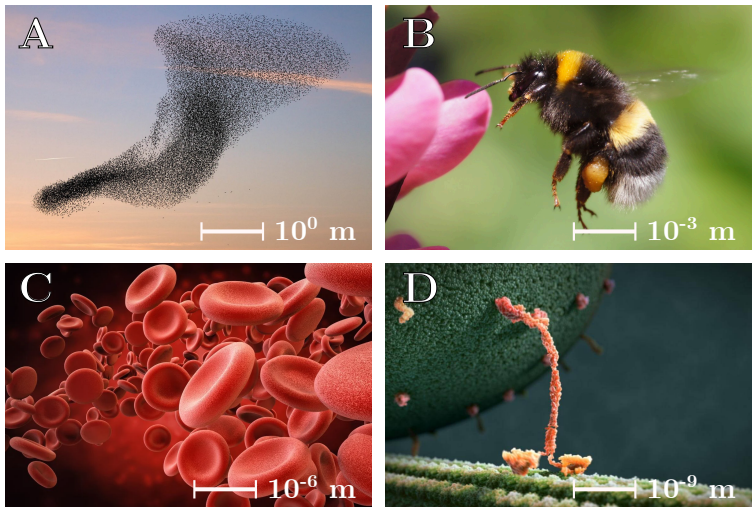


Figure 1.1: From the macroscopic to the mesoscopic scale, biophysics studies a wide variety of biological phenomena such as the collective behavior of bird flocks (A), insects flight (B), effects of mechanical deformation of red blood cells (C) and transport of cellular cargos toward microtubules opposite extremities by kinesin motor proteins (D).

at every level, from subatomic particles to weather systems, to ocean currents, to galaxies. Deterministic physical laws on the macroscopic scale left room for the random behavior of molecules on the smaller scale. The progressive discovery of similitudes between physical events, notably in mechanics, energetics, electricity, and corresponding processes occurring inside living cells, has been the vital motor of the increasing interest in biophysics. Over the last century, biophysics had increasing success, contributing to understanding a wide variety of biological phenomena (see Fig.1.1), from bumblebee flight "infamous" problem [2] and the collective behavior of bird flocks and fish schools [3], to the transport of cellular cargos along microtubules by kinesin motor proteins [4], and the link between red blood cells mechanics and pathologies [5].

In particular, studying single molecules (SMs) has become central to biology and physics research. SM experiments allow for measuring forces in the range of a few piconewtons and have a spatial resolution of a few nanometers. These are the ranges of forces and extensions typically involved in many biomolecular reactions where high-energy bonds are hydrolyzed and the energy released is subsequently used to perform mechanical work. Most of these reactions take place in an aqueous environment, so we can imagine that a specific molecular reaction is continuously influenced by the interaction with hundreds of water molecules per second. As the average kinetic energy carried by one water molecule is comparable with that of a biochemical process, such a process takes place in a highly noisy environment, i.e. in the presence of strong Brownian fluctuations. Under these conditions, it is reasonable to expect that, from time to time, water molecules with kinetic energy much higher than the average will impinge the molecular reaction. Such fluctuations can cause mutations during the replication processes of DNA when a new strand is synthesized from the parental strand, and the genetic information is transmitted to a new generation of cells. Will these large deviations affect the performance of a generic molecular reaction? In what way will they alter the molecular function? Even more interesting, are these deviations an integral part of the function and efficiency of the enzyme? Such questions are just a few among many others that biophysicists and statistical physicists are ready to confront.

1.1 BRIEF HISTORY OF BIOPHYSICS

Despite its success in modern years, biophysics is a relatively young field of science and has flourished chiefly over the last century. Probably, the first biophysicist in the contemporary sense was Luigi Galvani (Bologna, 1737 – 1798), who, around 1780, discovered animal electricity with his

famous experiments on frogs (from which the term ‘galvanic current’ for the electric currents generated by acid-salts solutions) [6]. During the XIX century, many prominent scientists became interested in biological phenomena. In the 1840s, the so-called "Berlin school of physiologists", including Hermann von Helmholtz, Emil DuBois-Reymond, Ernst von Brücke, and Carl Ludwig, started to conduct a "systematic investigation of an extensive field of physiological phenomena in accordance with the most rigorous physical methods" [7]. Their scientific approach to organic physics still sounds like an ideal description of the scope and aims of modern biophysics: “a vital phenomenon can only be regarded as explained when it has been proven that it appears as a result of the material components of living organisms, interacting according to the laws which those same components would follow in their interaction outside of living systems”. However, it is only with the discovery of X-rays by Wilhelm Röntgen in 1895 [8] that scientists were provided with a technique to investigate the matter at a mesoscopic level. These developments were fundamental in all fields of science. The use of X-rays, confined initially to the inorganic domain, would have an enormous impact on biology and medicine. In 1944, the physicist Erwin Schrödinger published the book *What is Life?* [9]. In this work, he predicts the structure of the gene as an "aperiodic crystal", building his model on fundamental biological observations. The book essentially suggested a physical approach to the questions concerning biology setting the basis of modern biophysics. A few years later, in 1953, Watson and Crick discovered the double-helix structure of the human genome by X-ray diffraction [10]. This achievement represents the starting point of biophysics, an independent field of science mainly focused on the structure and interaction of biomolecules and cells.

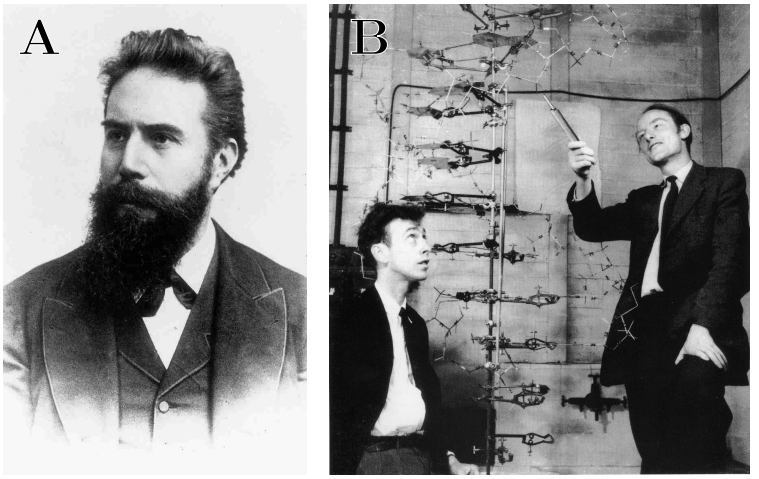


Figure 1.2: (**A**) Portrait of Wilhelm Conrad Röntgen (1845 – 1923), who discovered the X-rays, which set the basis for countless applications across all scientific fields. For his breakthrough, Röntgen has been awarded the inaugural Nobel Prize in Physics in 1901. (**B**) James Watson and Francis Crick show the double-helix structure of DNA in 1953.

1.2 FROM PHYSICS TO BIOLOGY AND BACK

The success of biophysics as an independent scientific discipline elicited a fundamental question: What is biophysics? Is it the study of biological systems through physical methods or the development of new physics by observing biological phenomena? A sharp division between these two definitions cannot be drawn. Physics contributes to biology by providing experimental techniques and mathematical tools. At the same time, biology provides new physical phenomena that physicists try to understand by formulating new concepts and laws.

An example of the latter approach is the experimental verification of the Worm-Like Chain (WLC) model. One of the first SM pulling experiments revealed that the elastic response of individual double-stranded DNA (dsDNA) molecules are excellently described by the WLC

model introduced in polymer theory by Kratky and Porod in 1949 [11]. By pulling an individual molecule, it was possible to experimentally measure the force as a function of the molecular extension, also called the force-extension curve (FEC) [12, 13]. These results verified the prediction of the worm-like chain model for the FEC and provided the first direct mechanical measurement of the persistence length of individual DNA molecules. In this case, little contribution was made to biology, but the experiment revealed a suitable environment to test a purely physical model. On the other hand, a clear example of a biological approach to biophysics is the discovery of the double helix of DNA. No new physics was applied in this case, as X-ray diffraction was already known.

Nevertheless, the development of biophysics in the last decades progressively left behind this dualistic view. The constant "dialogue" between physics and biology eventually shaped biophysics as an independent discipline defined by its scientific questions. At the same time, biophysicists have become a new kind of researcher. They do not want to explain the biological processes in detail or with an ideal model. Instead, they face scientific issues by extracting general laws of natural systems without excessive simplification [14].

1.3 SUMMARY OF THE THESIS

This doctoral work investigates nucleic acids' thermodynamic and kinetic properties. The main objective is the characterization of the energetics and the folding mechanisms driving the hybridization of DNA and RNA molecules. A rigorous study of these processes is key to understanding the diversity of behaviors observed for nucleic acids and predicting their main features. The thesis is organized into four main parts.

In Part I, we overview single-molecule force spectroscopy and some of the most common and relevant experimental techniques (Chapter 2).

Among these, we describe optical trapping with laser optical tweezers, the experimental method used to carry out this work. The experimental setup is also accurately described (Chapter 3). Then, we discuss the biological concepts and the statistical tools used in the thesis. This includes fluctuation relations (Chapter 4), nucleic acids structure, thermodynamic modeling of the unzipping experiments (Chapter 5), and the transition state theory in thermodynamic equilibrium (Chapter 6).

In Part II, we report the results of calorimetry force spectroscopy experiments on long DNA hairpins. We studied the temperature dependence of free energy, entropy, and enthalpy by carrying out unzipping experiments in the temperature range of $7 - 42°C$. Even though the effects of temperature are known to be non-negligible, an accurate characterization of the thermodynamics parameters at the single base pair level still needs to be improved. Therefore, we developed a powerful method to accurately assess the temperature dependence of the entropy and enthalpy parameters, ultimately permitting us to measure the specific heat change per base pair.

In Part III, we report the study of the energetics and kinetics of RNA folding, focusing on the complex mechanisms underlying RNA hybridization. By mechanically unzipping a long RNA hairpin, we derived the ten nearest-neighbor base pair RNA free energies in sodium and magnesium (Chapter 8). To characterize the irreversibility of the unzipping–rezipping process and the folding dynamics, we hypothesize that stem-loops structures forming along the unpaired RNA strands drive the folding (Chapter 9). This phenomenon is modeled by introducing a barrier energy landscape of the stem-loop structures forming along the complementary strands, which compete against the formation of the native hairpin.

Finally, in Part IV, the results of pulling experiments of short RNA hairpins at low temperatures are reported. After reaching $5 - 7°C$, short RNA sequences designed to fold as simple duplexes exhibit misfolded

states competing with the native fold. Despite different sequences forming different misfolded structures, all of them share common features: they are very compact and brittle, and their stability does not depend on the presence of monovalent or divalent ions. RNA cold misfolding appears to be a general phenomenon questioning our understanding of RNA folding.

Part I

PRELIMINARIES

2

SINGLE MOLECULE BIOPHYSICS

In molecular biophysics, accurate knowledge of the thermodynamics of nucleic acids (NAs) and proteins is essential to obtain reliable predictions of protein folding [15–18], DNA/RNA hybridization [19], and their interactions with enzymes and ions [20]. Over the past century, most of the knowledge on the functioning and structure of biomolecules has been obtained through bulk techniques such as UV absorbance, fluorescence, and calorimetry, among others [21]. A particularly relevant application regards the study of DNA and RNA hybridization, critical reactions in many biochemical processes, such as NA synthesis, RNA folding, and DNA amplification by PCR. The energy parameters used to model the hybridization reaction has been directly obtained from melting curves of oligos of varying sequence and length. Unified energy parameters have been derived from DNA and RNA melting temperature data obtained by many laboratories worldwide [22, 23] and are currently used as reference values by many prediction tools [24].

However, a typical bulk experiment involves studying a small volume (μl) of a molecular sample at (μM) concentration. Even in these conditions, more than $N \sim 10^{12}$ molecules are contained in each sample so that bulk techniques yield results that are incoherent temporal averages over a large population of molecules that are in different states (Fig.2.1, left). In these conditions, statistical fluctuations are of the order

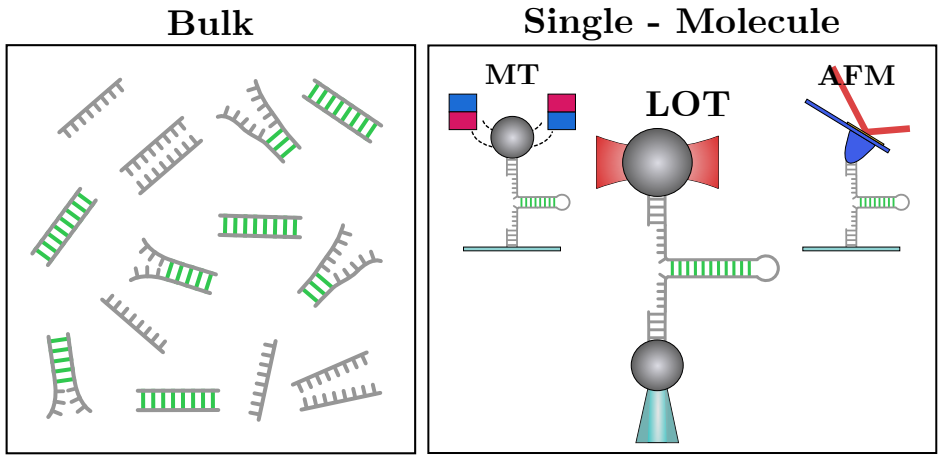


Figure 2.1: Information about biomolecules from bulk essays (**left**) is not a straightforward procedure and is obtained from the global behavior of a large number of molecules. On the contrary, single-molecule techniques (**right**) such as atomic-force microscopy (AFM), magnetic tweezers (MT), and laser optical tweezers (LOT), allow us to sample reactions one molecule at a time, characterizing the molecules at the microscopic level.

of $1/\sqrt{N}$ and become negligible in the large N limit. Therefore, the measured signal depends on the dominant species and reactions, limiting the capability of detecting fast events, rare non-native states, and reaction pathways [25]. For example, RNAs and proteins often become trapped in non-productive, misfolded structures [17, 26–30]. Such structures have been related to the development of many phenotype diseases such as Huntington's disease, fragile X-associated tremor ataxia syndrome, myotonic dystrophies, and spinocerebellar ataxias, among others [31–33]. In contrast, the development of single-molecule (SM) techniques allowed for the study of individual molecules giving access to their kinetics and the characterization of the coupling between the observed mechanical steps and the chemical reactions that drive them. In SM systems, the particle number, N, is of order 1 so that $1/\sqrt{N} \sim 1$: fluctuations are of the same order of magnitude as the measured interaction between the system and

the thermal bath [34, 35]. If the measurements are averaged over many events (or are carried out over long times), SM experiments approach the bulk limit to which results can be compared [36].

2.1 EXPERIMENTAL TECHNIQUES

Single-molecule measurements overcame the intrinsic limitations of bulk essays, revolutionizing the field of statistical physics and inspiring the development of new physical theories to understand non-equilibrium phenomena. Among the SM techniques developed in the past years, three of them stand out due to their widespread use and versatility: atomic force microscopy (AFM), magnetic tweezers (MT), and laser-optical tweezers (LOT) [37, 38] (Fig.2.1, right panel). Altogether, they cover the whole range of forces relevant to biomolecular reactions, from polymer entropic forces ($\sim 10^{-3}$pN) to the typical forces needed to break the covalent bonds of the proteins and NA ($\sim 10^{3}$pN). Here we will briefly discuss the main features of each technique and their typical applications (summarized in Table 2.1).

	AFM	MT	LOT
Force Range (pN)	$10^{1} - 10^{4}$	$10^{-2} - 10^{1}$	$10^{-1} - 10^{3}$
Spatial Resolution (nm)	0.1	1	1
Stiffness (pN/nm)	$10^{1} - 10^{5}$	10^{-6}	$10^{-2} - 10^{0}$
Temporal Resolution (s)	10^{-3}	$10^{-1} - 10^{-3}$	10^{-4}
Probe Size (μm)	$100 - 250$	$0.5 - 5$	$0.25 - 5$

Table 2.1: Comparison between different single-molecule techniques (AFM, MT, and LOT) [37].

2.1.1 *Atomic-Force Spectroscopy*

The AFM [39] consists of a very soft cantilever with a sharp tip at its end that is used to probe the roughness of a surface. The tip is moved to the vicinity of the surface (metallic or insulating) and deflects by an amount proportional to the tip's proximity to the surface. The most important application of the AFM is imaging, which can work in various modes: the contact mode, tapping mode, and jumping mode [40]. For example, in the tapping mode, the tip oscillates close to the sample surface. The oscillation amplitude is recorded and controlled by a feedback loop that keeps such amplitude constant. The amplitude decreases when passing over a bump, so the distance between the tip and surface is increased to keep the oscillation amplitude constant. When passing over a depression, the tip is moved to the surface. The map of the tip's distance from the sample provides an accurate topographic image of the surface. The AFM is also used to manipulate individual molecules. In this case, the surface and tip must be conveniently treated to work with biological samples. The surface has to be coated with the molecules to be manipulated.

The AFM covers forces in the range $10^1 - 10^4$pN, depending on the stiffness of the cantilever. Typical values of the stiffness are in the range $10^1 - 10^5$pN/nm. Although AFM is a very versatile tool, it has a few drawbacks for manipulating SMs. The most important one is probably the presence of undesired interactions between tip and substrate (Van der Waals, electrostatic, and adhesion forces) and the non-specificity of the attachments that often occur between tip and substrate. Moreover, thermal fluctuations limit spatial and force resolution in the AFM. When the cantilever stage is held at a constant position, the force acting on the tip and the extension between the tip and substrate fluctuate. The respective fluctuations are given by $\langle \delta x^2 \rangle = k_B T/k$ and $\langle \delta f^2 \rangle = k_B T k$ where k_B is the Boltzmann constant, T is the absolute temperature of

the environment, and k is the stiffness of the cantilever. At room temperature ($k_B T \approx 4.1$pNnm), this respectively gives $\sqrt{\langle \delta x^2 \rangle} \approx 20$nm and $\sqrt{\langle \delta f^2 \rangle} \approx 20$pN if we take $k \approx 100$pN/nm . Therefore, the signal-to-noise ratio of the force is small for forces of just a few tens of piconewtons corresponding to the force range characteristic of weak interactions. This makes AFM unsuitable for studying the mechanochemistry of weak interactions in the lower piconewton regime. In contrast, AFM is ideal for investigating strong to covalent interactions.

However, developments in AFM microscopy over the past two decades yielded remarkable improvements. Among these, high-speed AFM [41] revealed a powerful technique, allowing the direct visualization of the structure and dynamics of single protein molecules in action at high spatiotemporal resolution. This new tool has been successfully applied to various biological systems, from motor proteins to membrane proteins, antibodies, and enzymes [42,43].

2.1.2 *Magnetic Tweezers*

The MT works on the principle that a magnetized bead experiences a force when immersed in a magnetic field gradient $F = -\mu \nabla B$. A bead is trapped in the magnetic field gradient generated by two strong magnets. Molecules are attached to the surface of the magnetic bead on one end and a glass surface on the other. A microscopic objective with a CCD camera is used to determine the bead's position. Molecules are pulled by moving the translation stage that supports the magnets. Moreover, the rotation of the magnets permits twisting molecules: modifying the basic set-up by using a third bead to create a single chemical bond swivel allows the measurement of torques [44]. MT permits the measurement of very weak forces due to the low value of the stiffness of the magnetic trap ($\sim 10^{-6}$pN/nm). The typical force range is $10^{-2} - 10^{1}$pN, where

the force's maximum value depends on the magnetic bead's size. An advantage of MT with respect to AFM and LOT is the possibility to do experiments in passive constant-force mode, i.e. by keeping the magnet stage fixed. The force can be kept constant because the spatial region occupied by the bead is small enough for the magnetic field gradient to be considered uniform. A constant force protocol can also be implemented in the AFM and LOT setups by implementing force-feedback control mechanisms. The main drawback of feedback loops is their working frequency, typically limited to a few kilohertz, which does not allow for the detection of dynamical processes faster than milliseconds.

2.1.3 *Laser Optical Tweezers*

LOTs use the optical gradient force generated by a focused beam of light to trap an object with an index of refraction higher than the surrounding medium. In the basic experimental LOT set-up, a near-infrared laser is collimated by a high numerical aperture water immersion lens. Micron-sized polystyrene or silica bead is then trapped in the laser's focus by exerting forces in the range $10^{-1} - 10^{2}$pN depending on the size of the bead and the power of the laser. SM manipulation is possible by tethering the molecular construct between two beads: one is held to a fixed point while the other is optically trapped. Therefore, moving the optical trap makes it possible to exert a force on the molecule.

The trapping potential can be considered harmonic with excellent approximation. Therefore forces acting on the bead are directly proportional to the distance between the bead and the center of the trap, $F = kx$, where k is the stiffness constant of the trap. To determine the stiffness of the trap, noise measurements or Stokes force calibration are often used. The typical trap stiffness varies in the range $10^{-2} - 10^{0}$pN/nm ($10^{2} - 10^{4}$ times smaller than AFM tips), making LOT force resolution on

the order of 0.1pN. This basic setup is significantly improved by using dual counter-propagating laser beams passing through two identical objectives [45]. There are several advantages to this more complex setup. First, the axial scattering force is reduced. Second, the trapping forces that can be reached (up to 10^2pN) are higher than in the one-beam set-up. Finally, continued force calibration is not required because the force is directly measured from the total amount of light deflected by the bead. The bead position is measured using a light lever or reference beam, where a low-power light beam passing through a small lens in the chamber's frame is collected using additional position-sensitive detectors. This allows us to determine the molecular extension with a few nanometers of precision.

The experimental work in this thesis has been entirely carried out with LOT. A detailed description of the principles of optical trapping and the experimental setup is given in the next chapter.

3

OPTICAL TWEEZERS

In 1970, Arthur Ashkin demonstrated that the pressure of radiation exerted by a focused laser beam could be used to accelerate a micrometer-sized neutral particle suspended in liquids or gasses [46]. These first experiments set the basis for developing a groundbreaking technique. In 1986, Ashkin and colleagues reported the first realization of laser optical tweezers (LOT) [47] consisting of a focused laser beam controlling the position of a particle in three dimensions. Thanks to this work, in 2018, Ashkin was awarded the Nobel Prize in Physics [48]. Starting from the late 1980s, the impact of LOT in single-molecule biophysics started to grow. In 1990, Ashkin and colleagues used LOT to manipulate biological samples, starting from an individual tobacco mosaic virus and *E. Coli* bacterium [49]. Since then, LOT found many different applications across many fields of physics [50–52], nanotechnology [53], soft matter [54], and biology [55], among the others.

In particular, LOT revealed a powerful tool to investigate the properties of NAs. LOT permits the direct measurement of the mechanical work (and therefore of the free energy) needed to unfold DNA and RNA hairpins, rendering force spectroscopy a valuable tool for NA thermodynamics [56, 57]. Moreover, single-molecule manipulation sets a new bar for resolving complex molecular reactions [34], such as NAs' elastic

response [58, 59] and non-specific secondary structure formation [60], as well as their protein and NA folding [61–63].

3.1 PRINCIPLES OF OPTICAL TRAPPING

The interaction between a collimated laser beam and a dielectric particle can be modeled according to the physical law of conservation of momentum and ray optics (Fig.3.2, A). Given one ray of light coming from the laser beam, its linear momentum can be written as

$$\vec{p} = n\frac{h}{\lambda}\hat{e}_\lambda \,, \tag{3.1}$$

where n is the number of ray photons, h is the Plank constant, λ is the light wavelength, and $\hat{e}_\lambda$ is the unitary vector indicating the propagating direction of the ray. When the ray reaches the particle surface, a fraction of it is reflected, and the rest refracts according to the Snell law. The refracted ray of light propagates through the sphere until it reaches the other edge surface, where reflection and refraction occur again (red arrows in Fig.3.2). Usually, the fraction of the ray reflected on a transparent particle is much smaller than the transmitted one, so the former can be neglected. The amount of momentum transferred to the bead per unit of time defines the force applied to the particle

$$\vec{f} = \frac{d\vec{p}}{dt} \,. \tag{3.2}$$

This force can be decomposed into two components: a scattering force, f_{scat}, in the direction of light propagation, and a gradient force, f_{grad}, in the direction of the spatial light gradient (black arrows). Notice that the intensity of the force depends on the intensity of the incident light. In the case of a collimated Gaussian laser beam, the scattering force always

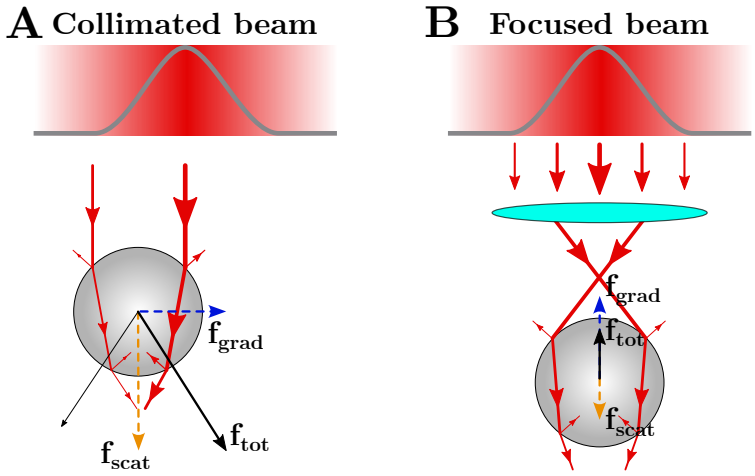


Figure 3.2: Ray optics description of the force exerted on a particle by an incident beam. (**A**) Interaction between a collimated Gaussian laser beam and a dielectric sphere. The refraction and reflection of a ray of light on the sphere surfaces (red arrows). The resulting radiation pressure can be decomposed into two components, the scattering force f_{scat} (orange arrow) and the gradient force f_{grad} (blue arrow). The resulting force, f_{tot} (black arrow), pushes the particle forward to the maximum intensity central region. (**B**) Optical trapping with a focused (not collimated) laser beam. The addition of all light rays causes a net gradient force (f_{grad}) that compensates for the scattering force (f_{scat}), generating a restoring axial force (f_{tot}) that pushes the particle toward the trap center.

pushes the particle along the beam direction so that the particle cannot be trapped in a region of space.

However, optical trapping can be achieved by focusing the light beam (see Fig.3.2, B). In this case, the rays of light reaching the particle create a force gradient that can be positive or negative depending on the position of the center of the sphere with respect to the focal point. Thus, the resulting force pulls the sphere toward the trap's center. Combining this axial restoring force with the radial component produces a 3-dimensional optical trap. For stable trapping in all three dimensions, the axial gradient component of the force (blue arrow) pulling the particle toward the focal

region must exceed the scattering component of the force (orange arrow) pushing it away from that region. This condition necessitates a steep gradient in the light, produced by sharply focusing the trapping laser beam to a diffraction-limited spot. This can be achieved using an objective with a high numerical aperture (NA).

Notice that, in a typical LOT experimental setup, the trapped beads have a diameter $d \approx 2 - 4\mu$m, while the beam wavelength is $\lambda \approx 1\mu$m so that $d \sim \lambda$. In this regime, geometrical optics correctly describes optical trapping (whereas it does not when $d \ll \lambda$).

3.2 EXPERIMENTAL SETUP

The experimental setup used in this thesis has been invented by Steve Smith and Carlos Bustamante [45, 64–66], which named it *miniTweezers*. The instrument is a miniaturized evolution of the original version by Ashkin, providing better stability and higher measurement resolution. The miniTweezers uses two counter-propagating laser beams to create a single optical trap [67, 68]. This is achieved using two microscope objectives with high NA focused on the same point. The light beam focused on one objective is collected by the opposite one after it emerges from the trap. The fundamental principle of the measure is based on detecting the change in the light momentum. One of the advantages of a dual-beam setup is that the trapped particle is subject to equal but opposite scattering forces that cancel out. This permits us to reduce the diameter of the laser beam (low NA) and collect all the deflected light exiting the trap using high NA objectives. Moreover, the Instrument calibration for force measurements is independent of several experimental conditions (bead size, index of refraction, etc.). Besides, the dual-beam optical trap presents other advantages to the single-beam trap [14]. First, the laser beams do not need to be highly focused, which minimizes

the effect of spherical aberration of the lenses. A lens with spherical aberration focuses the marginal rays of a laser beam more tightly than those near the optical axis, producing a blurred focal point. Since the intensity of a low NA laser beam is concentrated near the optical axis, such a beam is less affected by spherical aberration. Moreover, a low-focused beam has a longer focal distance, which makes it possible to focus the laser beam deeper inside the fluidics chamber. It reduces the hydrodynamic effects of the boundaries of the fluidics chamber (i.e., the coverslips) on the particle trapped in the optical trap. Moreover, using low-focused lasers reduces the heating of the medium by infrared absorption.

3.2.1 *The MiniTweezers Setup*

The schematic of the MiniTweezers optics is symmetric for each counter-propagating laser. As shown in Fig.3.3, the lasers partially share the same optical path but are generated, controlled, and measured independently. The instrument features two near-infrared laser diodes of power $P = 200\text{mW}$ and wavelength $\lambda = 845\text{nm}$. The absorption of this wavelength by water (the medium in which experiments are carried out) is quite low, preventing the heating of the sample. Each laser (Lumix LU0845M200) produces a linearly-polarized electromagnetic mode TEM00 with a Gaussian profile. Each laser is connected through a single-mode optical fiber to a device called *wiggler* (one for each device). The wigglers control the direction of the beam: piezoelectric actuators allow tilting of the optical fiber changing the direction in which the light emerges.

The optical paths of counter-propagating lasers A and B are equal but opposite. Here we describe the optical path of laser A (in green in Fig.3.3). After exiting the wiggler, a beam-splitter pellicle splits the light

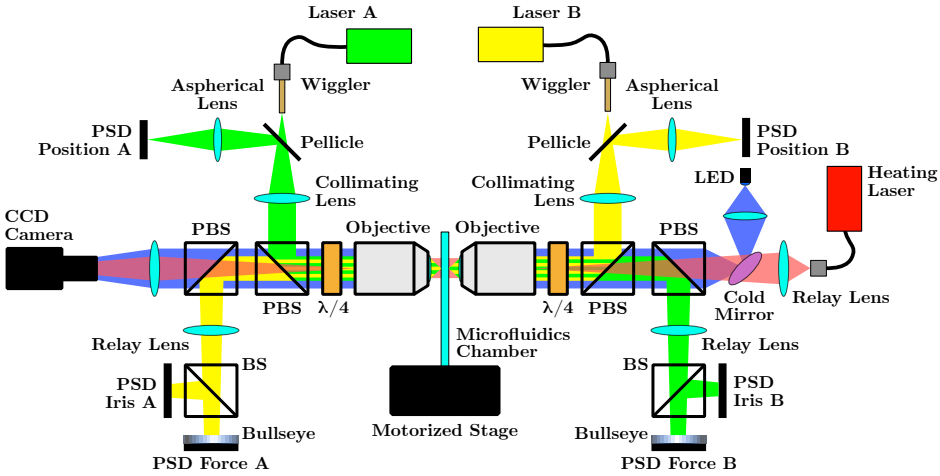


Figure 3.3: Schematics of the MiniTweezers with temperature control (see text). The optical paths of lasers A and B are depicted in green and yellow, respectively. The optical path of the imaging system is depicted in blue. Finally, the optical path of the heating laser is shown in red.

in two directions. About $\sim$ 8% of the incident light is redirected to a Position Sensitive Detector (PSD), the *light-lever*, measuring the beam position. The remaining $\sim$ 92% of the light goes through the pellicle and is collimated by a lens that forwards it to a polarizing beam-splitter (PBS).

The PBS selects the horizontally polarized light and redirects the beam toward a water-immersion microscope objective with NA=1.2 (Olympus UPLSAPO 60×W) that focuses the light forming the optical trap. Before the objective, a quarter-wave plate ($\lambda/4$) produces circular polarization, ensuring the scattered light is an average between parallel and perpendicular components.

The light exiting the trap is then collected by the opposite objective and is converted into vertically polarized light by another $\lambda/4$. The light continues on the path through two consecutive BPSs and a relay lens that redirects the light to the PSD, measuring the beam intensity.

Notice that the first BPS along the path is common to lasers A and B: it forwards the vertically polarized light of laser A at the second BPS, and at the same time, it redirects the beam by laser B toward the $\lambda/4$ and the corresponding objective. As the two beams have orthogonal polarizations (the exiting laser A beam is vertically polarized while the incoming laser B beam is horizontally polarized), they do not interact along the path.

The light that reaches the photodetectors is split into two parts using a (non-polarizing) beam-splitter. One part is redirected to a PSD detecting the deflection of the beam in the transverse direction (PSD Force A), permitting the measurement of the force along the x and y-axis (transverse force). The other part reaches a *bull's eye* (or bullseye) filter and is measured with a photo-diode (PSD Iris A). The bullseye is an attenuator with a suitable transmission profile: when the beam is axial, 100% of the light is transmitted. If the beam is off-axis, the transmitted light is attenuated, reducing the power that reaches the PSD. This attenuation signal is called *Iris* and allows for measuring the force along the z-axis (longitudinal force).

A blue LED ($\lambda = 470$nm) and a CCD camera (Watec WAT-902H3 SUPREME EIA) are used to form a microscope and view the experiment. From the LED, the light goes through a lens to create a Köhler illumination system at the focal plane of the laser beams. This allows for uniform illumination of the field of view. After the lens, a 45° dichroic long-pass filter (*cold mirror*) redirects the LED light toward the objectives.

After following identical (but opposite) optical paths, the optical trap is obtained by focusing the two laser beams on the same point. The trap is formed inside a *microfluidics chamber*, where the whole experiment is carried out. The chamber is placed between the objectives and is held by a *x, y, z motorized stage* that allows for position control. Even though the optical trap (and not the stage) is moved during the experiment, the

stage makes it possible to easily move the chamber during the experiment setup and the calibration procedure.

3.2.2 *The Temperature-Jump LOT*

A fundamental feature of the miniTweezers setup used in this work is the temperature-jump (TJ) LOT, which permits carrying out experiments in a broad range of temperatures [69,70]. To change the temperature of the system inside the microfluidics chamber, it has been used a wavelength $\lambda = 1435$nm (depicted in red in Fig.3.3).

The light emerges from a single-mode optical fiber with NA=0.14 and is focused by a relay lens on the back focal plane of the laser B objective. Notice that the *cold mirror* (see above) along the path lets the heating wavelength pass and only reflects the blue LED light. The objective expands the heating light creating a Köhler illumination cylinder of $\sim 33\mu$m that is used to heat the sample in the microfluidics chamber (Fig.3.4A). After exiting the chamber, the light focuses on the front focal plane of the opposite objective, which expands it again and forwards it to the CCD camera. Let us notice that the heating spot in the chamber cannot be seen with the mounted CCD camera because the photosensor is not sensible to the 1435nm wavelength. However, the laser position can be checked by replacing it with one of $\lambda = 975$nm (visible light). After the laser has been collocated in the center of the chamber, it is exchanged back with the heating laser.

The heating laser produces a non-uniform temperature profile inside the microfluidics chamber and a heat current. The wavelength of the heating laser has been chosen to minimize heat absorption and convective effects between regions of the medium (water) at different temperatures. In addition, the heat flux can be further reduced by minimizing the region's volume to be heated in the chamber [70]. Let us notice that

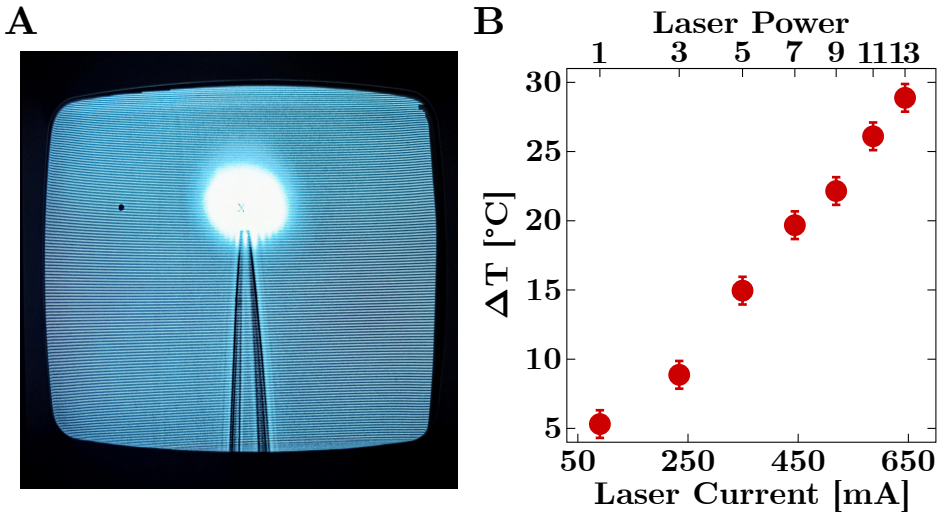


Figure 3.4: Heating laser in the temperature-jump LOT. (**A**) Köhler illumination cylinder ($\sim 33\mu$m) generated by the heating laser in the microfluidics chamber. The laser position is checked by replacing the $\lambda = 1435$nm heating laser (not visible by the CCD camera) with one of $\lambda = 975$nm (visible light). (**B**) Temperature change, ΔT, as a function of the heating laser power output. The laser output is controlled by four binary switches that can be combined to obtain a series of 16 discrete ΔT. Notice that not all the possible switches' combinations are shown here.

the standard miniTweezers setup features water immersion objectives. However, water cannot be used as it would interfere with the heating wavelength. Therefore, water has been replaced with Cargille Labs Refracting Index matching Liquid, Series AAA, N = 1.330. This liquid is placed between the objectives' lenses and the microfluidics chamber.

The output power of the heating laser is controlled by a binary set of switches incorporated into the power supply. The switches are labeled as 1, 2, 4, and 8, and each has the option to be in 1 (powered on) or 0 (powered off). By starting at $(0,0,0,0)$ (all switches are powered off, i.e. the output current of the heating laser is 0mA) up to $(1,1,1,1)$ (the laser power is maximum with an output current of 160mA), there are

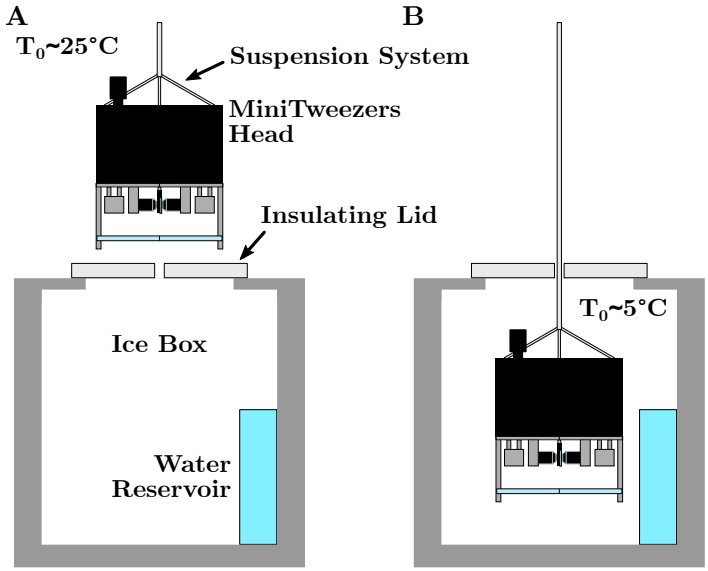


Figure 3.5: Temperautre-jump LOT experimental setup. (**A**) Experiments above room temperature. At $T_0 \sim 25°C$, the heating laser changes the temperature inside the microfluidics chamber up to $\sim 55°C$. (**B**) Experiments below room temperature. An ice box is used to lower the temperature of the miniTweezers, i.e. the "environment" temperature, to $T_0 \sim 5°C$. A previously frozen water reservoir keeps this temperature stable upon melting (see text). Inside the ice box, the heating laser can be used to raise the temperature in the proximity of the optical trap up to $\sim 30°C$.

14 different switches combinations. This gives 15 discrete temperature increases, ΔT, with respect to the environment temperature, T_0. Each laser power is identified by the sum of the numbers associated with the powered switches. For example, the combination $(1,0,0,0)$ correspond to the laser power 1 (switch 1 powered on and switches 2, 4 and 8 powered off), while $(0,0,0,1)$ and $(0,1,0,1)$ correspond to positions 8 and 10, respectively. Fig.3.4 shows the temperature change as a function of the laser power output (the corresponding switches' positions are also indicated).

The TJ-LOT allows for increasing the system temperature to a maximum of $\Delta T \approx +30°C$ with respect to T_0. Therefore, changing T_0 allows for exploring a broader range of temperatures. In particular, our miniTweezers setup can work above ($T_0 \approx 25°C$) and below ($T_0 \approx 5°C$) room temperature (Fig.3.5). To reach low temperatures, the whole instrument is placed inside a $1.5m^3$ top-opened ice box containing ~ 20 liters of water. After setting the experiments, the ice box is turned on until it reaches the (minimum) temperature of $\sim -30°C$, and the water is frozen into a block of ice. When the ice box is turned off, the ice starts to melt slowly, maintaining the temperature at a constant $4-5°C$ for several days. The LOT is then placed into the ice box using a pulley that controls a suspension system isolating the instrument head by vibrations. Inside the ice box, the temperature of the tweezers is measured by two thermometers that monitor the temperature evolution in different parts of the instrument. When the temperature inside the ice box becomes stable (generally after $4-5$ hours), it is possible to start the experiments. The temperature around the optical trap can be changed using the heating laser. Therefore, the current setup (ice box and TJ-LOT) permits us to explore a wide range of temperatures, spanning between $\sim 5°C$ and $\sim 55°C$ (the $5-30°C$ range inside the ice box plus the $25-55°C$ range at room temperature.)

3.2.3 *The Microfluidics Chamber*

The experiments are carried out in microfluidics chambers specifically designed to be used with the TJ miniTweezers (see Fig.3.6) and are realized manually.

The camber is made using two cover glasses (24mm $\times$ 60mm, No.2 VWR). One of them presents 6 holes for the buffer throughput. Between the two layers is placed a layer of Nescofilm (Nesco-Karlan) with a

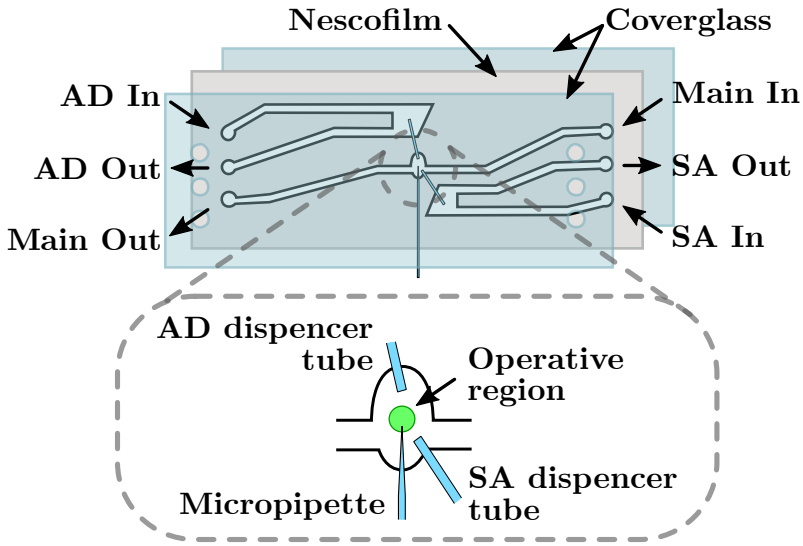


Figure 3.6: Design of the microfluidics chamber (see text).

suitable design for the TJ-LOT. Both the holes in the coverglass and the cutting of the Nescofilm are made with a laser cutter (Epilog Mini 18). Three channels are drawn in the Nescofilm (Fig.3.6, top). Each channel has a distinct input that allows for flowing different buffers (see the next section). Even though the channels' outputs are physically separated, all the buffer exiting the chamber is collected into the same waste container. As shown in Fig.3.6 (bottom), the upper and lower channels are connected to the central channel by glass dispenser tubes (King precision glass, Inc., inner diameter of 0.04mm, outer diameter of 0.10mm and length $\sim$ 6mm, glass type KG$-$33). Through these tubes, the sample can enter the central channel in a controlled way, drastically reducing undesired turbulences in the buffer flux and avoiding the sample overabundance in the chamber operative region.

A glass micropipette is placed on the Nescofilm layer, with the tip positioned in the central area of the main channel. The micropipette

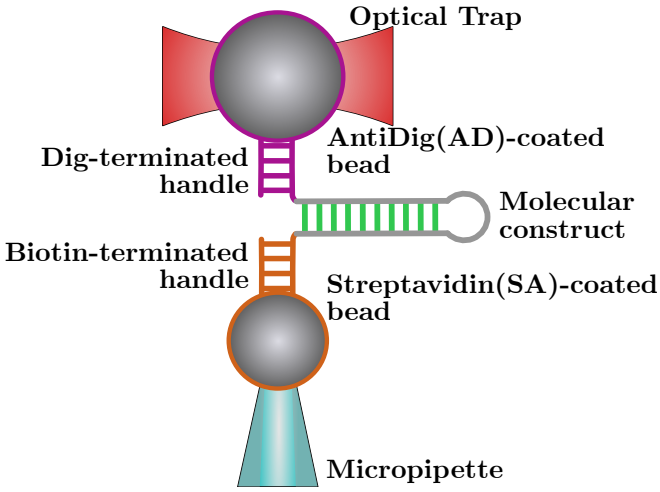


Figure 3.7: Unzipping experimental setup. The molecular construct is tethered between two beads in a dumbbell configuration: the antiDIG (AD)-coated bead is optically trapped, while the Streptavidin(SA)-coated one is held by air suction at the tip of a glass micropipette. The two tags specifically bind to the ds handles of the molecule, one labeled with a digoxigenin tail (DIG) and the other with biotin (BIO).

has a diameter of $\sim 1\mu$m and is produced by heating and pulling a glass tube (King precision glass, Inc., inner diameter of 0.04mm, outer diameter of 0.08mm, glass type KG$-$33). Notice that the design of the Nescofilm has been made to avoid the obstruction of the lower channel by the micropipette. A detailed description of the procedure to realize a microfluidics chamber can be found in [68].

3.3 PULLING EXPERIMENTS

Single-molecule unzipping experiments consist of controlling the position (or the force) of micron-sized particles (beads) that are coated with the molecule of interest (DNA, RNA, or proteins) inside a microfluidics chamber (see Sec.3.2.3). The typical experimental setup is shown in

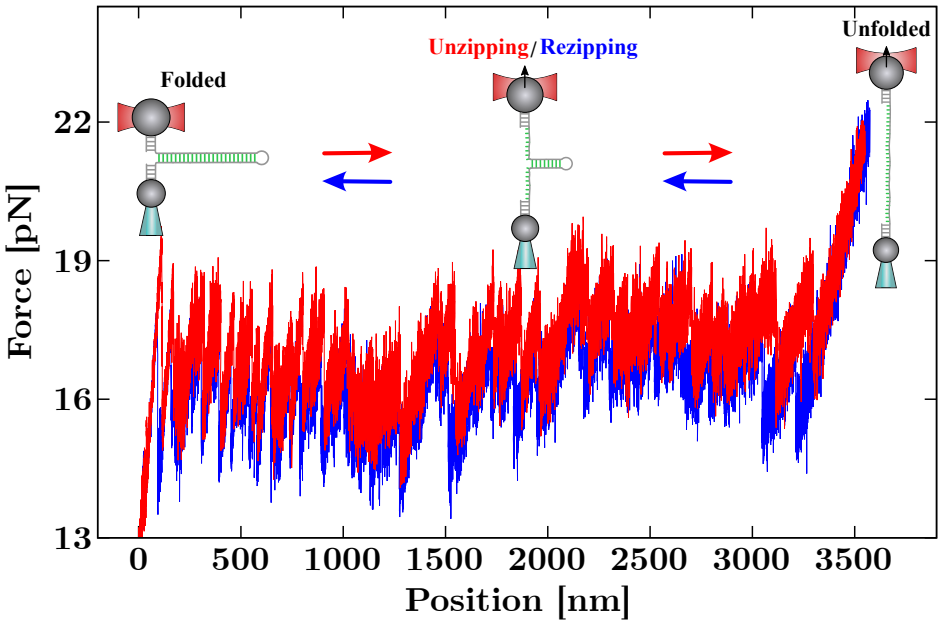


Figure 3.8: Unzipping/rezipping pulling protocol. Starting with the molecule in the completely folded configuration (the native state), the optical trap is moved with respect to the (fixed) micro-pipette at a constant speed. As the unzipping progresses, groups of bases open one after another in a stick-slip process. The reverse protocol (rezipping) starts when the molecule is completely unfolded. The FDC measured during unzipping (red) and rezipping (blue) exhibits a saw-tooth pattern that depends on the sequence of the hairpin.

Fig.3.7. The molecules are designed to end with double-stranded handles (with lengths varying from a few to hundreds of bases), one labeled with a digoxigenin tail (DIG) and the other with biotin (BIO). The two tags bind specifically to beads coated with anti-DIG (AD) and streptavidin (SA), respectively. The molecule is tethered between the two beads in a dumbbell configuration: the AD bead ($\sim 3\mu$m diameter) is optically trapped. In contrast, the SA one ($\sim 2\mu$m diameter) is held by air suction at the tip of a glass micropipette (Fig.3.7). The experiment setup proceeds as follows. AD beads (already coated with the molecule of interest) are

flown into the upper channel of the microfluidics chamber while the lower one is filled with SA beads. The central channel of the chamber is filled with the buffer in which the experiment has to be carried out. The glass dispenser tubes connecting the side channels to the main one allow the beads to enter the chamber's central region. The optical trap is used to (manually) select an SA and an AD bead and bring them to the starting position of the experiment. The different sizes of the AD and SA beads permit us to distinguish them during the experiment.

In an unzipping experiment, the optical trap is moved with respect to the (fixed) micro-pipette at a constant speed. At the beginning of the protocol, the molecule is folded into its native double-stranded (ds) hairpin configuration (Fig.3.8). As the optical trap moves away from the pipette, the force applied to the hairpin increases until the intramolecular bonds at the beginning of the stem break open. As the unzipping progresses, groups of new bases sequentially open one after another. Unzipping is a stick-slip process consisting of the succession of an elastic deformation (stick) followed by the release of groups of bases that collectively unfold cooperatively (slip), resulting in sudden force jumps. The unfolding protocol proceeds until the hairpin is fully unzipped and the single-strand (ss) form is fully stretched (Fig.3.8). At this point, the reverse process starts (rezipping), and the molecule refolds starting from the loop until the native ds hairpin has been reformed. Upon rezipping, groups of bases are cooperatively absorbed into the stem resulting in sudden increases in force. The force-distance curve (FDC) measured during unzipping and rezipping exhibits a saw-tooth pattern that depends on the sequence of the hairpin.

4

FLUCTUATION RELATIONS

In macroscopic systems, fluctuations represent just minor deviations from the average behavior. In a gas of molecules, where the number of particles $N \sim 10^{12}$, relative energy deviations are on order $1/\sqrt{N}$ with respect to the average value. This implies that the typical magnitude of the deviations is of the order of 10^{-6} making them negligible (and experimentally not visible) in bulk experiments. In contrast, the advances in SM manipulation during the last 20 years granted access to events occurring at the level of individual molecules. In this microscopic scale, work measurements and thermal fluctuations are of the same order, and repetitions of the same experiment may lead to different outcomes [35].

The study of such systems takes the name of *Nonequilibrium thermodynamics of small systems* [71] and *Stochastic thermodynamics* [72]. Since the 90*s*, the search for a thermodynamic description of small systems has become increasingly active. The results obtained during these years went under the name of fluctuation theorems (FTs) and laid the foundations for a mathematical description of energy fluctuations for nonequilibrium systems [35, 73].

The development of the FTs lead to crucial breakthroughs in SM biophysics: the measurement of the free energy of formation of DNA and RNA [22, 23, 74, 75]; the study of the stability of the proteins native domains [76]; the measurement of mechanical torque in rotary mo-

tors [77]; the computation of free energy landscapes from work measurements [78, 79]; the reconstruction of the free-energy branches for a molecular construct exhibiting multiple stable states [56, 80], among others.

4.1 CROOKS AND JARZYNSKI RELATIONS

Let us consider a system (Eg., a single molecule) initially ($t = 0$) in thermal equilibrium at state A for a given value $\lambda(0) = \lambda_A$ of the control parameter. In the case of LOT, the control parameter is the trap-pipette distance. An experimental time-dependent forward (F) protocol, $\lambda_F(t)$, is applied by increasing λ during a time interval t at a constant pulling rate, $v = d\lambda/dt$: the system starts in state A at λ_A and ends in state B at an arbitrary $\lambda(t) = \lambda_B$. The mechanical work, W, done along this process is:

$$W = \int_0^t v f dt = \int_{\lambda_A}^{\lambda_B} f d\lambda \,. \tag{4.3}$$

Now suppose that the time-reversed (R) experimental protocol is performed: the system starts at equilibrium in state B at λ_B and the control parameter changes according to the time-reverse transformation $\lambda(t-s)$ until it reaches state A at λ_A. The work measured during the reverse process is then

$$W = \int_0^t (-v) f ds = -\int_{\lambda_A}^{\lambda_B} f d\lambda \,. \tag{4.4}$$

The Crooks Fluctuation Theorem (CFT) [81] relates the mechanical work done on a system during the forward and reverse processes, and it reads:

$$\frac{P_F(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta G_{AB}}{k_B T}\right), \tag{4.5}$$

where $P_F(W)$ and $P_R(-W)$ respectively are the probability distributions of the work measured in the forward and reverse process, and $\Delta G_{AB} = G(\lambda_B) - G(\lambda_A)$ is the Gibbs free-energy difference between state A at λ_A and state B at λ_B.

The CFT implies the well-known Jarzynski equality [82]. By multiplying Eq.(4.5) by $P_R(-W)$ and integrating over W, one gets

$$\left\langle \exp\left(-\frac{W}{k_B T}\right)\right\rangle_F = \exp\left(-\frac{\Delta G}{k_B T}\right), \tag{4.6}$$

where $\langle \cdots \rangle_F$ is the average over the forward trajectories. Note that the analogous equality holds for the reverse process.

A fundamental consequence of Eq.(4.5) is that the value of the work, W^*, by which the probabilities of the forward and the reverse processes are equal, i.e. $P_F(W^*) = P_R(-W^*)$, gives the free-energy difference $W^* = \Delta G_{AB}$. This value, which also corresponds to the reversible work, is always the same no matter how far from equilibrium the system is during the experimental process. A consequence of Eq.(4.6) is the second law inequality, $\langle W \rangle \geq \Delta G_{AB}$. Notice that the CFT predicts the possibility of having trajectories where $W < \Delta G_{AB}$. In this case, transient violations of the second law of thermodynamics are observed. However, such (rare) events do not imply that the second law does not hold for small systems, as thermodynamic quantities are not evaluated over one but an infinite number of trajectories.

4.2 FREE ENERGY ESTIMATORS

In general, the extraction of free-energy differences from irreversible work measurements can be obtained using *bidirectional* or *unidirectional* free-energy estimators. The former, as the CFT in Eq.(4.5), can be applied when both the forward and the reversed protocols are feasible. How-

ever, when dissipation and hysteresis effects between the forward and the reversed processes are large, the work distributions in the l.h.s of Eq.(4.5) separate from each other up to a point where they do not cross anymore, and only unidirectional estimators can be used. Typically, the combination of information from the forward and reversed protocols provides less biased free-energy estimates. Here we briefly introduce the free-energy estimators used in this thesis. Their extension to the out-of-equilibrium case where multiple intermediates appear during the forward and reverse processes is discussed in Part III.

Let us first rearrange the terms in Eq.(4.5) to get

$$P_F(W)\exp\left(-\frac{W}{k_\mathrm{B}T}\right) = P_R(-W)\exp\left(-\frac{\Delta G_{AB}}{k_\mathrm{B}T}\right). \tag{4.7}$$

Let us now multiply both terms in Eq.(4.7) by a generic function $\phi(W)$, which gives

$$\left\langle \exp\left(-\frac{W}{k_\mathrm{B}T}\right)\phi(W)\right\rangle_F = \langle\phi(-W)\rangle_R \exp\left(-\frac{\Delta G_{AB}}{k_\mathrm{B}T}\right), \tag{4.8}$$

where $\langle\cdots\rangle_{F(R)}$ is the average over all possible forward (reverse) trajectories. By properly choosing the function $\phi(W)$, it is possible to obtain different free-energy estimators. Notice that for $\phi(W) = 1$, the Jarzynski equality in Eq.(4.6) is recovered.

4.2.1 *Bennett Acceptance Ratio Method*

The Bennett Acceptance Ratio (BAR) method [83] is a bidirectional free-energy estimator, i.e. it uses work measures from both the forward and the reverse protocols. It is obtained from Eq.(4.8) by choosing the function

$$\phi(W) = \frac{1}{1 + \frac{n_F}{n_R} \exp\left(\frac{W - \Delta G_{AB}}{k_B T}\right)}, \tag{4.9}$$

where $n_{F(R)}$ is the number of forward (reverse) work measurements. It has been proved that the use of Eq.(4.9) in Eq.(4.8) minimizes the statistical variance of the estimation of ΔG_{AB} [83, 84]. Hence, for a given set of forward (reverse) work measurements $\{W_i^{F(R)}\}_{i=1,\dots,n_{F(R)}}$, the transition free-energy from state A to state B is given by

$$\frac{u}{k_B T} = z_R(u) - z_F(u), \tag{4.10}$$

where

$$z_R(u) = \log\left[\frac{1}{n_R}\sum_{i=1}^{n_R}\left(\frac{\exp -\frac{W_i^R}{k_B T}}{1 + \frac{n_F}{n_R}\exp -\frac{W_i^R + u}{k_B T}}\right)\right] \tag{4.11a}$$

$$z_F(u) = \log\left[\frac{1}{n_F}\sum_{i=1}^{n_F}\left(\frac{1}{1 + \frac{n_F}{n_R}\exp \frac{W_i^F - u}{k_B T}}\right)\right]. \tag{4.11b}$$

The value of u that satisfies the transcendental equation (4.10) corresponds to ΔG_{AB}.

4.2.2 *Jarzynski Free-Energy Estimator*

When information from both forward and reverse protocols is not available, it is possible to infer the ΔG_{AB} from unidirectional work measure-

ments. The Jarzynski free-energy estimator can be obtained from Eq.(4.8) by choosing $\phi(W) = 1$. In this case, for a given set of n forward (or reverse) work measurements $\{W_i\}_{i=1,\dots,n}$ one gets

$$\Delta G_{AB} = k_B T \log \left[\frac{1}{n} \sum_{i=1}^{n} \exp\left(-\frac{W_i}{k_B T} \right) \right]. \tag{4.12}$$

Notice that the exponential in Eq.(4.12) is dominated by the lowest work values, and it is crucial to have a good sampling of rare events to avoid the biased estimation of ΔG_{AB}. In practice, a large number of trajectories, n, is required to get a reliable estimation of the energy [85]. This implies that unidirectional estimators have a slower convergence to the "true" value of ΔG_{AB} than bidirectional ones [86].

5

UNZIPPING OF NUCLEIC ACIDS

The discovery of nucleic acids and their importance to heredity took more than a century. In 1865, Gregor Mendel's observations on the breeding of peas laid the foundations of genetics. In 1869, Friedrich Miescher found evidence of the presence of a new substance (i.e. the nucleic acids) in the cell's nucleus. In 1900, Hugo de Vries noticed that changes in the coloring of primroses are not gradual but occur abruptly as spontaneous mutations. Hermann Müller and L. G. Stadler discovered that the mutation rate could be increased by exposing the gamete cells to X-rays. In 1926, Thomas Morgan suggested that genes are not separated entities but grouped in the chromosomes.

In 1928, Frederick Griffith showed that the pneumonia bacterium, *Diplococcus pneumoniae*, is virulent when its polysaccharide coat is intact. Based on Griffith's result, in 1944, Oswald Avery, Colin MacLeod, and Maclin McCarthy carried out the following experiment: they inoculated mice with avirulent bacteria (i.e. no polysaccharide coating) and heat-killed bacteria. The experiment showed that removing the polysaccharide coat had no effect, as successive generations of bacteria had an intact coating and were virulent. However, if the bacterium's DNA is removed or denatured, its lethal potential is destroyed. Avery and his colleagues had established that DNA is the repository of the genetic message and that changes to an organism's observable traits (phenotype) do not affect

the information transmitted to the descendants (genotype). In 1953, Watson and Crick discovered the structure of DNA using X-ray diffraction. All these observations led to the discovery of the genetic code by Khorana, Holley, and Nirenberg, which was awarded the Nobel prize in 1968.

At that time, it was clear that DNA is the carrier of genetic information and that such information is transcripted from the DNA in an irreversible process. In 1970, Francis Crick wrote [87]: *"The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred back from protein to either protein or nucleic acid."*. Since then, molecular biology has undergone enormous progress. Although molecular activities such as reverse transcription, ribozymes, and post-translational modification significantly alter the linearity of the central dogma, its general meaning is still valid nowadays.

In this chapter, we introduce the biological structure of nucleic acids (DNA and RNA) and briefly discuss their main features and differences (Sec.5.1). In Sec.5.2, we discuss the model describing the energetics of NAs hybridization. In Sec.5.3 is described the theoretical modeling of NAs unzipping. Finally, in Sec.5.4 and 5.5 we discuss the theoretical modeling of the experimental FDC.

5.1 STRUCTURE OF NUCLEIC ACIDS

The fundamental building block of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) is the nucleotide (Fig.5.9). A nucleotide is made of a cyclic furanoside-type sugar (β-D-ribose in RNA and β-D-2′-deoxyribose in DNA) that is phosphorylated in the 5′ position and forms a glycosyl linkage in correspondence of the C_1' position with one of four different heterocycles. The heterocycles are divided into purines, i.e., Adenine (A) and Guanine (G), and pyrimidines, i.e., Cytosine (C), Thymine (T),

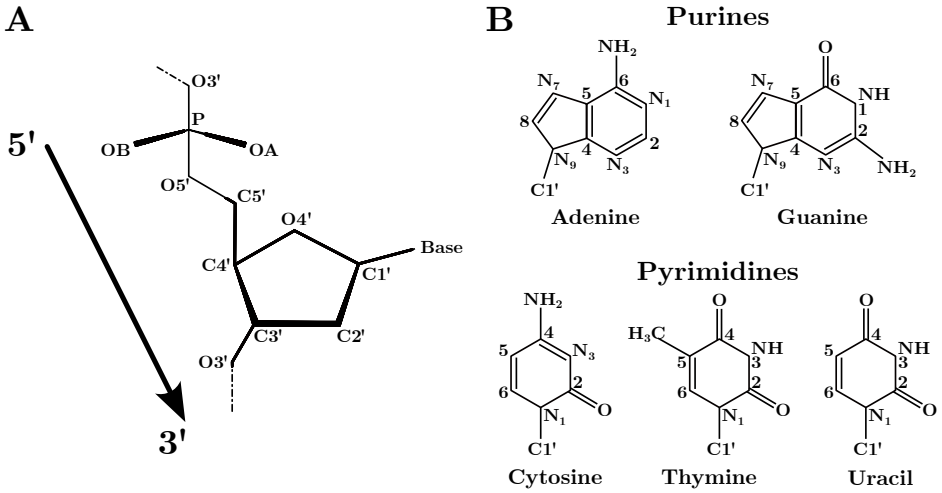


Figure 5.9: Structure of the nucleotides. (**A**) The basic structure of a nucleotide (see text). By convention, the directionality of the chain is given from the 5′ to the 3′ end. (**B**) The double-ring (purines) and single-ring (pyrimidines) nitrogenous bases make the DNA and RNA structure.

and Uracil (U). The latter is only present in the RNA structure and is replaced by thymine in DNA. Two nucleotides concatenate by forming a bond between the phosphoric group of the first nucleotide and the third carbon of the second nucleotide. The concatenation of multiple nucleotides forms a phosphate–deoxyribose backbone of linked bases featuring a phosphate group on one terminus (5′) and a hydroxyl group on the other one (3′). The resulting polynucleotide chain has asymmetric ends (5′ and 3′), and it is conventionally described given as a sequence of bases in the direction 5′ → 3′ (see Fig.5.9A).

The secondary structure of NA results from the hybridization of two complementary nucleotides sequences having antiparallel directions (i.e. a 5′ → 3′ strand pairs with a 3′ → 5′ strand) that are linked together by non-covalent interactions [88]. The canonical pairing between nucleotides is given by the Watson–Crick complementarity rules [89] and account for

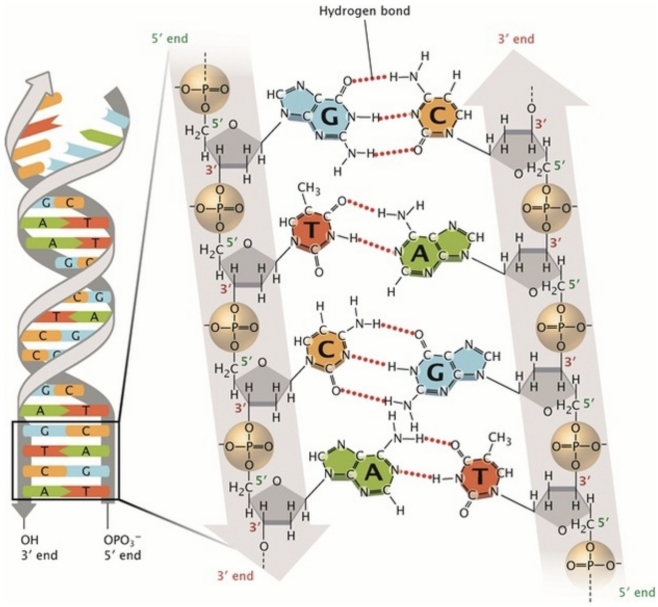


Figure 5.10: DNA secondary structure. Two complementary DNA strands with antiparallel directions hybridize according to the Watson–Crick base-pairing interactions. An analogous representation can be made for dsRNA by substituting T with U. Image from [90].

purine–pyrimidine bonding: A links to T (U in the RNA case) and G to C with three and two hydrogen bonds, respectively. The two strands bond through base-pairing, forming the characteristic NAs helical structure. Although hydrogen bonding is responsible for the specificity of the interaction of the bases, most of the stability of the structure is due to base stacking [91]. Stacking is an intermolecular interaction that tends to arrange molecules in a pile. Two forces stabilize base stacking: the hydrophobicity of the aromatic rings of the bases and the London dispersion of the dipoles (induced in the bases). Generally, the stacking interaction between purines (A and G) is stronger than the one between pyrimidines (C, T, and U).

The tertiary structure of a biomolecule represents its three-dimensional structure, i.e. the spatial localization of the atoms. In the case of NAs, the two strands form a double-helix (Fig.5.10). The backbones of the two single strands face each other and twist along the central axis of the molecule. The bases of one strand are paired with the complementary ones and are localized in the cavity left between the two backbones. The outer envelope of the double helix exhibits two helical grooves: the major and the minor groove having different widths and depths.

Compared to DNA, RNA exhibits more complex behavior. The replacement of deoxyribose for ribose and thymine for uracil makes RNA catalytic due to the reactive polarizable 2′–OH group of ribose. Ribose also induces significant changes at the level of base stacking interactions between contiguous bases. In their double-stranded forms, NAs form distinct right-handed double helices, B-form and A-form. Although DNA can adopt both A-form and B-form, RNA can only be found in A-form. The predominant form of dsDNA is the B-form, which consists of a right-handed double helix with a rise along the axis of 0.34nm/bp and a helix diameter of 2.0nm requiring 10 bases to complete a whole turn. In contrast, the dsRNA bases are tilted by approximately 19 degrees relative to the helical plane, and the interphosphate distance is smaller (0.28nm/bp) [92], $\sim$ 20% wider and shorter than the DNA B-form. These structural differences generate stacking between inter-strand bases and tighter water molecular bridges between phosphates and bases in RNA. Overall, base stacking tends to be stronger in RNA than in DNA. Base stacking is due to the Van der Waals attractive forces of the fluctuating dipole-dipole interactions between contiguous bases. Much weaker than the covalent nature of hydrogen bonding, the latter's effect is minimized upon secondary structure formation due to the compensation effect of hydrogen bonding with water. Overall, base stacking and hydrogen bonding contribute equally to RNA helix stabilization, albeit the $1/r^6$

dependence of Van der Waals forces makes stacking strongly sensitive to the inter-base distance, r. Therefore, RNA structure strongly depends on RNA stacking between intra-strand and inter-strand bases, making RNA folding prediction a difficult problem.

5.2 THE NEAREST-NEIGHBOR MODEL

Nucleic acids are polymeric chains of monomers (nucleotides) organized into increasingly complex structures. From the single-stranded form to the complex tertiary double-helix structure, the nature of such interactions defines the physico-chemical properties of a biomolecule. As described in the previous section, the formation of such structures is governed by the specific Watson–Crick base-pairing and the stacking between adjacent base pairs (see Fig.5.11A). The specific and non-specific interactions defining the duplex energetics can be described according to the Nearest-Neighbor (NN) model [22,93–95].

In the NN model, the base-pairing energy of two complementary bases only depends on the base itself and the first neighbor located in the same strand (in the $5' \rightarrow 3'$ direction). The nearest-neighbor base-pair (NNBP) are denoted as $XY/\overline{X}\,\overline{Y}$, where $X, Y = A, C, G, U$, and $\overline{X}(\overline{Y})$ is the complementary base of $X(Y)$ and $XY/\overline{X}\,\overline{Y}$ is the NNBP resulting from hybridizing dinucleotides $5' - XY - 3'$ and $5' - \overline{Y}\,\overline{X} - 3'$. The energies of $XY/\overline{X}\,\overline{Y}$ and $\overline{Y}\,\overline{X}/YX$ are equal due to complementary strand symmetry.

According to the NN model, the total hybridization free energy of a duplex, ΔG_0, is given by the sum over all the NNBP motifs along the sequence:

$$\Delta G_0(N) = \sum_{i=1}^{N} \Delta g_i \tag{5.13}$$

where Δg_i is the free energy of motif i. Notice that the NNBP energies are negative, as they are defined as the free-energy loss upon hybridizing

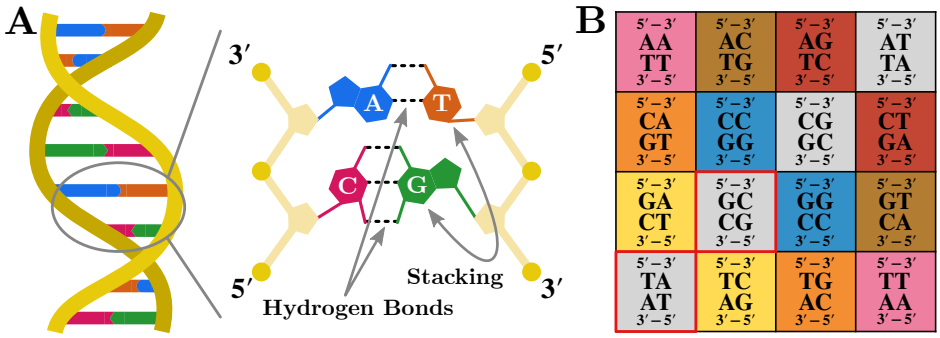


Figure 5.11: Nearest neighbour model. (**A**) The energetics of the (complementary) nucleotides forming the double-helix structure only depends on the base itself and the first neighbor in the same strand. (**B**) Matrix of the 16 NNBP motifs according to the NN model. Degenerate energies have the same cell color. Out of the 10 independent parameters, circular symmetry allows expressing two NNBP energies (TA/AT and GC/CG – red-bordered cells) as a linear combination of the others (see Eq.(5.14)). An analogous matrix can be written for the RNA case by changing T for U.

a base-pair, i.e., $\Delta g_i = g_i^H - g_i^O < 0$, where $g_i^H(g_i^O)$ is the free-energy of the hybridized (open) motif. There are 16 different motifs accounting for all possible combinations of adjacent NNBPs (see Fig.5.11B). This number is reduced from 16 to 10 by considering the degeneracy of the free energies due to the Watson–Crick complementarity. It is possible to further reduce this number from 10 to 8 independent parameters by considering the circular symmetry of the NN model [96, 97]. This symmetry yields additional self-consistent relations so that out of the 10 NNBP energies, 2 can be expressed as linear combinations of the remaining 8 [74, 97, 98]. The two motifs TA/AT (UA/AU) and GC/CG are those usually expressed as a function of the others:

$$\begin{aligned} \Delta g_{\mathrm{TA(UA)}} &= \Delta g_{\mathrm{CG}} + \frac{1}{2}\left(\Delta g_{\mathrm{AC}} + \Delta g_{\mathrm{GA}} - \Delta g_{\mathrm{AG}} - \Delta g_{\mathrm{CA}}\right) \\ \Delta g_{\mathrm{GC}} &= \Delta g_{\mathrm{AT(AU)}} + \frac{1}{2}\left(\Delta g_{\mathrm{GA}} + \Delta g_{\mathrm{CA}} - \Delta g_{\mathrm{AG}} - \Delta g_{\mathrm{AC}}\right) . \end{aligned} \tag{5.14}$$

For example, the RNA sequence $5' - \mathrm{CUUAGC} - 3'$ forms a duplex with its complementary strand, $5' - \mathrm{GCUAAG} - 3'$. The total energy of this sequence is $\Delta g_{\mathrm{CU/GA}} + \Delta g_{\mathrm{UU/AA}} + \Delta g_{\mathrm{UA/AU}} + \Delta g_{\mathrm{AG/UC}} + \Delta g_{\mathrm{GC/CG}}$ with $\Delta g_{\mathrm{CU/GA}} = \Delta g_{\mathrm{AG/UC}}$ due to complementary strand symmetry.

The accurate measure of the NNBP free energies is key for correctly estimating the total free energy of formation of the duplex native state. The 10 independent parameters have been extracted from melting experiments of short duplexes of varying sequences and lengths [23, 99–101] for both DNA and RNA and are accessible in the Mfold server [24]. Single-molecule techniques allow for much more accurate free-energy measurements than bulk experiments. However, to derive the 10 (8 if circular symmetry is considered) NNBP parameters from unzipping experiments, it is fundamental to have a theoretical model of the unzipping process to predict the experimental FDC.

5.3 MODELING UNZIPPING EXPERIMENTS

In unzipping experiments at a controlled position (see Sec.3.3), the trap–pipette distance, λ, is steadily increased (unzipping) or decreased (rezipping) by moving the optical trap. As the trap-pipette distance changes, the increasing (decreasing) force, f, exerted on the molecule causes the number of open bases, n, to change and the hairpin to unfold (refold). As shown in Fig.5.12, the total distance, λ, can be written as

$$\lambda(f) = \begin{cases} x_{\mathrm{b}}(f) + x_{\mathrm{h}}(f) + x_{\mathrm{d}}(f) + \lambda_0 \,(\text{unzipping/rezipping}) \\ x_{\mathrm{b}}(f) + x_{\mathrm{h}}(f) + x_{\mathrm{ss}}(f) + \lambda_0 \;\; (\text{unfolded state})\,, \end{cases} \tag{5.15}$$

depending on whether the stem of the molecule is formed during unzipping–rezipping (Fig.5.12, left) or completely unfolded (Fig.5.12, right). The total distance is a sum of various contributions [102]: $x_{\mathrm{b}}(f)$

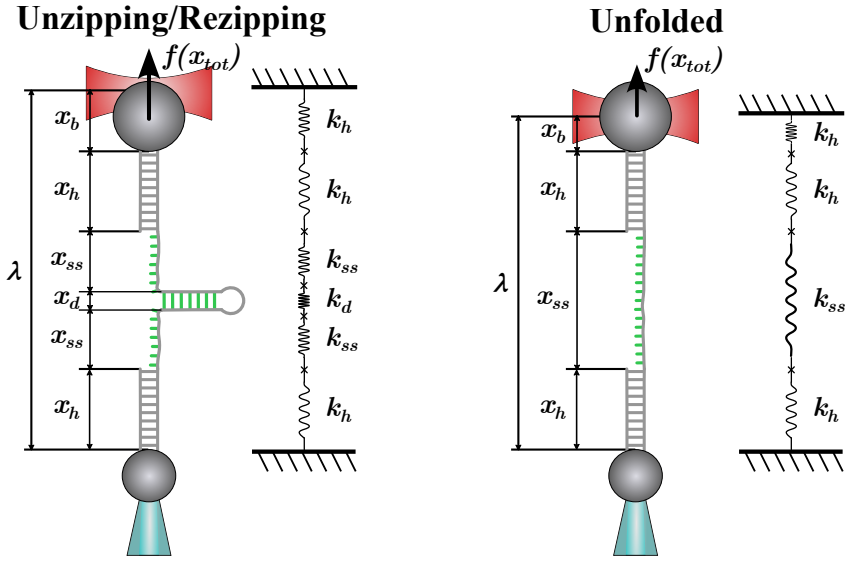


Figure 5.12: Unzipping experiment at controlled position. By moving the optical trap, a force is applied to the molecular construct (see text). The trap–pipette distance, x_{tot}, is the result of several contributions (left): the distance of the bead from the trap center (x_b), the extension of the handles (x_h), the extension of the double-stranded molecule along the force axis (x_d), and the extension of the single-stranded molecule (x_{ss}). Notice that when the molecule is completely unfolded (right), $x_d = 0$.

is the displacement of the bead from the center of the optical trap, $x_{\mathrm{h}}(f) = x_{\mathrm{h}_1}(f) + x_{\mathrm{h}_2}(f)$ accounts for the sum of the elongations of the two double-stranded handles, $x_{\mathrm{ss}}(f)$ is the end-to-end extension of the single-stranded unfolded molecule, and $x_{\mathrm{d}}(f)$ is the average extension of the folded hairpin. This last term is defined as the extension of the NA double-helix projected along the force axis [103] and is usually called "hairpin diameter", d (typically $d = 2$nm for DNA and RNA hairpins [104]). Notice that the λ in Eq.(5.15) is a relative quantity so that it is defined up to an additive constant, λ_0.

The total distance, λ, can be described as the extension of a series of different springs, each with a different elastic constant, k (see Fig.5.12).

Upon changing λ, the extension variation of each term in the r.h.s of Eq.(5.15) depends on their physico-chemical properties. In general, a small variation $\delta\lambda$ produces a small change in the applied force δf. The extent of this variation is the *effective stiffness* of the system $k_{\text{eff}} = \delta f / \delta\lambda$ and it equals the slope of the experimental FDC. Therefore, according to Eq.(5.15), the inverse effective stiffness can be written as

$$\begin{aligned} \frac{1}{k^{\text{F}}_{\text{eff}}(f)} &= \frac{1}{k_{\text{b}}(f)} + \frac{1}{k_{\text{h}}(f)} + \frac{1}{k_{\text{d}}(f)} \,(\text{unzipping/rezipping}) \\ \frac{1}{k^{\text{U}}_{\text{eff}}(f)} &= \frac{1}{k_{\text{b}}(f)} + \frac{1}{k_{\text{h}}(f)} + \frac{1}{k_{\text{ss}}(f)} \,(\text{unfolded state})\,, \end{aligned} \tag{5.16}$$

where $k_{\text{b}}(f)$ is the stiffness of the bead in the optical trap, $k_{\text{h}}(f)$ is the sum of the two handles' stiffness, $k_{\text{d}}(f)$ corresponds the molecular stiffness of the folded molecule and $k_{\text{ss}}(f)$ stands for the stiffness of the single-stranded molecule. Let us study in detail each one of these components.

The bead in the optical trap is modeled as a Hookean spring of stiffness k_{b}:

$$f(x_{\text{b}}) = k_{\text{b}} x_{\text{b}}\,. \tag{5.17}$$

The extension upon orienting the double helix is modeled as a dipole of length equal to the helix diameter, d, that aligns along the force axis:

$$x_{\text{d}}(f) = d\left[\coth\left(\frac{fd}{k_{\text{B}}T}\right) - \frac{k_{\text{B}}T}{fd}\right], \tag{5.18}$$

where k_{B} is the Boltzmann constant and T is the system temperature.

Finally, the extension of double-stranded handles, x_{h}, and of the single-stranded molecule, x_{ss}, are modeled as elastic polymers. These terms are

typically described by the inextensible or extensible Worm-Like Chain (WLC) model and its interpolation formula [105]. In the inextensible case

$$f(x) = \frac{k_B T}{4 l_p} \left[\left(1 - \frac{x}{n l_d} \right)^{-2} - 1 + 4 \frac{x}{n l_d} \right], \tag{5.19}$$

where x is the average extension of the molecule ($x = x_{ss}$ for the unfolded hairpin, $x = x_h$ for the double-stranded handles). l_p is the persistence length, i.e. the typical distance along the polymer backbone over which there is an appreciable bending due to thermal fluctuations, and l_d is the interphosphate distance, i.e. the typical distance between consecutive bases (base-pairs in the double-stranded case) along the NA backbone. Eq.(5.19) is usually written as a function of the contour length, $L_c = n l_d$, i.e. the end-to-end distance of a fully straightened polymer of n bases (base-pairs). Notice that computing $x_h(f, n)$ and $x_{ss}(f, n)$ requires inverting Eq.(5.19) [102], which is not an easy task and the solution depends on the system parameters. The full computation is reported in Appendix A. By recalling that in general $k = \partial f / \partial x$, the stiffness of the polymer can be obtained by differentiation of Eq.(5.19), so that

$$k(x) = \frac{\partial f(x)}{\partial x} = \frac{k_B T}{2 L_c P} \left[\left(1 - \frac{x}{L_c} \right)^{-3} + 2 \right]. \tag{5.20}$$

Given Eq.(5.19), it is also possible to take into account the elastic deformation of the stretched polymer by performing the substitution $L_c \to L_c(1 + f/Y)$, with Y the Young modulus of the stretchable polymer [106, 107], i.e. the resistance to deformation of the system to applied uniaxial stress. The contour length becomes force-dependent in this case, and the corresponding model is called the *extensible* WLC. By contrast, Eq.(5.19), where L_c is constant, is known as the *inextensible* WLC. The latter has been shown to describe the elastic properties of single-stranded

nucleic acids (ssDNA and ssRNA) with good accuracy [59] while the former has for long been the standard to model the elastic properties of double-stranded nucleic acids in the entropic regime.

The persistence length l_p is a measure of the mechanical stiffness of the polymer being strongly sensitive to environmental conditions (e.g. ionic strength, temperature, solvation, etc.). Polymers with $l_p \gg L_c$ effectively behave as rigid rods, whereas if $l_p \leq L_c$ polymers are bent at the scale of the contour length by thermal forces. It is important to mention that l_p does not only depend on the ionic concentration and temperature [69] (as predicted by polyelectrolyte theories) but also on experimental parameters such as contour length [59], suggesting that l_p is an effective parameter. For example, at 1M NaCl, recent single-molecule studies have shown that, for short (a few tens bases) ssDNA molecules, $l_p = 1.35$nm [108] whereas for long ssDNA $\sim$ 13kbp $l_p = 0.76$nm [109]. On the other hand, for short ssRNA molecules $l_p = 0.75$nm [110] and for long $\sim$ 1kbp ssRNAs $l_p = 0.83$nm [59, 60]. These values are significantly lower than for double-stranded nucleic acids (dsDNA and dsRNA) where $l_p = 50$nm for dsDNA [13] and $l_p \simeq 60$nm for dsRNA molecules [111].

5.4 COMPUTATION OF THE SYSTEM FREE-ENERGY

Single-molecule unzipping experiments are reversible processes in which a molecular construct (NA, protein, etc.) is repeatedly pulled between two fixed positions by moving the optical trap. As discussed in Chapter 4, the amount of energy needed to stretch the molecular construct depends on the initial and final positions of the system: at the initial state, i.e. at position λ_0, the molecule is subject to a (minimum) force $f_0 = f_{\min}$ while at the final state (at λ_1) the corresponding applied force is $f_1 = f_{\max}$.

In general, given the model of the experimental setup discussed in Sec.5.3, the total system free energy can be written as

$$\Delta G = \Delta G_0 + \Delta G_{\mathrm{m}} + \Delta G_{\mathrm{b}} + \Delta G_{\mathrm{h}}\,, \tag{5.21}$$

where ΔG_0 is the free energy of formation of the molecule at zero force, i.e. the free energy difference between the folded state (completely hybridized hairpin) and (unfolded) random coil conformations and ΔG_i ($i = \mathrm{m}, \mathrm{b}, \mathrm{h}$) is the reversible work difference between the initial state (at position λ_0 where the hairpin is subject to a minimum force $f_0 = f_{\mathrm{min}}$) and final state (at position λ_1 and force f_1) of each setup element (molecule optically trapped bead and handles, respectively)

5.4.1 *Molecular Stretching Contribution*

The molecular contribution $\Delta G_{\mathrm{m}}(x)$ in Eq.(5.21) accounts for the reversible work needed to stretch the molecular construct, i.e. ssNA and dsNA,

$$\Delta G_{\mathrm{m}}(x) = \int_0^{x_{\mathrm{ss}}(f_1)} f_{\mathrm{ss}}(x)\,dx - \int_0^{x_{\mathrm{d}}(f_1)} f_{\mathrm{d}}(x)\,dx\,, \tag{5.22}$$

where $f_{\mathrm{ss}}(x)$ and $f_{\mathrm{d}}(x)$ are the force applied to the stretched ssNA (Eq.(5.19)) and dipole (Eq.(5.18)), respectively. The two terms on the r.h.s of Eq.(5.22) respectively account for the reversible work to stretch the single-stranded hairpin and orientate the molecular diameter along the force axis at $f = 0$ up to f_1. However, in unzipping experiments at a controlled position, the force is the independent variable, so the calculation of the molecular contributions requires computing the integral

$$\Delta G_{\mathrm{m}}(x) = \int_0^{x_{\mathrm{m}}(f_1)} f_{\mathrm{m}}(x)dx = f_1\, x_{\mathrm{m}}(f_1) - \int_0^{f_1} x_{\mathrm{m}}(f')df'\,, \tag{5.23}$$

where $x_m = x_{ss}, x_d$ given by Eqs.(5.19) and (5.18), respectively. Notice that the second term in the r.h.s of Eq.(5.22) is zero when the molecule is completely unfolded.

5.4.2 *Bead and Handles Contributions*

The two terms $\Delta G_b(x) + \Delta G_h(x)$ in Eq.(5.21) correspond to the sum of the reversible work required to displace the bead from the center of the optical trap (ΔG_b) and the reversible work needed to stretch the double-stranded handles (ΔG_h). The trap contribution can be easily computed by integrating Eq.(5.17), which gives

$$\Delta G_b(x) = \frac{1}{2} k x_b^2 \,. \tag{5.24}$$

The dsNA term is modeled according to the WLC (see Sec.5.3). Therefore, the handles contribution, ΔG_h, is computed by integrating Eq.(5.19) according to (5.23). Notice that this term needs to be counted twice, as two double-stranded handles are used in the experimental setup.

5.5 PREDICTION OF THE UNZIPPING CURVE

The model of the NA elastic response described in Sec.5.3 and 5.4 allows for the computation of the equilibrium force applied to the molecular construct at each instant of the unzipping protocol. This ultimately gives a theoretical prediction of the equilibrium FDC of a hairpin sequence.

At each position of the optical trap, λ, the total system free energy is given by Eq.(5.21) so that

$$\begin{aligned} \Delta G_{tot}(x_{tot}, n) = & \Delta G_0(n) + \Delta G_b(x_b) + \Delta G_h(x_h) + \\ & + \Delta G_{ss}(x_{ss}, n) + \Delta G_d(x_d) \,, \end{aligned} \tag{5.25}$$

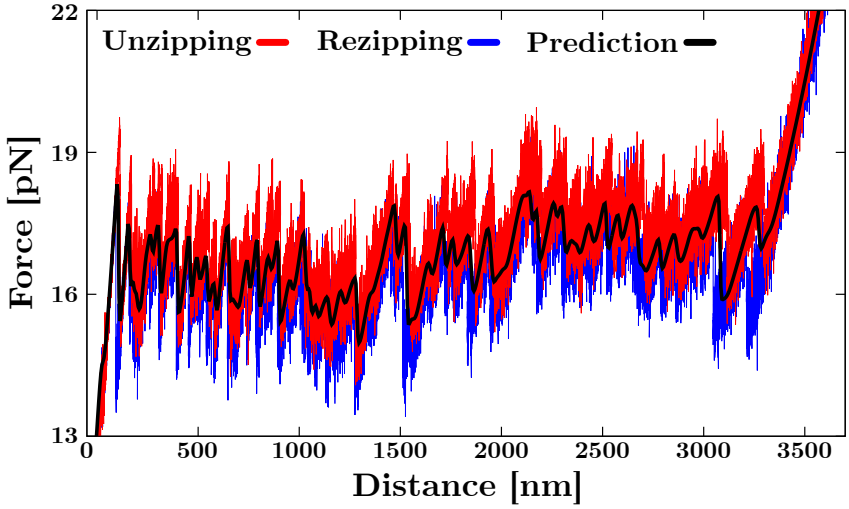


Figure 5.13: Prediction of the theoretical FDC of a 3.6kbp DNA hairpin at 1M NaCl, 25°C (black) compared to experimental unzipping and rezipping data (red and blue curves, respectively).

where $x_{\text{tot}} \equiv \lambda$ is the total system extension in Eq.(5.15) and n is the number of open base-pairs.

5.5.1 *Computation of the Equilibrium FDC*

Let us consider the case where thermal fluctuations are not accounted for in the FDC computation. Thus, at a given value of x_{tot}, the system is always in the state of minimum energy, $\Delta G_{\text{eq}}(x_{\text{tot}}) = \Delta G_{\text{tot}}(x_{\text{tot}}, n^*)$.

To compute the equilibrium free energy of the system, let us first introduce the system partition function, Z. At each x_{tot}, this is defined as the sum over all the possible states, i.e., all the possible sequences of n open base pairs, which is

$$Z(x_{\text{tot}}) = \sum_{n=0}^{N} \exp\left(-\frac{\Delta G_{\text{tot}}(x_{\text{tot}}, n)}{k_{\text{B}}T}\right), \tag{5.26}$$

where N is the total number of base pairs of the sequence. Finally, by recalling that $\Delta G = -k_B T \ln Z$, the equilibrium force is given by:

$$f_{eq}(x_{tot}) \equiv \frac{\partial \Delta G(x_{eq})}{\partial x_{tot}} = -k_B T \frac{\partial \ln Z(x_{tot})}{\partial x_{tot}}. \tag{5.27}$$

Computing Eq.(5.26) requires solving the transcendental equation (5.15) (that can be performed numerically) with respect to f and then computing Eq.(5.21) for all $n \in [0, N]$. The value n^* minimizing the equilibrium free-energy $\Delta G_{eq} = \Delta G_{tot}(x_{tot}, n^*(x_{tot}))$ gives the most probable number of open base-pairs at a given x_{tot}. Eventually, the computation of the equilibrium force in Eq.(5.27) gives a theoretical prediction for the unzipping curve of a given sequence (see Fig.5.13).

5.5.2 *Equilibrium Free Energy*

The free energy in Eq.(5.25) is the sum of two terms: the hybridization energy, which depends only on n, and the elastic contribution that depends on both n and x_{tot}. For a hairpin of N bases, n is a discrete variable ranging from 0 (native state) to $N-1$ (totally unfolded), which gives a set of $N-1$ possible functions for each value of x_{tot}. By increasing x_{tot}, the system minimizes the total free energy by undergoing a cooperative opening of base pairs. Let us suppose that the system starts with n_1 open base pairs. When x_{tot} is increased, the elastic term in Eq.(5.25) also increases until $\Delta G_{tot}(x_{tot}, n_1) \equiv \Delta G_{tot}(x_{tot}, n_2)$, where $n_2 > n_1$ (Fig.5.14A, top). Even though the total energy of these two states is the same, the energetic internal balance is different (Fig.5.14A, bottom). If x_{tot} keeps increasing, the state n_2 becomes more stable: the system minimizes the elastic free energy and switches to state n_2 by releasing $\Delta n = n_2 - n_1$ base-pairs. Notice that, despite opening Δn bp increases the system's energy, the released ssNA lowers the elastic contribution. In general,

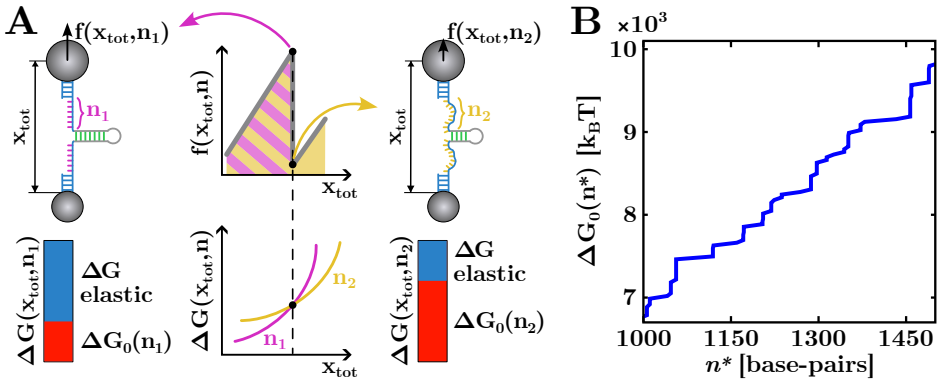


Figure 5.14: (**A**) Equilibrium of energies in a force rip. The molecule has n_1 open bp before the force rip (left) and n_2 open bp after the rip (right). At the force rip (black dots), the total free energy of the system is the same for both states. Still, the elastic and the hybridization contributions are different: the system switches to the configuration that minimizes the elastic energy and releases $\Delta n = n_2 - n_1$ base pairs. (**B**) Equilibrium free-energy of formation, $\Delta G_0(n^*)$, as a function of the number of open base pairs, n^* (only a subregion of the total function has been shown to point out the profile pattern). Each jump is due to the release of Δn bases that convert from dsNA (hybridized) to ssNA.

$\Delta G_{\text{el}} \gg \Delta G_0$ so the global balance of the state n_2 is lower than the one of n_1. Therefore, the equilibrium free energy is a step function increasing with x_{tot} (Fig.5.14B). At each discontinuity, several base pairs open, and the system transitions to an equilibrium state of n^* open bases. At each discontinuity of the equilibrium, energy corresponds to a drop in the system's elastic energy and then a rip along the equilibrium FDC.

6

TRANSITION STATE THEORY AND FREE ENERGY LANDSCAPE

Biological processes, such as molecular unzipping/rezipping and protein-protein or protein-DNA interaction, are chemical reactions in which some reactants (R) convert into products (P). Accurate knowledge of the relation between the reaction's kinetic rates and the system's macroscopic behavior is key to interpreting the experimental observation.

Given a chemical reaction $R \rightleftarrows P$, when the conversion from reactants to product (and vice-versa) is in equilibrium, we can define the equilibrium concentration of products and reactants, $K_{\text{eq}} = [P]_{\text{eq}}/[R]_{\text{eq}}$. In 1884, van't Hoff proposed the famous relation to describe the temperature dependency of equilibrium concentration [112],

$$\frac{d \log K_{\text{eq}}}{dT} = \frac{\Delta H}{k_{\text{B}} T^2}\,, \tag{6.28}$$

where ΔH is the system enthalpy change over converting one reactant molecule into one of the products. A few years later, Arrhenius considered that the reactants must acquire a minimum energy [113], called *activation energy*, to transform into products successfully. The reactants in the active form R^* must be in equilibrium with the reactants according to $R \overset{K_{\text{eq}}}{\rightleftarrows} R^* \rightarrow P$. Therefore, the kinetic rate constant, $k_{\rightarrow}$, of the chemical

reaction $R \rightleftarrows P$ is proportional to the normal-to-active equilibrium constant, $k_{\rightarrow} \propto K_{\mathrm{eq}}$. Based on van't Hoff's work, in 1889, Arrhenius proposed the following equation:

$$\frac{d \log k_{\rightarrow}}{dT} = \frac{\Delta E^{\ddagger}}{k_{\mathrm{B}} T^2}\,, \tag{6.29}$$

where $\Delta E^{\ddagger}$ is the temperature-independent energy change between an activated (R^*) and a normal reactant molecule (R). By solving Eq.(6.29), one gets

$$k_{\rightarrow} = k_0 \exp\left(-\frac{\Delta E^{\ddagger}}{k_{\mathrm{B}} T}\right) \tag{6.30}$$

where k_0 is a constant of integration often called *attempt rate*.

Arrhenius' work laid the basis for the modern transition state (TS) theory. Nowadays, the TS theory methods are widely used to study reactions in equilibrium conditions. Given their high temporal (ms) and spatial resolution (nm), SM pulling experiments (see Sec.3.3) provide an excellent playground to apply these methods to biological processes.

6.1 THE BELL-EVANS MODEL

To model unzipping experiments with the TS theory, the effect of force must be included to model the kinetic rates. In the Bell-Evans (BE) model [114, 115], the unfolding/refolding reaction is described as a thermally activated process in which the system crosses a TS energy barrier. For a two-state system, the molecular free-energy landscape (FEL) is made of a single kinetic barrier linearly decreasing with force as $B = B_0 - f x^{\ddagger}$. The term B_0 stands for the height of the kinetic barrier at zero force, and $x^{\ddagger}$ is the distance between the initial state and the TS. From Eq.(6.30), the

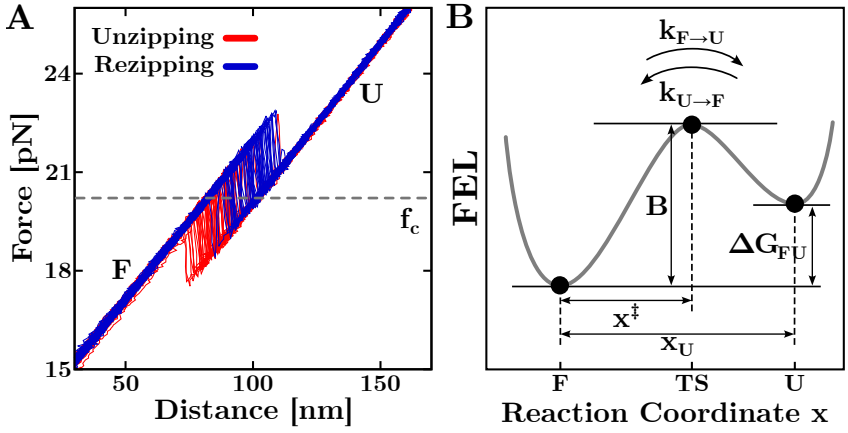


Figure 6.15: Transition state theory (BE model) applied to a two states system. (**A**) Unzipping/rezipping FDCs by pulling experiments with LOT of the RNA hairpin CD4L4 at 4mM $MgCl_2$. The unzipping is a two states process: upon pulling, the molecule switches between the folded (F) and the unfolded (U) state. The coexistence force between unzipping (red) and rezipping (blue) FDCs is $f_c \approx 20$pN (dashed grey line). (**B**) Molecular FEL according to the BE model (see text). The system must overcome the energy barrier B at a distance $x^\ddagger$ to switch from U to F (and vice-versa).

kinetic rates of an unzipping process starting in the folded (F) state and ending in the unfolded (U) state (and vice-versa), can be written as

$$k_{F\to U} = k_m \exp\left(\frac{f x^\ddagger}{k_B T}\right) \tag{6.31a}$$

$$k_{U\to F} = k_m \exp\left(\frac{\Delta G_{FU} - f(x^\ddagger - x_U)}{k_B T}\right), \tag{6.31b}$$

where $k_m = k_0 \exp(-B_0/k_B T)$, being k_0 the pre-exponential factor, $x^\ddagger$ ($x^\ddagger - x_U$) is the relative distance between state F (U) and the TS at coexistence force f_c, and $\Delta G_{FU} = f_c x_{FU}$ is the free-energy difference between states F and U (see Fig.6.15).

According to Eqs.(6.31a) and (6.31b), the logarithm of the unfolding and folding kinetic rates are linear with force. However, the experimental

observations deviate from the expected behavior, exhibiting curvature over large force intervals [116, 117]. It has been shown [118] that the description of the kinetic rates as given by the BE model is only valid in the vicinity of the coexistence force f_c, where the following relation holds:

$$\log\left(\frac{k_{F\to U}(f)}{k_{U\to F}(f)}\right) = (f - f_c)\frac{x_{FU}}{k_B T}\,, \tag{6.32}$$

since $\Delta G_{FU} = f_c \Delta x_{FU}$ to a first order.

6.1.1 *Folding and Unfolding Probability*

Let us consider a two-states molecular system pulled at a force linearly increasing with time as $f = rt$, being r the experimental pulling rate. The probability, $P(f)$, of the system remaining in the folded state (F) is given by

$$\frac{dP(f)}{df} = -\frac{k_{F\to U}}{r}P(f)\,. \tag{6.33}$$

By substituting Eq.(6.31a) into Eq.(6.33), one gets

$$\log\left(-r\log P(f)\right) = \log\left(\frac{k_m k_B T}{r x^\ddagger}\right) + \log\left[\exp\left(\frac{f x^\ddagger}{k_B T}\right) - 1\right], \tag{6.34}$$

where $k_m = k_0 \exp(-B_0/k_B T)$ and $x^\ddagger$ is the distance to the transition state. By computing the derivative with respect to f of the latter expression, one gets the average rupture force and the rupture force standard deviation [118]:

$$\langle f_{\text{rupt}}\rangle = \frac{k_m k_B T}{r x^\ddagger}\log\left(\frac{r x^\ddagger}{k_m k_B T}\right), \tag{6.35}$$

and

$$\sigma_f = \log\left(\frac{3+\sqrt{5}}{2}\right)\frac{k_B T}{r x^\ddagger}\,. \tag{6.36}$$

An analogous computation can be done for the reverse process, i.e., the probability of the system remaining in the unfolded state (U). In this case, the results are identical but k_m has to be replaced by $k_0 \exp\left(\Delta G_{FU}/k_B T\right)$ and $x^\ddagger$ by $x_U - x^\ddagger$.

6.2 CONTINUOUS EFFECTIVE BARRIER ANALYSIS

In the BE model, the height of the kinetic barrier is assumed to decrease linearly with the applied force ($B(f) = B_0 - fx^\ddagger$). This hypothesis is relaxed in the kinetic diffusion (KD) model, in which the folding reaction is assumed to be a diffusive process in a one-dimensional force-dependent FEL. The Continuous Effective Barrier Approach (CEBA) is based on the KD model. It can be used to extract the force-dependent behavior of the kinetic barrier from unzipping experiments [116, 117, 119]. In CEBA, the effective barrier between the native (N) and the unfolded state (U), $B(f)$, is derived by imposing the detailed balance between the unfolding ($k_{FU}(f)$) and folding ($k_{FU}(f)$) kinetic rates (see Eqs.(6.31)):

$$k_{F\to U}(f) = k_0 \exp\left(-\frac{B(f)}{k_B T}\right) \tag{6.37a}$$

$$k_{U\to F}(f) = k_{F\to U}(f) \exp\left(\frac{\Delta G_{FU}(f)}{k_B T}\right), \tag{6.37b}$$

where k_0 is the attempt rate, $B(f)$ is the effective barrier at force f, is and $\Delta G_{FU}(f)$ is the folding free energy at force f. The latter term is given by

$$\Delta G_{FU}(f) = \Delta G_0 - \int_0^f \left(x_U(f') - x_F(f')\right) df', \tag{6.38}$$

where ΔG_0 is the folding free energy between F and U at zero force, and the integral accounts for the free energy change upon stretching the molecule in state U (F) at force f.

We can derive two estimates for $B(f)$ by computing the logarithms of Eqs.(6.37a) and (6.37b), which give

$$\frac{B(f)}{k_B T} = \log k_0 - \log k_{F\to U}(f) \tag{6.39a}$$

$$\frac{B(f)}{k_B T} = \log k_0 - \log k_{U\to F}(f) + \frac{\Delta G_{FU}(f)}{k_B T}\,. \tag{6.39b}$$

By imposing the continuity of two estimations of $B(f)$ in Eqs.(6.39), we can measure the folding free energy at force f, $\Delta G_{FU}(f)$. As the free energy stretching contribution to the latter term (see Eq.(6.38)) can be measured from the FDC unfolded branch, we ultimately get a direct estimate of the folding free energy at zero force, ΔG_0.

Part II

SINGLE-MOLECULE CALORIMETRY OF LONG DNA HAIRPINS

7

DNA THERMODYNAMICS FROM UNZIPPING EXPERIMENTS

The duplex stability is the result of the compensation of two quantities, a favorable enthalpy (ΔH) and an unfavorable entropy (ΔS), related by the thermodynamics equation $\Delta G = \Delta H - T\Delta S$, being T the environment temperature. However, molecular stability is not only determined by the direct interactions between the nucleotides forming the duplex backbone. For example, in proteins, an important contribution is given by the hydrophobicity of buried amino acid side chains exposed to the aqueous environment. These effects result in a significant variation of the system's heat capacity (ΔC_p) [120], yielding a strong temperature dependence of enthalpy and entropy.

For a long time, the effects of temperature dependence on enthalpy and entropy on DNA stability have been considered negligible, and the process of DNA melting, i.e. the separation by heating of the dsDNA into two single strands, was assumed as temperature independent. This assumption mainly relied on the fact that, when the first experiments were carried out, the scanning calorimeters could not detect ΔC_p. Over time, the development of the experimental techniques gave access to instruments of increasing precision and accuracy, which showed the significant role played by ΔC_p for DNA hybridization.

During the past 20 years, several studies have been carried out both in bulk assays [121–125] and in SM pulling experiments [69, 126] to assess the effects of temperature change on DNA hybridization and measure the specific heat change, ΔC_p. This yielded a multitude of different results, with ΔC_p values spread over several orders of magnitude (between $\approx 10^0 - 10^2\,\mathrm{cal\,mol^{-1}K^{-1}}$ per bp) depending on the experimental technique and DNA sequence. To our knowledge, a direct measurement of the temperature effects on DNA hybridization at the NNBPs level is still lacking.

The 10 NNBP DNA parameters have been measured from melting experiments of short DNA duplexes by many laboratories worldwide [95, 127–132]. These results have been unified by Santalucia *et al.* [22] in the so-called *Unified Oligonucleotide* (UO) set. More recently, the 10 (8 if circular symmetry is applied) NNBP DNA energies have been derived from SM pulling experiments of long DNA hairpins with LOT in different monovalent and divalent salt conditions [74, 133]. However, all these experiments have been carried out at a fixed temperature, and the measured NNBP DNA entropies and enthalpies were considered temperature-independent ($\Delta C_p \approx 0\,\mathrm{cal\,mol^{-1}K^{-1}}$).

Here, we derive the 8 independent NNBP DNA free energies, entropies, enthalpies, and specific heat changes by unzipping a 3.6kbp DNA hairpin with short handles (29bp). The experiments were carried out in a buffer of 1M NaCl, 10mM Tris-HCl (pH 7.5), and 1mM EDTA in the temperature range between 7°C and 42°C. The experiments have been carried out using the temperature-jump LOT setup described in Sec.3.2.2. At each temperature, we pulled $5-6$ different molecules for a total of $40-50$ unzipping/rezipping complete cycles. In Fig.7.1, we show the FDCs averaged over all the trajectories obtained at each temperature.

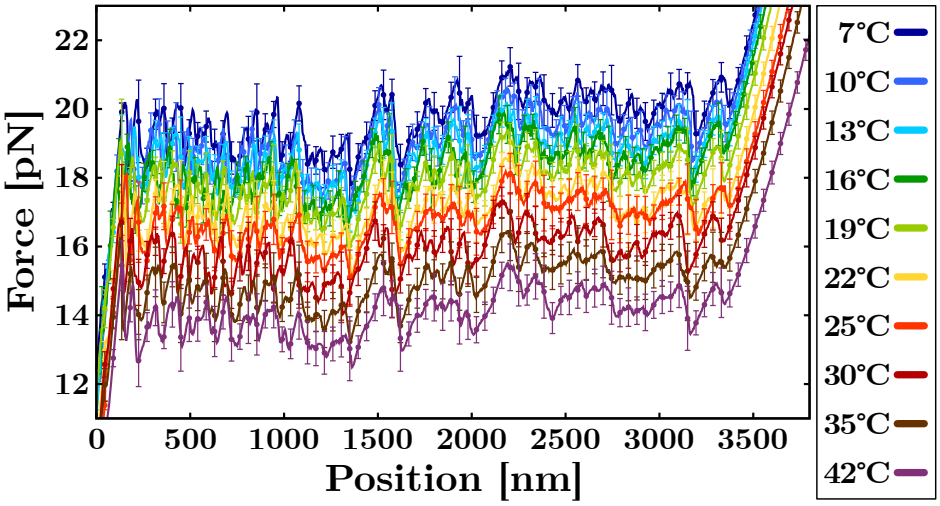


Figure 7.1: Experimental FDCs measured by pulling a 3.6kbp DNA hairpin at 1 M NaCl in the temperature range $[7, 42]^{\circ}$C. At each temperature, the FDC results from averaging over all the measured molecules. The error bars (plotted for a fraction of the total data points) show the molecule-to-molecule variability.

7.1 TEMPERATURE DEPENDENCE OF THE DNA FDCS

The energetics and the mechanical properties of DNAs are strongly dependent on the environment's temperature. In Fig.7.1, we show the FDCs at different temperatures obtained by unzipping the same 3.6kbp DNA hairpin. Let us define the average unzipping force, f_{m}, as the average force (evaluated between the first "peak" and the last "valley" of the FDC) at which the hairpin unzips/rezips (red line in Fig.7.2). The temperature change affects the stabilization of the NN base pairing and stacking interactions: the lower the temperature, the more stable the DNA double-helix. As a result, unzipping occurs at a higher f_{m} upon lowering T (see Fig.7.3A and Table 7.1). A linear fit to data gives the f_{m} temperature dependence, $\partial f_{\mathrm{m}}/\partial T = -0.165 \pm 0.03$pN/K. Interestingly, while f_{m} changes with the temperature, the FDC sawtooth pattern, i.e.

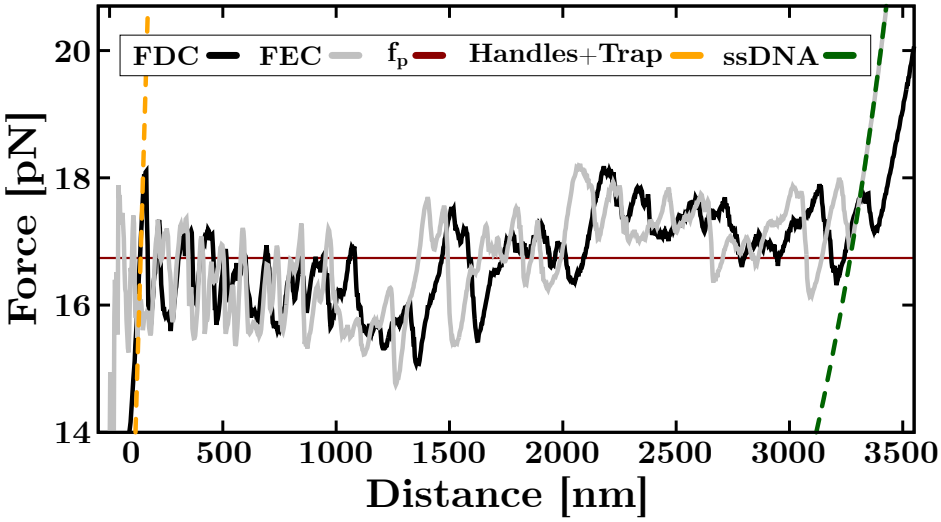


Figure 7.2: Unzipping FDC (black line) and FEC (grey line) at $T = 25°C$. The FEC is computed by subtracting the measured bead and trap elastic contributions (orange dashed line) to the FDC (see text). The fit of the FEC unfolded branch (green dashed line) allows for measuring the ssDNA elastic properties at temperature T. Notice that the average unzipping force, f_m, (red line) remains unchanged upon computing the FEC.

the succession of slopes and rips, remains the same over the whole temperature range.

At the same time, the hairpin extension increases upon increasing T. To assess the effects of temperature on the ssDNA elastic properties, we first need to compute the hairpin extension from the measured FDCs. The FDC is a function of the total trap-pipette distance, which includes contributions from the optical trap and the dsDNA handles (see Sec.5.3). These two terms can be evaluated by using the *effective stiffness* method [102] (see Eqs.(5.16)). According to it, the stretching contributions of the experimental setup can be approximated by the effective stiffness, $k_{\text{eff}}^{-1} \approx k_{\text{h}}^{-1} + k_{\text{b}}^{-1}$, where k_{h} and k_{b} are the stiffness of the dsDNA handles and the optically trapped bead, respectively. The use of short handles

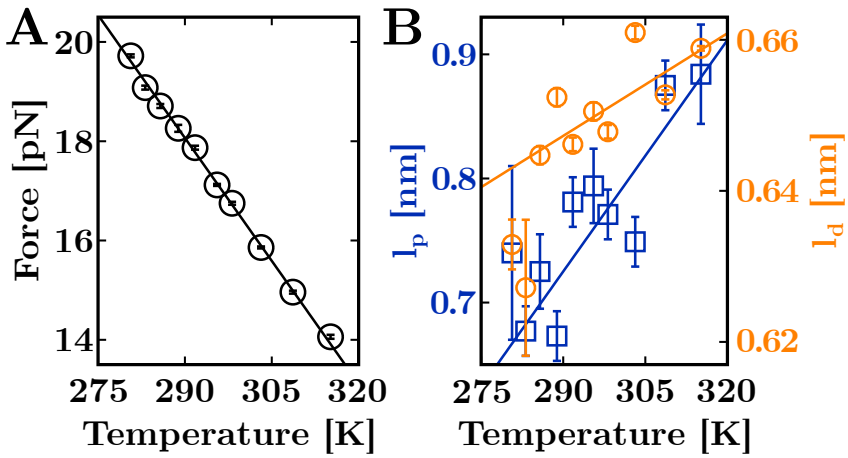


Figure 7.3: Temperature dependence of the DNA mechanical properties measured by unzipping the 3.6kbp DNA hairpin in the temperature range [280, 315] K ([7, 42]°C). (**A**) Change of the average unzipping force, f_m, measured from the experimental FDCs at each temperature. A fit to data (solid line) gives $\partial f_m/\partial T = -0.165 \pm 0.03$pN/K. (**B**) Temperature dependence of the persistence length, l_p, (blue), and the interphosphate distance, l_d, (orange). Linear fits to data (blue and orange lines, respectively) give $\partial l_p/\partial T = (6.2 \pm 1.4) \cdot 10^{-3}$nm/K and $\partial l_d/\partial T = (4.5 \pm 1.0) \cdot 10^{-4}$nm/K. The measured values of f_m, l_p, and l_d are reported in Table 7.1.

(29bp) makes the evaluation of the stretching terms easier. The stiffness of short handles is much larger as compared to the trap stiffness ($k_h \gg k_b$), implying that $k_{eff} \approx k_b$. Moreover, if the force varies in a relatively narrow range ($f_{max} - f_{min} \lesssim 10$pN), trap stiffness can be considered nearly force-independent so k_{eff} is constant along the folded branch of the FDC. Therefore, we can estimate k_{eff} by fitting the slope preceding the first FDC rip to the linear equation $f = k_{eff}x$ (orange dashed-line in Fig.7.2). This allows us to compute the (effective) contribution of the handles and optical trap, x_{eff}, to the total trap-pipette distance, λ. By computing $\lambda - x_{eff}$, we obtain the force-extension curve (FEC) corresponding to the extension of the ssDNA upon unzipping (grey dashed-line in Fig.7.2).

$T \pm 1$ [°C]	$T \pm 1$ [K]	f_m [pN]	l_p [nm]	l_d [nm]	Δs_{ss} [cal mol^{-1}K^{-1}]
7	280	19.72 (3)	0.74 (7)	0.647 (3)	3.33(2)
10	283	19.08 (4)	0.68 (2)	0.631 (9)	3.18(2)
13	286	18.71 (4)	0.73 (3)	0.662 (1)	3.07(2)
16	289	18.26 (7)	0.67 (2)	0.672 (1)	2.95(2)
19	292	17.87 (4)	0.78 (2)	0.655 (1)	2.84(2)
22	295	17.12 (2)	0.79 (3)	0.657 (1)	2.67(2)
25	298	16.75 (3)	0.77 (2)	0.647 (1)	2.58(2)
30	303	15.86 (2)	0.75 (2)	0.665 (1)	2.39(2)
35	308	14.96 (3)	0.88 (2)	0.639 (1)	2.21(1)
42	315	14.06 (4)	0.88 (4)	0.641 (1)	2.02(1)

Table 7.1: Experimentally measured average unzipping force, f_m, persistence length, l_p, interphosphate distance, l_d, and ssDNA entropy per base, Δs_{ss}, (see text) in the studied temperature range (reported both in Celsius and Kelvin degrees). The errors (in brackets) refer to the last digit.

We derived the elastic properties of the ssDNA by fitting the FEC unfolded branch, i.e. the slope following the last FEC rip, to the WLC model in Eq.(5.19) (green dashed-line in Fig.7.2). The measured persistence length, l_p, and interphosphate distance, l_d, at each experimental temperature are shown in Fig.7.3B and reported in Table 7.1. As temperature increases, the ssDNA persistence length, l_p (blues squares), varies from $l_p^{280K} = 0.74 \pm 0.07$nm to $l_p^{315K} = 0.88 \pm 0.04$nm ($\approx +30\%$) in the studied T range. A linear fit to data gives $\partial l_p/\partial T = (6.2 \pm 1.4) \cdot 10^{-3}$nm/K (blue line). Analogously, the interphosphate distance, l_d (orange circles), also depends on temperature as $\partial l_d/\partial T = (4.5 \pm 1.0) \cdot 10^{-4}$nm/K (orange line). As the l_d variation is of $\mathcal{O}(10^{-3})$nm over the whole temperature range, it can be considered approximately constant.

7.2 NNBP FREE-ENERGIES AND TEMPERATURE

We derived the 8 NNBP DNA free energies at 1M NaCl from the experimental FDCs at each studied temperature (see Fig.7.1). According to the NN model, the NA energetics of hybridization is described by 16 NNBP

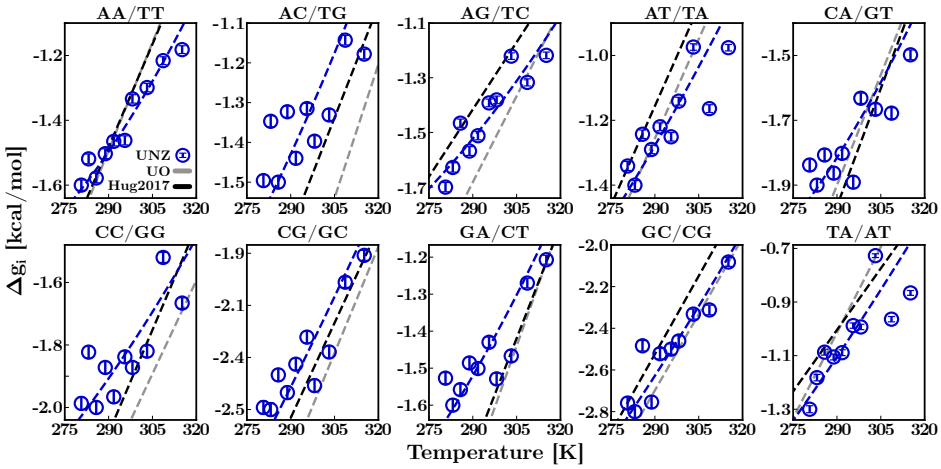


Figure 7.4: The 10 NNBP DNA free energies measured from unzipping a 3.6kbp hairpin in the temperature range $[280, 315]$ K (see text). The free energy of the two motifs GC/CG and TA/AT has been computed by applying the circular symmetry relations. A fit to data (blue line) has been added to compare with the temperature dependence of the free energies predicted with the UO 10 NNBP values (grey line) and the 8 NNBP values of Huguet *et al.* [74] (black line).

parameters (see Sec.5.2). The number of parameters reduces to 10 due to Watson-Crick complementarity and eventually to 8 independent values by considering the circular symmetry relations given in Eqs.(5.14).

The NNBP DNA energies have been measured using a Monte-Carlo optimization algorithm [74, 133]. Starting with an initial guess of the 8 NNBP DNA energies, at each step of the optimization, a random increment of the energies is proposed, and a prediction of the FDC is generated. The error in approximating the experimental FDC with the theoretical one, E, drives a Metropolis algorithm: a change of the energy parameters is accepted if the error difference to the previous step is negative ($\Delta E < 0$). Otherwise ($\Delta E > 0$), the proposal is accepted if $\exp(-\Delta E/T) < r$ with r a random number uniformly distributed $r \in U(0,1)$. The algorithm continues until convergence is achieved, i.e. until ΔE is smaller than a given threshold.

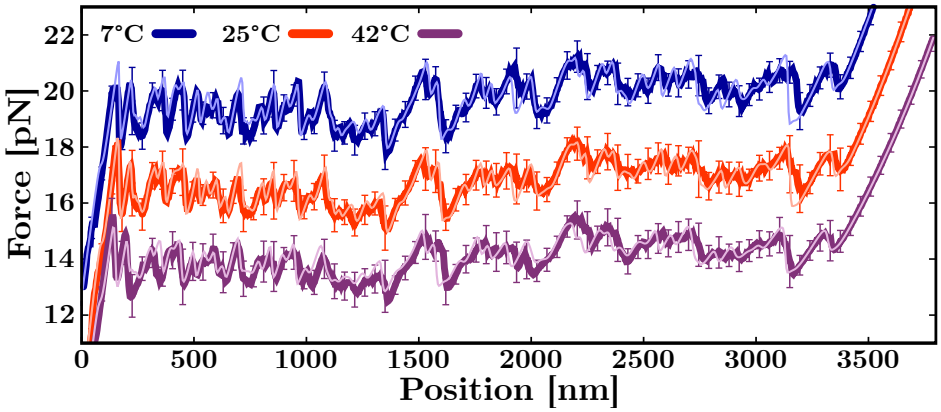


Figure 7.5: Comparison of the experimental average FDCs (dark-colored lines) with the theoretical predictions (light-colored lines) obtained by using the measured NNBP DNA free energies at 7°C, 25°C, and 42°C. Analogous results have been obtained for the other temperatures experimentally studied.

We measured the NNBP free energies at each experimental temperature using this algorithm. To generate the FDC prediction needed by the Monte-Carlo algorithm at each T, we used the ssDNA elastic parameters experimentally measured at the respective temperature (see Table 7.1). The measured NNBP values are shown in Fig.7.4 and reported in Table C.1 of Appendix C. Notice that the free energies of NNPB GC/CG and TA/AT have been computed by applying the circular symmetry relations. Remarkably, our results (blue circles and line) agree (within errors) both with the UO set (black line) and the Huguet *et al.* values (grey line) over the range $[280, 315]$ K. Notice that the agreement with Huguet *et al.* is better than with the UO set. Moreover, we tested the accuracy of the extracted 8 NNBP energies at each T by computing the theoretical FDC at different temperatures. The results are shown in Fig.7.5.

7.3 DNA THERMODYNAMICS AT THE SINGLE BASE-PAIR LEVEL

In an unzipping experiment, the mechanical work done along the forward (reverse) process is given by Eq.(4.3) (Eq.(4.4)) so that $W = \int_{\lambda_A}^{\lambda_B} f d\lambda$ (see Sec.4.1), where λ_A and λ_B are two arbitrary values of the control parameter λ (the trap-pipette distance).

Let us now consider an FDC unzipping curve: we can approximate the typical sawtooth pattern by a horizontal line corresponding to the average unzipping force, f_m. In the plateau region, the contributions to the total distance from the optical trap and the ds handles ($x_b(f_m)$ and $x_h(f_m)$ in Eq.(5.15)) are approximately constant. Thus, the change in λ when moving from λ_A to λ_B corresponds to the change in the ssDNA extension so that $\Delta\lambda = \lambda_B - \lambda_A \approx \Delta x_{ss}(f_m)$.

Therefore, by recalling that the total system's free energy is the sum of the double-helix hybridization energy, ΔG_0, plus the elastic contributions (see Eq.(5.21)), we can write

$$W = f_m \Delta\lambda \equiv \Delta G = \Delta G_0(T) + \int f_{ss}(x) dx \,, \tag{7.1}$$

where $\Delta G_0(T)$ is the hybridization free energy at temperature T between positions λ_A and λ_B. The integral accounts for the stretching of the ssDNA, modeled with the WLC in Eq.(5.19). Integrating Eq.(7.1) by parts and taking into account that $\Delta\lambda \approx \Delta x_{ss}$, gives

$$\Delta G_0(T) = -\int_0^{f_m(T)} \Delta x_{ss}(f, T) df \,. \tag{7.2}$$

Let us notice that, following the convention by which the unfolded (initial) state is taken as a reference, the hybridization free energy of NAs is taken as negative ($\Delta G_0 = G_F - G_U$). In this case, ΔG_0 is maximal when the stem is completely formed, and it decreases upon unzipping the molecule. This

convention is adopted for the definition of all thermodynamic quantities in this thesis chapter.

7.3.1 *The Clausius-Clapeyron Equation*

The entropy change, ΔS, for a system subjected to mechanical deformation can be obtained from the Clausius-Clapeyron equation [134, 135] where f and λ are equivalent to pressure and volume. At constant pressure and for a given value of the control parameter, λ, the free energy can be written as $dG = -SdT - \lambda df$. Let us consider an unzipping process starting at λ_A and ending at λ_B. Therefore, at equilibrium, $-S_B dT - \lambda_B df = -S_A dT - \lambda_A df$ so the Clausius-Clapeyron equation reads

$$\frac{\partial f}{\partial T} = -\frac{S_B - S_A}{\lambda_B - \lambda_A} = -\frac{\Delta S}{\Delta \lambda}\,. \tag{7.3}$$

Notice that $\Delta S = S_F - S_U$ in Eq.(7.3) is defined as negative according to the convention by which the unfolded (initial) state is taken as a reference.

To derive the entropy, let us recall that $\Delta S = -\partial \Delta G/\partial T$ and that, at f_m, $\Delta\lambda \approx \Delta x_{ss}$. Thus, from Eqs.(7.2) and (7.3), we eventually obtain the entropy change at zero force,

$$\Delta S_0(T) = \frac{\partial f_m(T)}{\partial T}\Delta x(f_m(T)) + \int_0^{f_m(T)} \frac{\partial \Delta x(f,T)}{\partial T} df\,. \tag{7.4}$$

The second term on the r.h.s of Eq.(7.4) accounts for the (positive) contribution, ΔS_{ss}, to stretch the ssNA and orient the molecule dipole from zero up to force $f_m(T)$. To compute this term, we used the measured elastic parameters (see Sec.7.1) to predict the temperature dependence of the stretched ssDNA according to the WLC. The measured ΔS_{ss} per base (Δs_{ss}) at each T are shown in Fig.7.6 and reported in Table 7.1. A fit to data according to the thermodynamic equation $\Delta s_{ss}(T) =$

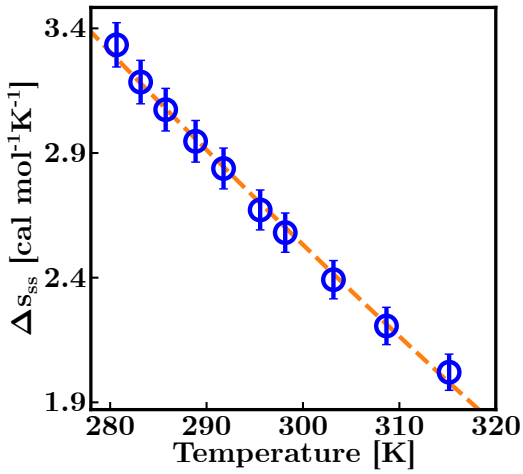


Figure 7.6: Temperature dependence of entropy change per base (Δs_{ss}) to stretch the ssDNA and orient the molecule dipole from zero up to force $f_m(T)$ (see Eq.(7.4)). The measured values at each T are reported in Table 7.1.

$\Delta s_{ss,0} + \Delta c_p^{ss} \log(T/T_m)$ (orange dashed line), gives the ssDNA specific heat change per base at zero force, $\Delta c_p^{ss} = -11.2 \pm 0.2\,\text{cal}\,\text{mol}^{-1}\text{K}^{-1}$.

7.3.2 *Derivation of the NNBP Entropies*

Given the DNA FEC at temperature T, we first measured the coordinates of all the FEC peaks, i.e. the position and the force at which the rips occur (orange dots in Fig.7.7). Then, we computed the distance between all possible combinations of the positions of the peaks, $\Delta x = x^B - x^A$, where A(B) denotes the initial (final) point of a segment. This gives the set of segments $\{\Delta x_k\}$ with $k = 1, 2, \ldots, K$ at temperature T. By knowing the dependency of the unzipping average force with the temperature $\partial f_m/\partial T = -0.165 \pm 0.03\text{pN/K}$ (see Fig.7.3A in Sec.7.1), the Clausius-Calpeyron relation in Eq.(7.4) can be applied to measure the total entropies set, $\{\Delta S_{0,k}^{exp}\}$.

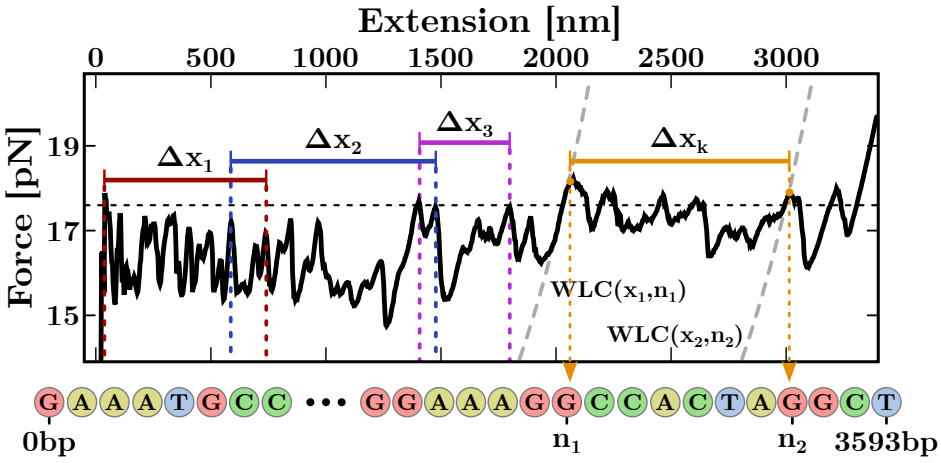


Figure 7.7: Application of the Clausius-Clapeyron relation (see Eq.(7.4)) to the experimental FDC (see text).

Moreover, we computed the number of open bases, $(n^{\mathrm{A}}, n^{\mathrm{B}})_k$, corresponding to the extremal points, $(x^{\mathrm{A}}, x^{\mathrm{B}})_k$, of each Δx_k. Given the coordinates of a peak, $(x, f)_k^{\mathrm{A(B)}}$, we obtained $n_k^{\mathrm{A(B)}}$ by inverting the WLC (see Appendix A) passing for point $(x, f)_k^{\mathrm{A(B)}}$ (dashed grey lines in Eq.(5.19)). Notice that this computation requires the knowledge of the ssDNA elastic properties at temperature T (see Fig.7.3B in Sec.7.1)). Given $(n^{\mathrm{A}}, n^{\mathrm{B}})_k$ for segment Δx_k, we identified the corresponding NNBP motifs that are released along the hairpin sequence (see Fig.7.7). Notice that the Clausius-Clapeyron equation discussed in Sec.7.3.1 can be directly applied to experimental curves (FDCs or FECs) as it relates the system entropy with an extension change. In fact, at the average unzipping force, we can assume $\Delta\lambda_k \approx \Delta x_{\mathrm{ss},k}$. However, estimating the number of open bases, n_k, at each point, x_k, from unzipping data requires accounting for the elastic contributions of the optical trap and dsDNA handles. Therefore, in what follows, we will work with the experimental FECs, computed as described in Sec.7.1.

In Sec.5.2 we introduced the NN model according to which the hairpin free-energy of formation can be described by considering all possible motifs given by two complementary bases (according to the Watson-Crick complementarity rules) and their nearest-neighbors along the $5' \to 3'$ direction. This allows us to describe the NA energetics with 16 NNBP parameters, reduced to 8 once the appropriate symmetry relations are considered. Therefore, the hairpin hybridization free energy is given by $\Delta G_0 = \sum_i^N \Delta g_{0,i}$, where N is the total number of base pairs of the sequence and $\Delta g_{0,i}$ denotes the free energy of the i-th NNBP motif. The NN model can be straightforwardly extended to the total system entropy and enthalpy, which gives $\Delta S_0 = \sum_i^N \Delta s_{0,i}$ and $\Delta H_0 = \sum_i^N \Delta h_{0,i}$, respectively. Hereafter, we will adopt the lightened notation $\Delta q_{0,i} \to \Delta q_i$ for the thermodynamic parameter, Δq_i, of NNBP motif i at zero force.

By combining the NN description of the total hairpin entropy and the Clausius-Clapeyron equation, the total entropy of each segment, Δx_k (with $k = 1, 2, \ldots, K$), can be written as

$$\Delta S_k^{\text{exp}} = \sum_{i=\text{AA,CA,...}} c_{k,i} \Delta s_i + n_k \Delta s_{ss} \,, \tag{7.5}$$

where the sum runs over all NNBP motifs, Δs_i is the NNBP entropy of motif i, $c_{k,i}$ is the multiplicity coefficient of motif i, i.e. the number of times motif i appears in the interval $(n^{\text{A}}, n^{\text{B}})_k$, n_k is the total bp number of segment k, and Δs_{ss} is the ssDNA stretching entropy (see Eq.(7.4) in Sec.7.3.1). Notice that, by construction, $\sum_i c_{k,i} \equiv n_k$.

The derivation of the 8 independent NNBP entropies of DNA corresponds to solving the non-homogeneous linear system of K equations and $I = 8$ parameters (being $K \gg I$) in Eq.(7.5). Different methods can be used to do this computation. Given its fast convergence when handling large data sets [136, 137], here we chose a stochastic gradient descent (SGD) optimization algorithm (see Appendix B for more details). The

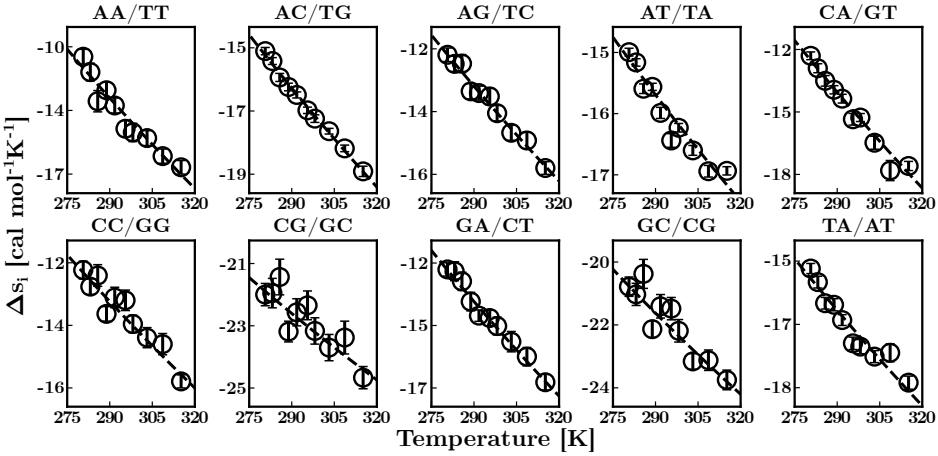


Figure 7.8: The 10 NNBP DNA entropies measured from unzipping a 3.6kbp hairpin in the temperature range $[280, 315]$ K (see text). The entropy of the two motifs GC/CG and TA/AT has been computed by applying the circular symmetry relations. A fit to data (grey dashed line) according to Eq.(7.7a) allows for extracting the specific heat change per motif (see Sec.7.4).

measured NNBP DNA entropies are reported in Fig.7.8 as a function of the temperature in the studied range, $[280, 315]$ K. The complete set of results is reported in Table C.2 of Appendix C.

7.3.3 *Measure of the NNBP Enthalpies*

We derived the NNBP enthalpies from the experimental measurements of the free energies and entropies per motif. Let us first recall that at a given environment temperature, T (in Kelvin degrees),

$$\Delta g_i = \Delta h_i - T\Delta s_i \,, \tag{7.6}$$

where Δg_i, Δh_i, and Δs_i are the free energy, enthalpy, and entropy of motif i, respectively.

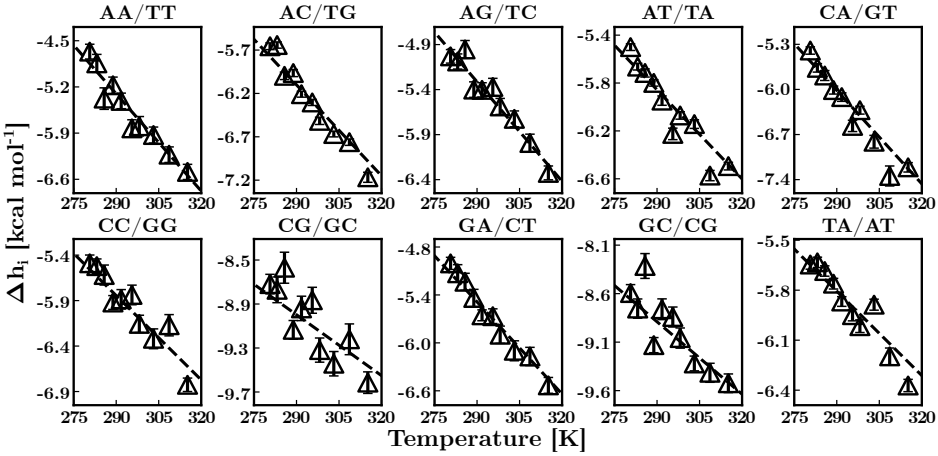


Figure 7.9: The 10 NNBP DNA enthalpies measured from unzipping a 3.6kbp hairpin in the temperature range $[280, 315]$ K (see text). The enthalpy of the two motifs GC/CG and TA/AT has been computed by applying the circular symmetry relations.

Given Eq.(7.6), it is straightforward to get the NNBP entropies as $\Delta h_i = \Delta g_i + T\Delta s_i$. The measured parameters are shown in Fig.7.9 and reported in Table C.3 of Appendix C.

7.4 THE NNBPS SPECIFIC HEAT CHANGE

The entropy and enthalpy of NNBP motif i depend on temperature according to the thermodynamic relations

$$\Delta s_i(T) = \Delta s_{0,i} + \Delta c_{p,i} \log(T/T_{m,i}) \tag{7.7a}$$

$$\Delta h_i(T) = \Delta h_{0,i} + \Delta c_{p,i}(T - T_{m,i})\,, \tag{7.7b}$$

where $\Delta c_{p,i}$ is heat capacity change at constant pressure, T is the environment temperature (in Kelvin degrees), and $T_{m,i}$ is the melting temperature, i.e. the temperature (in Kelvin) at which motif i can be found in the

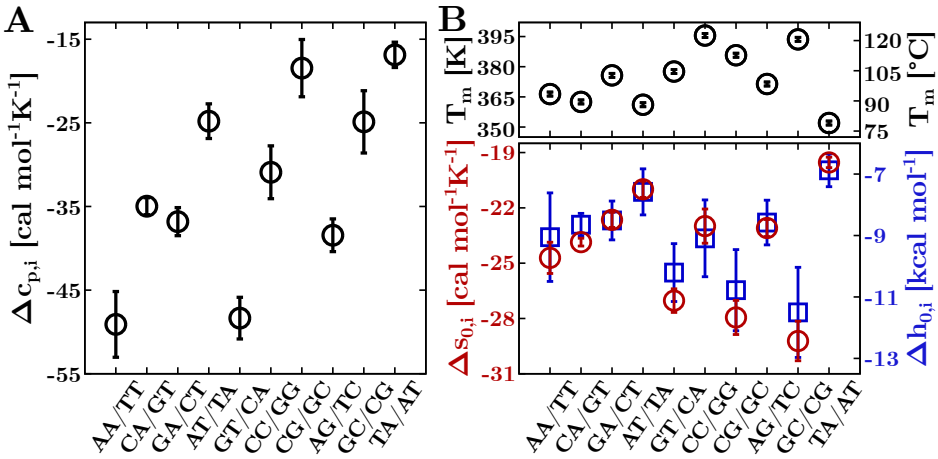


Figure 7.10: DNA NNPB thermodynamics. (**A**) Measured heat capacity change per motif. (**B**) Melting temperatures (top), entropies, and enthalpies at T_m (bottom) for each of the 10 NNBP parameters. Notice that results for motifs GC/CG and TA/AT have been derived by applying circular symmetry relations.

melted or hybridized state with equal probability. Finally, $\Delta s_{0,i}$ and $\Delta h_{0,i}$ are the entropy and enthalpy at $T = T_{m,i}$, respectively.

Given Eq.(7.7a), we measured the heat capacity change per motif, $\Delta c_{p,i}$, by fitting the experimentally measured entropies to the equation $A_i + \Delta c_{p,i} \log(T)$, being $A_i = \Delta s_{0,i} - \Delta c_{p,i} \log(T_{m,i})$ (grey dashed line in Fig.7.8). The results are shown in Fig.7.10A and Table 7.2 (column 1).

To derive $\Delta s_{0,i}$, $\Delta h_{0,i}$ and $T_{m,i}$ from the experimental data, let us fit the NNBP enthalpies to Eq.(7.7b), which can be rewritten as $B_i + \Delta c_{p,i} T$, where $B_i = \Delta h_{0,i} - \Delta c_{p,i} T_{m,i}$ (grey dashed line in Fig.7.9). Then, by plugging in Eqs.(7.7) in Eq.(7.6), one gets

$$\begin{aligned} \Delta g_i(T) &= \Delta h_i(T) - T\Delta s_i(T) = \\ &= \Delta h_{0,i} + \Delta c_{p,i}(T - T_{m,i}) - T\left(\Delta s_{0,i} + \Delta c_{p,i} \log\left(\frac{T}{T_{m,i}}\right)\right) \\ &= B_i + \Delta c_{p,i} T - T\left(A_i + \Delta c_{p,i} \log(T)\right) . \end{aligned} \tag{7.8}$$

NNBP	$\Delta c_{p,i}$ [cal mol^{-1}K^{-1}]	$\Delta s_{0,i}$ [cal mol^{-1}K^{-1}]	$\Delta h_{0,i}$ [kcal mol^{-1}]	$T_{m,i} \pm 1$ [K]	[°C]
AA/TT	−49(4)	−24.7(0.8)	−9.1(1.4)	366	93
AC/TG	−35(1)	−23.9(0.2)	−8.6(0.4)	363	89
AG/TC	−37(2)	−22.7(0.4)	−8.5(0.6)	376	103
AT/TA	−25(2)	−21.0(0.4)	−7.6(0.8)	361	88
CA/GT	−48(2)	−27.0(0.6)	−10.2(0.9)	378	104
CC/GG	−31(3)	−23.0(0.9)	−9.1(1.3)	396	122
CG/GC	−18(3)	−27.9(0.9)	−10.8(1.3)	386	113
GA/CT	−38(2)	−23.1(0.5)	−8.6(0.7)	371	98
GC/CG	−25(4)	−29.2(1.1)	−11.5(1.5)	394	120
TA/AT	−17(2)	−19.5(0.3)	−6.9(0.5)	352	79

Table 7.2: Experimentally measured heat capacity change ($\Delta c_{p,i}$), entropy ($\Delta s_{0,i}$) and enthalpy ($\Delta h_{0,i}$) at the melting temperature ($T_{m,i}$) for each NNBP motif i (both in [K] and [°C]). The experimental error is reported in brackets. Motifs GC/CG and TA/AT have been derived by applying circular symmetry relations.

By definition, the free energy is zero at the melting temperature ($T = T_{m,i}$). As all parameters in Eq.(7.9) have been determined by fitting entropies and enthalpies, we measured the NNBP melting temperatures by searching for the temperature at which $\Delta g_i(T_m) \equiv 0$ kcal/mol. The measured values are shown in Fig.7.10B (top) and Table 7.2 (column 4). Given $T_{m,i}$ per motif, a new fit of the entropies and enthalpies to Eqs.7.7 allowed to derive $\Delta s_{0,i}$ and $\Delta h_{0,i}$. The results are reported in Fig.7.10B (bottom) and Table 7.2 (columns 2 and 3, respectively).

7.5 PREDICTION OF DNA MELTING TEMPERATURES

We tested the validity of our NNBP thermodynamics parameters by computing the melting temperatures, T_m, for the set of 92 DNA oligonucleotides (with a length between 10bp and 30bp) that were initially studied in bulk experiments by Owczarzy *et al.* [138].

The melting temperature of NAs is defined as

$$T_m = \frac{\Delta H_0}{\Delta S_0 + R \log\left[\frac{c_T}{4}\right]}, \tag{7.9}$$

where ΔH_0 (ΔS_0) is the total enthalpy (entropy) of the oligo at T_m, $R = 1.987\,\mathrm{cal\,mol^{-1}K^{-1}}$ is the ideal gas constant, and $[c_T]$ is the oligonucleotide concentration in [M] (in this case $[c_T] = 2 \cdot 10^{-6}$ M [138]). The factor $1/4$ must be included for non-self-complementary molecules. For each oligo, ΔH_0 (ΔS_0) has two contributions: the NNBP contribution and the initiation term. The initiation term depends on the sequence's first and last bp. We used the initiation terms reported in [22], assuming that these values depend on the temperature.

We solved Eq.(7.9) by minimizing the loss function $(T - T^*)^2$. At each iteration of the algorithm, a new temperature T is proposed, and the corresponding $\Delta H(T)$ and $\Delta S(T)$ are computed. Then, the latter values are used to compute T^* according to Eq.(7.9). At the end of the optimization, the temperature minimizing the loss function corresponds to the melting temperature, T_m. Our predictions (T_m^{unz}) for the set of 92 oligos at 1M NaCl are reported in Table C.4 of Appendix C together with the experimentally measured values [138] (T^{exp}) and the predictions by the UO set (T_m^{UO}) and the Huguet *et al.* parameters [133] (T_m^{Hug}).

To check the accuracy of our results, we defined the error function

$$\chi^2 = \frac{1}{N}\sum_{j=1}^{N}\left(T_{m,j}^{\mathrm{exp}} - T_{m,j}^{pred}\right)^2, \tag{7.10}$$

which accounts for the difference between the experimental measure $T_{m,j}^{\mathrm{exp}}$ [138] and the predicted value $T_{m,j}^{\mathrm{pred}}$ (being *pred* = unz, UO, Hug) for each of the $N = 92$ oligos. Our predictions give $\chi^{\mathrm{unz}} = 5.1(^\circ\mathrm{C})^2$ with an average temperature error, $\langle T_m^{\mathrm{unz}}\rangle = 2^\circ$C. Remarkably, our predictions agree with the experimental measurement (T^{exp}) within errors. However,

let us notice that a comparison with the other data sets ($\chi^{\mathrm{UO}} = 9.1(^{\circ}\mathrm{C})^2$ and $\chi^{\mathrm{Hug}} = 3.4(^{\circ}\mathrm{C})^2$, respectively [74]) shows that our results are more accurate than Mfold but are less accurate than Huguet *et al.*. This is because we used temperature-independent values for the initiation factors enthalpy and entropy [22, 74] to compute the oligos T_m. Even though our current methodology does not allow for measuring the initiation factors, the error committed is small, and our predictions agree with the literature.

7.6 DISCUSSION AND CONCLUSIONS

We studied the temperature dependence of the 8 NNBP DNA independent parameters (free energies, entropies, and enthalpies) by carrying out calorimetric unzipping experiments of a 3.6kbp DNA hairpin with a temperature-jump LOT in the range $[7, 42]^{\circ}\mathrm{C}$. We used the Clausius-Clapeyron thermodynamic relation (Eq.(7.3)) to extract the 8 NNBP entropy parameters from the unzipping curves at each temperature (see Sec.7.3.1). To do this, we developed a suitable stochastic gradient descent (SGD) algorithm to relate the experimentally measured entropies with the NN model prediction (see Sec.7.3.2 and Appendix B). We notice that an accurate characterization of the temperature dependence of the ssDNA elastic response is crucial to this task (Sec.7.1). In particular, we found that the T-dependence of the ssDNA entropy is significant, and its contribution cannot be neglected compared to the NNBP one (see Table 7.1). Given the 8 NNBP free energies measured from the experimental FDCs at each T (Sec.7.2), we obtained the 8 NNBP enthalpy parameters (Sec.7.3.3), thoroughly characterizing the NNBP DNA thermodynamics. These results allowed us to derive the specific heat change (Δc_p) for the 8 NNBP parameters (Sec.7.4).

To our knowledge, this is the first time that the specific heat variation is measured at the single bp level. Ultimately, we tested the validity of our results by computing the melting temperatures for the set of DNA hairpins experimentally studied by Owczarzy *et al.* [138]. Finally, let us notice that the measured Δc_p values ($\approx -32\,\mathrm{cal\,mol^{-1}K^{-1}}$ per motif) are in agreement with several studies reported by the literature [121, 124]. However, bulk experiments yield a broad range of measured Δc_p ($[-20, -160]\,\mathrm{cal\,mol^{-1}K^{-1}}$ per bp), depending on the experimental technique, setup, and molecular structure. DNA calorimetric bulk experiments are mostly done with short hairpins to avoid interactions between the unpaired ssDNA (knots formation, base-pairing between different hairpins) during the heating/cooling protocols. In short DNA sequences (a few tens of bases), stacking interactions are detrimental to the DNA folding process. The number of stacked bases is often comparable to the total number of bases in the sequence, largely affecting the ssDNA elastic properties and, ultimately, the free energy of hybridization. In contrast, in long DNA hairpins (a few kbp), stacking can involve several consecutive bases (a maximum of 11 consecutive purines has been found in our 3593bp sequence), but it weakly affects the overall folding dynamics as the number of stacked bases is small compared to the total number of bases in the sequence. By overcoming the limitations of bulk techniques, single-molecule calorimetry experiments on long DNAs permit an accurate estimation of the NNBP thermodynamic DNA parameters.

Part III

ENERGETICS AND KINETICS OF RNA FOLDING

8

THE ENERGETICS OF RNA

Unzipping experiments allows for studying the physico-chemical properties of NAs, from the thermodynamics of duplex formation to the folding of secondary and tertiary structures. In particular, DNA hybridization finds diverse applications in DNA nanotechnology, the construction of DNA origami, molecular robots, DNA walkers, switches, and nanomotors [139–143]. In the past decades, DNA unzipping permitted testing the validity of the NN model [22, 93–95] and to extract the NNBP free-energy parameters at different salt conditions [74, 133]. A precise knowledge of these parameters is key to unraveling hidden energy codes in molecular evolution [144]. In contrast, SM experiments on RNAs are much more complex than for DNAs, so most of the RNA characterization (especially at the level of single motif) relies on bulk experiments carried out at the end of the past century [23, 99–101]. In addition, the difficulties in preparing molecular constructs suitable for SM techniques and the complex phenomenology typical of RNAs make it hard to study in pulling experiments.

Here we derive the ten NNBP RNA energies from unzipping experiments on a 2kbp RNA hairpin in monovalent (sodium) and divalent (magnesium) salt conditions. To our knowledge, the NNBP RNA free-energies in magnesium have never been measured before. The large number of parameters of the NN model (10 or 8, if circular symmetry is

applied) requires a sufficiently long hairpin to infer the NNBP energies from unzipping experiments reliably. Two are the main difficulties of these experiments: first, the molecular synthesis of a long RNA hairpin is challenging; second, the FDC along the RNA sequence alternates reversible unzipping regions with irreversible ones that exhibit hysteresis, and multiple long-lived intermediates [145, 146]. Compared to DNA, where unzipping is practically reversible, a similar derivation of the RNA energies from irreversible FDCs is not possible. Here we derive the full equilibrium FDC in RNA by the piecewise assembly of the reversible parts and the reconstructed equilibrium ones for the irreversible regions. These are obtained by repeatedly unzipping and rezipping the RNA hairpin in these irreversible regions and using statistical physics methods based on fluctuation theorems.

Moreover, we validate the 100/1 salt-equivalence rule of thumb, stating that the concentration of a divalent salt equals 100-fold that of a monovalent one [110]. We demonstrate that NNBP free-energy parameters for a given magnesium concentration equal those in $77(\pm 49)$-fold sodium. We verify this phenomenological result by measuring the energies in sodium and magnesium at the level of individual NNBP.

Finally, we investigate the irreversibility and hysteresis in the FDCs, hypothesizing that they are caused by the formation of stem-loop structures along the unpaired single strands. Remarkably, the hysteresis along the unzipping-rezipping pathway directly correlates with the barrier energy landscape (BEL) defined by the stem-loops forming at the junction separating single strands and duplex. The analysis of the hairpin sequence in correspondence with the regions of large irreversibility shows that base stacking and base-pairing within the ssRNA promote the formation of stem-loop structures transiently stabilized at forces as high as 20pN. The stem-loops mechanism explains the slow kinetics and multiple trapping

conformations observed in RNA folding, with implications for the RNA folding problem [15, 145, 147–149].

8.1 UNZIPPING OF LONG RNA HAIRPINS

We use the LOT setup described in Chapter 3 to unzip a 2027bp RNA hairpin with short (29bp) hybrid DNA/RNA handles. A long RNA sequence is essential to extract the NNBP energy parameters. The synthesis of such a molecular construct required the development of a protocol tailored for the OT setup (see Appendix D.1 for details).

The experiments have been performed in buffers containing 100mM Tris-HCl (pH 8.1), 1mM EDTA and 500mM NaCl (monovalent salt) or 100mM Tris-HCl (pH 8.1) and 10mM $MgCl_2$ (divalent salt). Notice that the ionic strength of the buffers has to be corrected by adding 100mM Tris-HCl $\equiv$ 52mM $[Mon^+]$. The measured FDCs in sodium and magnesium are shown in Figs.8.1 and 8.2, respectively.

The data show that changing from $[Na^+]$ to $[Mg^{++}]$ strongly increases the irreversibility and hysteresis of the FDC. This makes experimentally inaccessible the beginning (the first 200bp, between 400nm and 650nm) and the end (the last 600bp, between 1800nm and 2200nm) of the FDC in magnesium. The observed hysteresis occurs in correspondence of specific regions along the FDC, each limited by at-equilibrium left (L) and right (R) states, and exhibiting intermediate states I_p, with $p = 1, \dots, P$. To efficiently sample these intermediates, we repeatedly unzipped and rezipped the RNA between the two limit positions (L, R), typically collecting a hundred trajectories per region. We identified 8 irreversible regions in sodium (Fig.8.1, insets) and 3 in magnesium (Fig.8.2, insets). Regions in sodium are numbered from 1 to 8. In magnesium, regions are numbered from 2 to 4/5 to underline the matching of the RNA sequences in those regions in sodium and magnesium, as evidenced by the number of

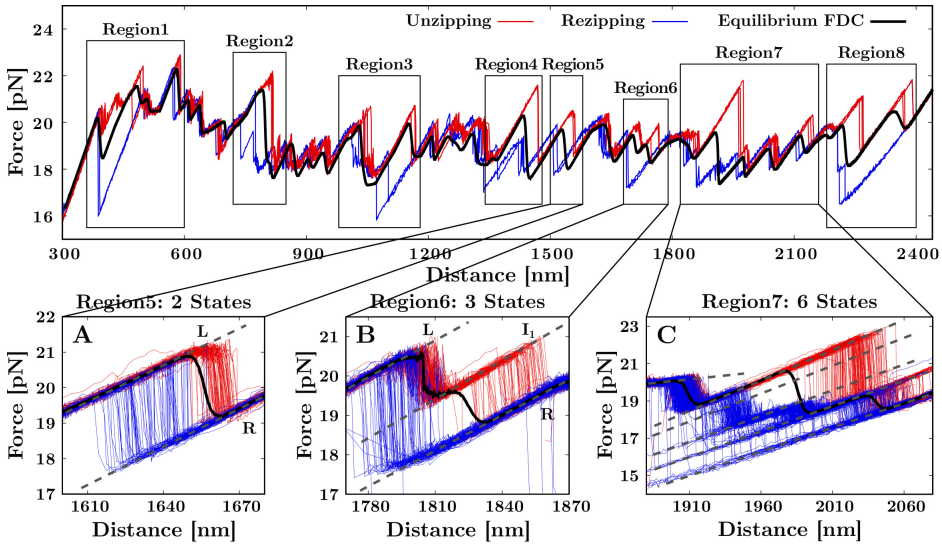


Figure 8.1: Unzipping/rezipping FDCs (red/blue) in 500mM NaCl. Black frames separate the different irreversible regions. The insets show repeated pulling cycles in regions of increasing complexity. The intermediate states (dashed grey lines) and the recovered equilibrium FDC (solid black line) are also shown. (**A**) shows a 2-states region (L/R) with no intermediates. Inset (**B**) shows a 3-states region (L,R and the intermediate I_1), while (**C**) reports a 6-states region (L,R and intermediates I_p with $p = 1, \ldots, 4$). The equilibrium FDC in the main box (black line) results by merging the reversible FDCs obtained for each region.

opened base pairs. The larger hysteresis observed in magnesium makes regions 4 and 5 in sodium merge into a single irreversible region (4/5). As explained above, the missing regions (1 and $6-8$) in magnesium result from their inaccessibility. Although a few regions do not contain intermediates (e.g., region 5 in Fig.8.1A), most exhibit more than one. The level of complexity of the unzipping-rezipping FDCs can be high, e.g., region 3 in magnesium shows 7 states (5 intermediates plus L and R, Fig.8.2B).

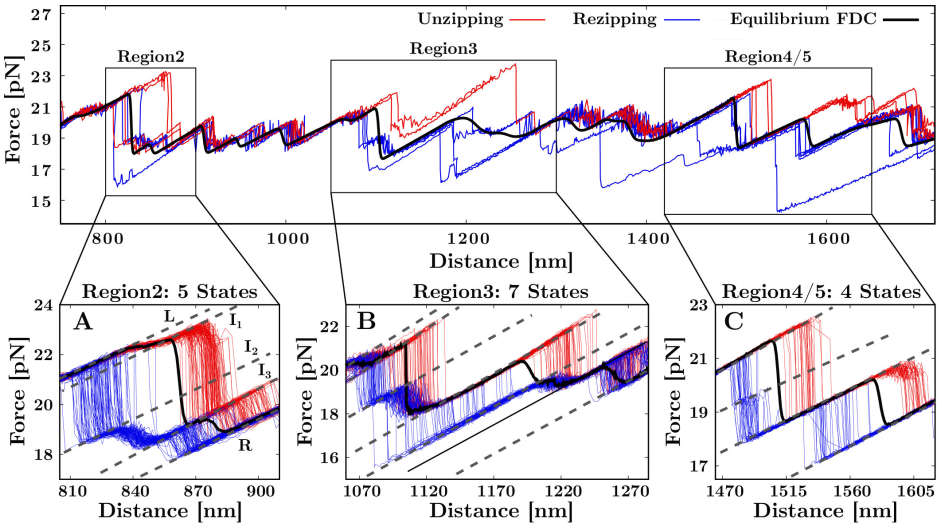


Figure 8.2: Unzipping/rezipping FDCs (red/blue) in 10mM $MgCl_2$. Black frames separate the different irreversible regions. The insets show repeated pulling cycles in regions of different complexity. Intermediate states (dashed grey lines) and the equilibrium FDC (solid black line) are shown. (**A**) shows a 5-states region (L,R and intermediates I_1, I_2, I_3), (**B**) shows a 7-states region (L,R and intermediates I_p with $p = 1, \ldots, 5$) and (**C**) shows a 4-states region (L,R and intermediates I_1, I_2). The equilibrium FDC in the main box (black line) results by merging the reversible FDCs obtained for each region.

8.2 RECOVERY OF THE EQUILIBRIUM FDC

RNA unzipping is an out-of-equilibrium process exhibiting a large hysteresis between unfolding and refolding FDCs. Therefore, an equilibrium FDC (black line in Figs.8.1 and 8.2) had to be computed from the experimental data. To do this, we applied the extended fluctuation relations (EFR), which have been introduced to recover the free energy of thermodynamic branches [56], kinetic states [80], and ligand binding energies [150, 151].

Given an irreversible region of the FDC limited by starting (left, L) and ending (right, R) equilibrated states, let $\mathcal{S}$ be the set of all the states

in that region $\mathcal{S} = (I_0 = L, I_1, I_2 \ldots, I_P, I_{P+1} = R)$ being $(I_1, \ldots, I_P)$ the partially equilibrated intermediates. During the experimental forward process (F) the trap position λ is moved at a constant speed starting in I_0 at λ_0 and ending in I_p at λ. Similarly, in the time-reversed protocol (R), the trap position is moved back at the same speed, starting in I_p at λ and ending in I_0 at λ_0. Thus, in analogy with Eq.(4.5), the extended fluctuation theorem (EFR) reads

$$\frac{\phi_F^{I_0 \to I_p}}{\phi_R^{I_p \to I_0}} \frac{P_F^{I_0 \to I_p}(W)}{P_R^{I_p \to I_0}(-W)} = \exp\left[\frac{W - \Delta G_{I_0 I_p}(\lambda)}{k_B T}\right], \tag{8.1}$$

where $P_F^{I_0 \to I_p}(W)$ ($P_R^{I_p \to I_0}(-W)$) is the partial distribution of the work W measured along the F (R) protocol, $\Delta G_{I_0 I_p}(\lambda) = G_{I_p}(\lambda) - G_{I_0}(\lambda_0)$ is the free-energy difference between states I_p at λ and I_0 at λ_0 and $\phi_F^{I_0 \to I_p}$ ($\phi_R^{I_p \to I_0}$) is the fraction of paths starting in I_0 (I_p) at λ_0 (λ) and ending in I_p (I_0) at λ (λ_0). k_B is the Boltzmann constant and T is the environment temperature.

To compute the free-energy, $\Delta G_{I_0 I_p}$, from bidirectional work measurements, let us define the extended Bennett acceptance ratio (EBAR) method. By using Eq.(4.9) and repeating the same computation shown in Sec. 4.2.1, one gets

$$\frac{u}{k_B T} = -\log\left(\frac{\phi_F^{I_0 \to I_p}}{\phi_R^{I_p \to I_0}}\right) + z_R(u) - z_F(u)\,, \tag{8.2}$$

where $z_R(u)$ and $z_F(u)$ are, once again, given by Eqs.(4.11). The logarithmic correction in the r.h.s of Eq.(8.2) accounts for the fraction of trajectories that end in state p. We notice that, in general, this number can be small, i.e. the number $n_{F(R)}$ of forward (reverse) trajectories that are in state p for a given λ is small, ranging from 1 to $20 - 25$ per state. However, it has been shown that they lead to reasonable free energy

estimates when applying the extended fluctuation relations to analogous cases [80, 150].

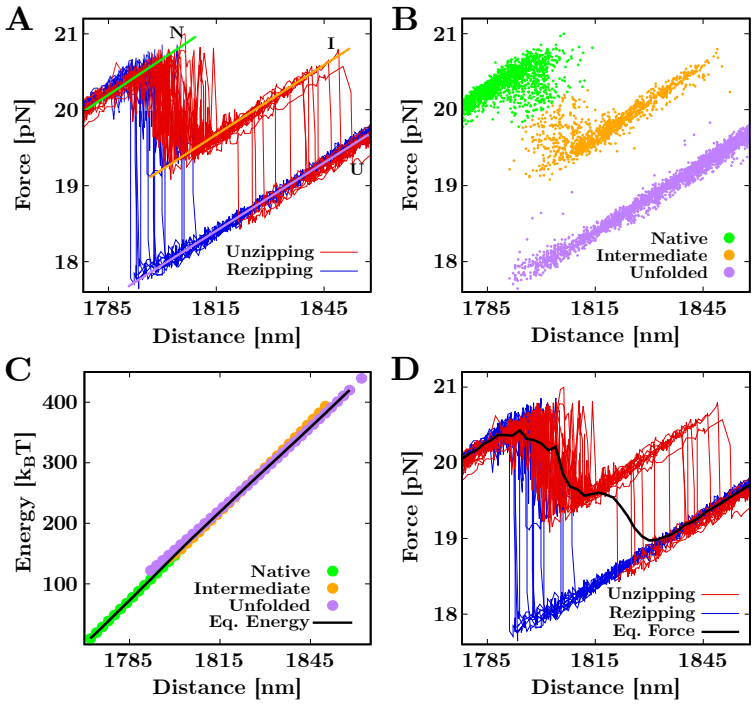


Figure 8.3: Reconstruction of the equilibrium FDC in the 3-states region measured in 500mM NaCl (from Fig.8.1). The protocol involves four main steps. **(A)** Given the unzipping/rezipping experimental trajectories, each state (Native, Intermediate, Unfolded) is identified and labeled (green, orange, and purple lines, respectively). **(B)** The experimental points are assigned to the closest state through a least-square fit. **(C)** The free energy of each state (U, I, N) is computed as the result of the combination of the Bennett and Jarzynski equations (see text). The equilibrium free energy between all the states (black line) is computed through (8.4). **(D)** Eventually, the equilibrium FDC (black line) is recovered by computing (8.5).

Thus, Eq.(8.2) is a self-consistent relation that returns the free energy of the transition $I_0 \to I_p$ at position λ by using information from both the folding and refolding trajectories. However, for given λ, some intermedi-

ates may only occur in the forward (reverse) process. This implies that $n_R = 0$ ($n_F = 0$) in Eqs.(4.11) and the EBAR cannot be computed. Because hysteresis affects the forward and the reverse processes differently, the number of intermediates is generally different between unzipping and rezipping, making EBAR unsuitable for our purpose. This is clearly shown in Fig.8.3A, where the unzipping FDCs always exhibit three states, whereas most of the rezipping trajectories directly go from R to L.

Let us introduce the extended Jarzynski estimator (EJE) to solve this problem. Given Eq.(4.12), it reads

$$\left\langle \exp -\frac{W_{F(R)}}{k_B T} \right\rangle_{F(R)} = \phi_{F(R)}^{S_{0(p)} \to S_{p(0)}} \exp\left(-\frac{\Delta G_{I_0 I_p}(\lambda)}{k_B T} \right), \tag{8.3}$$

which allows computing $\Delta G_{I_0 I_p}(\lambda)$ by only taking into account the trajectories of the forward (reverse) protocol that visits state I_p (unidirectional estimator). However, unidirectional free-energy estimators are slow-converging, requiring a larger number of trajectories than bidirectional estimators to obtain reliable free-energy measures (see Sec. 4.2.2). Therefore, to correct the bias of the measures obtained with the EJE, we used the results of the EBAR as a reference. This has been obtained with the following method.

Let us consider a state I_p stretching over the positions set $\{\lambda\}$. Firstly, we computed the energies $\{\Delta G(\lambda)_{I_0 I_p}\}^{\text{EBAR}}$ by solving (8.2) for all those λ having $n_F(\lambda), n_R(\lambda) \neq 0$. Then, the energies of the forward, $\{\Delta G_{I_0 I_p}(\lambda)\}_S^{\text{EJE}}$, and reverse, $\{\Delta G_{I_p I_0}(\lambda)\}_R^{\text{EJE}}$, processes have been separately computed for all the λ by using (8.3). By using the energy values computed with the EBAR $\{\Delta G_{I_p I_0}(\lambda)\}^{\text{EBAR}}$ as a reference, the closest intersection point with the Jarzynski energy sets $\{\Delta G_{I_p I_0}(\lambda)\}_F^{\text{EJE}}$ ($\{\Delta G_{I_p I_0}(\lambda)\}_R^{\text{EJE}}$) is computed. Eventually, the bias is corrected by applying a rigid shift to $\{\Delta G_{I_p I_0}(\lambda)\}_F^{\text{EJE}}$ ($\{\Delta G_{I_p I_0}(\lambda)\}_R^{\text{EJE}}$) to match the Bennett set. The bias correction is shown

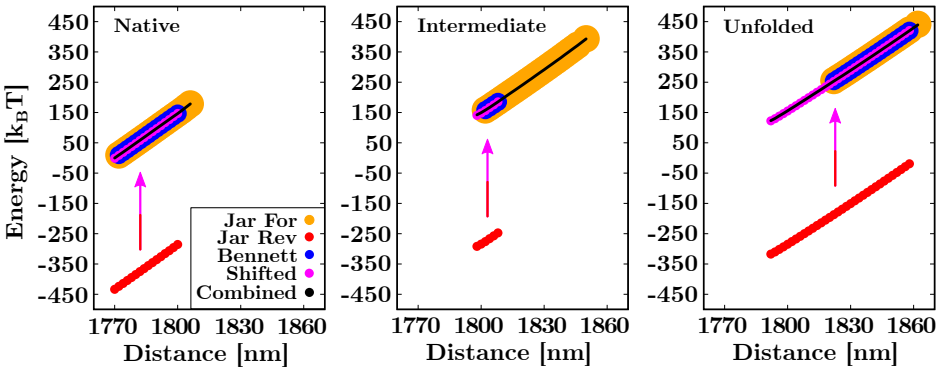


Figure 8.4: Free energy computation of the 3-states (Native, Intermediate, Unfolded) region showed in Fig.8.3. The free energy of each state is computed by combining (8.2) and (8.3). EBAR method (blue dots) only holds if $n_F, n_R \neq 0$ for each position λ, often leading to free energy estimations limited to a restricted data fraction (see the Intermediate state panel). The forward and reverse EJE are used to compute the energies of the forward (yellow dots) and reverse (red dots) trajectories for each λ. The (biased - see text -) results are eventually corrected (pink dots) according to the computed EBAR values, used as reference. This procedure gives the complete free-energy set of each state (black dots).

in Fig.8.4 for each state (Native, Intermediate, Unfolded) exhibited by the 3-states region in Fig.8.3.

Given the set $\{\Delta G_{I_0 I_p}(\lambda)\}$ for all the intermediates, I_p, the equilibrium free energy of one region (black line in Fig.8.3C) is given by

$$\Delta G_{\text{eq}}(\lambda) = -k_B T \log \left(\sum_{p=0}^{P+1} \exp - \frac{\Delta G_{I_0 I_p}(\lambda)}{k_B T} \right) \tag{8.4}$$

for any λ. Eventually, the equilibrium FDC (black line in Fig.8.3D) is recovered by computing

$$f_{\text{eq}}(\lambda) = \frac{\partial}{\partial \lambda} \Delta G_{\text{eq}}(\lambda) \,. \tag{8.5}$$

This method has been used to compute the equilibrium FDC in all 8 irreversible regions in NaCl and the 4 irreversible regions in $MgCl_2$. Eventually, the piecewise merging of the equilibrated FDC segments recovered the full equilibrium FDC. This allowed us to reconstruct the equilibrium FDCs for 7 molecules in sodium and 4 in magnesium.

8.3 DERIVATION OF THE NNBP RNA ENERGIES

We derived the RNA NNBP and loop energies from the equilibrium FDCs in sodium and magnesium. Let us remember that in the NN model (see Sec.5.2), the free energy of formation ΔG_0 of a NA duplex is defined as the sum over all couples of neighboring bases along the sequence, $\Delta G_0 = \sum_i \Delta g_{0,i}$ with $\Delta g_{0,i}$ the free energy of NNBP motif i. This gives 16 different NNBP motifs that reduce to 10 due to Watson-Crick complementarity. It is possible to further reduce this number from 10 to 8 independent parameters by considering the circular symmetry relations given in Eqs.(5.14). The 10 NNBP energies have been measured from melting experiments of short RNA duplexes of varying sequence and length [23, 99–101]. These values are accessible in the Mfold server [24]. Hereafter we will refer to such energies as the RNA Mfold values.

The NNBP energies have been measured using a Monte-Carlo Metropolis optimization based on comparing the equilibrium FDC with a theoretical prediction of the FDC (see Sec.5.5). The method has already been introduced in Sec.7.2 and is analogous to the one used in the DNA case [133], where equilibrium FDCs can be obtained within the experimental pulling speeds. Starting with an initial guess of the 8 independent NNBP parameters, at each step of the optimization, a random increment of the energies is proposed, and a prediction of the FDC is generated. The error made in approximating the experimental curve with the theoretical one, E, drives a Metropolis algorithm: a change of the energy parameters

NNBP	(1) $\Delta g^{Na}_{500mM,i}$	(2) $\Delta g^{Mg}_{10mM,i}$	(3) $\Delta g^{Na}_{0,i}$	(4) $\Delta g^{Mg}_{0,i}$	(5) $\Delta g^{Mg}_{14mM,i}$	(6) Mfold
AA/UU	−0.99 (6)	−1.11 (1)	−1.06 (6)	−1.57 (5)	−1.14 (7)	−1.12
CA/GU	−1.81 (6)	−2.12 (1)	−1.88 (6)	−2.58 (5)	−2.15 (7)	−2.14
GA/CU	−2.45 (7)	−2.77 (2)	−2.52 (7)	−3.23 (5)	−2.80 (7)	−2.73
AU/UA	−1.20 (4)	−1.06 (4)	−1.27 (4)	−1.52 (6)	−1.09 (8)	−1.09
GU/CA	−2.43 (6)	−2.53 (6)	−2.50 (6)	−2.99 (7)	−2.56 (9)	−2.41
CC/GG	−3.33 (4)	−3.21 (4)	−3.40 (4)	−3.67 (6)	−3.25 (8)	−3.26
CG/GC	−2.45 (7)	−2.35 (4)	−2.56 (7)	−2.81 (6)	−2.38 (8)	−2.23
AG/UC	−2.16 (5)	−1.96 (5)	−2.23 (5)	−2.42 (7)	−2.00 (9)	−1.93
GC/CG	−2.94 (8)	−2.95 (2)	−3.01 (8)	−3.41 (5)	−2.99 (8)	−3.82 [−2.77]
UA/AU	−1.03 (10)	−1.26 (7)	−1.10 (10)	−1.72 (8)	−1.29 (10)	−1.36 [−1.37]
Loop	0.16 (3)	—	0.09 (3)	—	—	0.14

Table 8.1: NNBP and loop RNA energies at $T = 298$K. (**Columns** 1**,** 2) NNBP energies in 500mM NaCl and 10mM $MgCl_2$, respectively. The last two values (GC/CG, UA/AU) have been computed with circular symmetry. (**Columns** 3**,** 4) NNBP values at the standard conditions of 1M NaCl and 1M $MgCl_2$, respectively. (**Column** 5) NNBP energies in magnesium reported at the concentration equivalent to 1M $Na^+ \equiv$ 14mM Mg^{++}. (**Column** 6) The Mfold 10 NNBP energies at 1M NaCl. NNBP values computed with circular symmetry are also reported (square brackets). Notice that the free energy of the loop in magnesium is not given (see text). All energies are in kcal/mol and have been reported with the statistical error computed over the different molecules (brackets).

is accepted if the error difference to the previous step is negative ($\Delta E < 0$). Otherwise ($\Delta E > 0$), the proposal is accepted if $\exp(-\Delta E/T) < r$ with r a random number uniformly distributed $r \in U(0,1)$. The algorithm continues until convergence is achieved, i.e. until ΔE is smaller than a given threshold. Let us note that because of the high number of parameters, only experimental data from the unzipping of long molecules (a few kbp) allow for an accurate estimation of the NNBP energies. The algorithm relies on the sawtooth pattern characteristic of the sequence to accept or reject an energies proposal: the longer the sequence is, the more accurate the values of the NNBP energies are. The elastic parameters of the model include the persistence and contour lengths of the hybrid DNA/RNA

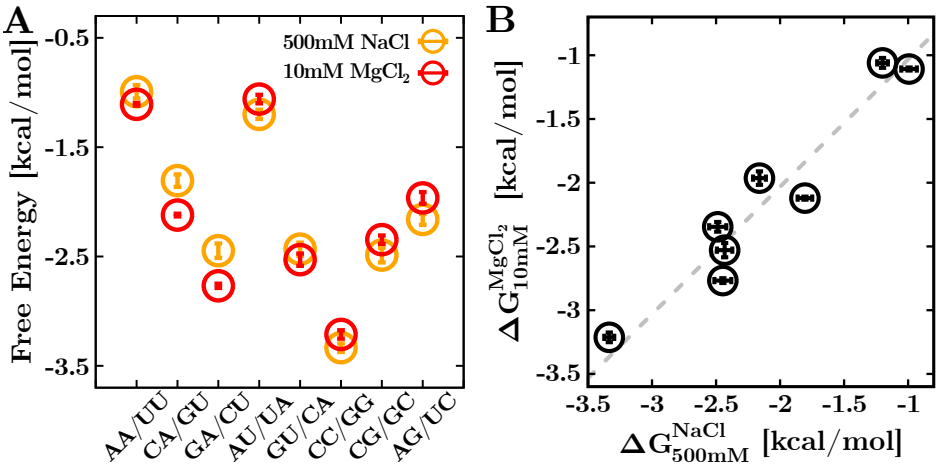


Figure 8.5: NNBP free-energy parameters. (**A**) Measured energies at 500mM NaCl (orange) and 10mM $MgCl_2$ (red). (**A**) Validation of the 100/1 salt equivalence rule. A fit to data according to Eq.(8.6) (gray line) gives the salt equivalence factor $a = 77 \pm 49$ (see text). Notice that here only the 8 NNBP independent parameters are shown.

handles ($P_{\mathrm{DNA/RNA}} = 10$nm and $L_{\mathrm{DNA/RNA}} = 7.8$nm) and those of the ssRNA ($l_p = 0.805$nm and interphosphate distance $l_d = 0.68$nm). The results, averaged over the different molecules, are summarized in Table 8.1 (columns 1, 2) and plotted in Fig.8.5A.

8.4 THE SALT EQUIVALENCE RULE

The measured NNBP energies support the validity of a salt equivalence rule between sodium and magnesium. To derive the equivalence rule, we plotted the measured energies in $[Mg^{++}] = 0.01$M as a function of the energies in $[Na^{+}] = 0.5$M fitting them to the relation

$$\Delta g_i^{\mathrm{Mg}}([\mathrm{Mg}^{++}]) = \Delta g_i^{\mathrm{Na}}([\mathrm{Na}^{+}]) - m \cdot \log\left(\frac{[\mathrm{Na}^{+}]_{\mathrm{eq}}}{[\mathrm{Na}^{+}]}\right), \qquad (8.6)$$

where $[Na^+]_{eq} \equiv a \times [Mg^{++}]$ is the magnesium concentration in sodium equivalents and a is the equivalence factor. $\Delta g_i^{Mg}([Mg^{++}])$ and $\Delta g_i^{Na}([Na^+])$ are the experimentally derived energies of motif i in (Mg^{++}) and (Na^+) at the respective salt concentrations in molar units. Finally, $m = 0.10 \pm 0.01$ kcal/mol is the NNBP-homogeneous monovalent salt correction parameter experimentally derived in [110],

$$\Delta g_i^{Na}([Na^+]) = \Delta g_{0,i}^{Na} - m \cdot \log([Na^+]) . \tag{8.7}$$

A least-squares fit to data gives $a = 77 \pm 49$ (Fig.8.5B), which is compatible with the value $a \approx 100$ of previous studies [110]. We expect that (8.6), with a constant over a broad range of magnesium concentrations, holds if Mg^{++} correlations and competitive effects between sodium and magnesium are weak. This implies diluted magnesium solutions, i.e. $[Mg^{++}] < 0.05$M [152, 153]. With added sodium, Mg^{++} effects dominate when $R = \sqrt{[Mg^{++}]}/[Na^+] > 0.22 M^{-1/2}$ [138], which is the case in our experimental conditions ($R = 2 M^{-1/2}$).

Given the measured energies (columns 1,2 in Table 8.1), we calculated the NNBP and loop values at the reference salt conditions of 1M NaCl ($\Delta g_{0,i}^{Na}$) and 1M $MgCl_2$ ($\Delta g_{0,i}^{Mg}$). By combining (8.6) and (8.7), we get

$$\Delta g_i^{Mg}([Mg^{++}]) = \Delta g_i^{Na}([Na^+]_{eq}) = \Delta g_i^{Na}(a \times [Mg^{++}]) . \tag{8.8}$$

The resulting energies in sodium and magnesium are given in columns 3 and 4 of Table 8.1, respectively.

For a direct comparison with the Mfold set, we use (8.8) to report the energies at 14mM $Mg^{++} \equiv$ 1M Na^+ (Column 5 in Table 8.1), obtained by using (8.8). Column 6 in the table shows the ten independent RNA Mfold energies plus the free energy of the loop. The last two NNBP values (indicated in brackets) are obtained from the circular symmetry relations applied to the other eight Mfold parameters. Notice that the Mfold value

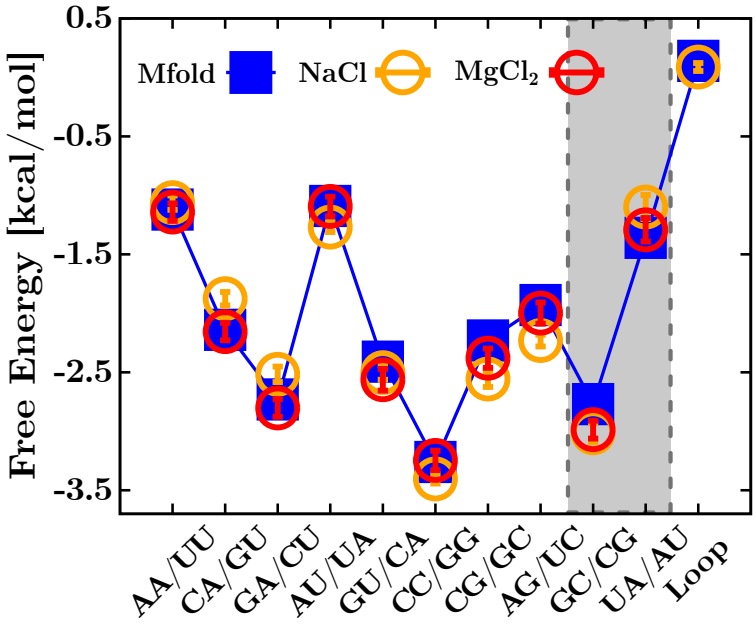


Figure 8.6: Comparison of the Mfold energies (blue) with the 1M NaCl and (the equivalent) 14mM $MgCl_2$ free-energy sets. The last two parameters (GC/CG and UA/AU – grey band) result from applying the circular symmetry relations. Notice that the free energy of the loop has not been measured in magnesium (see text).

for GC/CG (-3.82) is very different from our value in sodium (-3.01, column 3). This discrepancy arises from using eight parameters in our model, while Mfold uses ten. Interestingly, by applying the circular symmetry property to the Mfold set, we get for GC/CG the value -2.77, which is in better agreement with our value (-3.01). Notice that we do not give the free energy of the loop in magnesium, as this value cannot be measured due to the inaccessibility of the last part of the unzipping curve. Results in Table 8.1 (columns 3, 5, and 6) are plotted in Fig.8.6, which shows the overall agreement between the unzipping free-energy values and those of Mfold. For the total hybridization free energy of the RNA hairpin, the unzipping values predict $\Delta G_0^{\mathrm{Na}} = 4031$ kcal/mol (1M

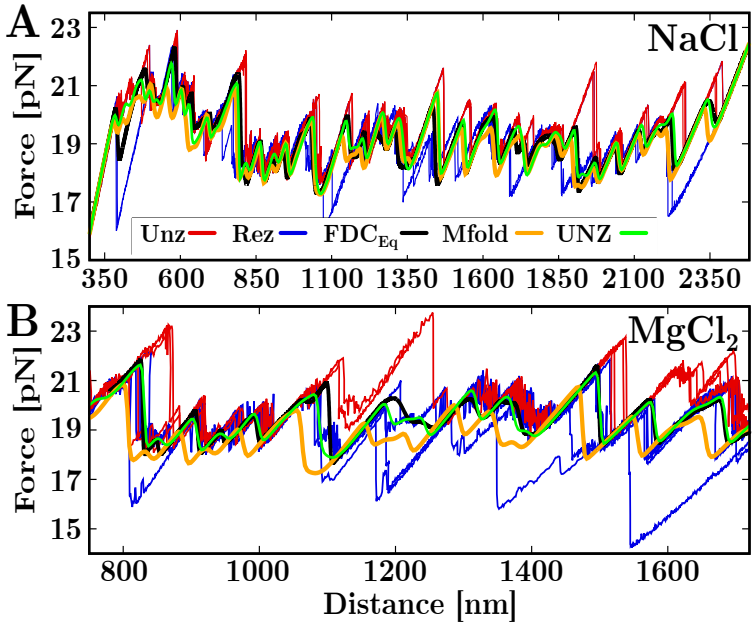


Figure 8.7: Comparison of the unzipping, rezipping, and equilibrium FDCs (in red, blue, and black, respectively) at 500mM NaCl (**A**) and 10mM $MgCl_2$ (**B**) with the theoretical prediction obtained from Mfold (orange), and the measured energies reported in columns 1 and 2 of Table 8.1 (green). Notice that in magnesium, the comparison is limited to the experimentally accessible region of the molecule (see text).

NaCl) and $\Delta G^{\mathrm{Mg}}_{14\mathrm{mM}} = 4082$ kcal/mol (14mM of equivalent $MgCl_2$). These numbers compare well to the Mfold value in sodium $\Delta G^{\mathrm{Mfold}} = 4086$ kcal/mol (1% relative error). In Fig.8.7A and B, we show the predicted FDCs from the measured energies in sodium and magnesium (columns 1 and 2 in Table 8.1), respectively. Notably, our predictions better agree with the experimental data than Mfold, particularly for magnesium (green and orange lines versus the black). Ultimately, a comparison of the theoretical FDCs predicted by the Mfold set with those obtained with our energies at 1M NaCl and 14mM $MgCl_2$ is shown in Fig.8.8.

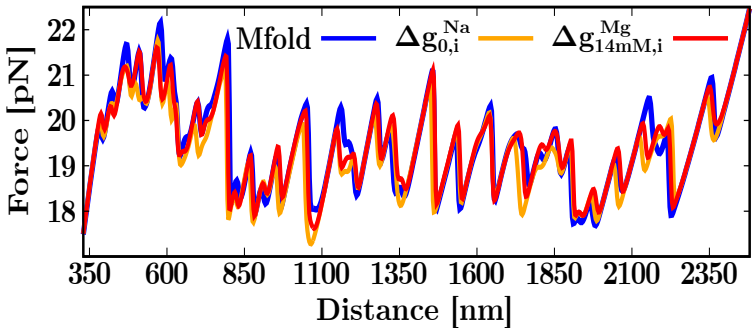


Figure 8.8: Comparison between the theoretical FDCs computed at the equivalent salt conditions $\Delta g^{Na}_{0,i}$ (orange), $\Delta g^{Mg}_{14mM,i}$ (red) and Mfold (blue) (columns 3, 5, and 6 of Table 8.1).

8.5 CONCLUSIONS

Detailed knowledge of the energetics of RNA hybridization is key to determining the thermodynamic stability of RNA structures, from dsRNA to tertiary RNAs, essential in many biophysical processes. By repeatedly unzipping and rezipping the RNA, we measured the sequence-dependent FDCs in sodium and magnesium. The large hysteresis observed along the FDCs demanded nonequilibrium physics methods to derive the fully reversible FDC from the irreversible pulling data. Quasi-static RNA unzipping experiments are not feasible as the lifetime of the intermediates requires exceedingly low pulling speeds. Estimates based on the Bell-Evans model range from 0.1nm/s to 1pm/s for irreversible hairpin segments of $30 - 40$ bp.

By using an optimization algorithm, we derived the free energies of the ten nearest-neighbor base pairs (NNBP) in RNA (Fig.8.6), finding good agreement with the Mfold values reported for sodium. The highest difference between our energies and Mfold is found for CG/GC in sodium, a relevant motif prone to methylation that accumulates in many regulatory regions [154, 155]. Moreover, the results for magne-

sium show the validity of a general salt equivalence rule 80/1 for which 10mM Mg^{++} corresponds to 800mM Na^{+} (Fig.8.5). Although the scope of this result has been tested in a single salt condition, its validity should span the dilute salt regime where cooperative salt effects are negligible ($[Mg^{++}] < 0.05$M) and competition effects with sodium are weak ($R = \sqrt{[Mg^{++}]}/[Na^{+}] > 0.22\,M^{-1/2}$). A salt equivalence rule has been disputed based on experimental data obtained in bulk experiments using atomic emission spectroscopy in buffer equilibrated samples [156]. Although this technique can determine the fraction of cations that are dissociated and bound to the RNA, it does not provide a direct measurement of free energies. Here we have demonstrated the validity of an 80/1 salt equivalence rule at the level of individual NNBP motifs. This is the most direct confirmation of the 100/1 rule of thumb for the equivalence of the non-specific binding energy of sodium and magnesium in RNA structures.

9

KINETICS OF RNA AND STEM-LOOPS MODEL

To characterize the irreversibility observed along the RNA unzipping FDCs (Figs.8.1 and 8.2), we hypothesize a scenario where stem-loop structures forming along the unpaired RNA strands slow down the stem formation and kinetically trap the system into the observed off-pathway metastable states, I_p [75]. The size of the force jumps in Figs.8.1, 8.2 indicate the number $\Delta n \in [50 - 150]$ of unzipped-rezipped bps between consecutive states, I_p. The large stacking free energy of RNA loops facilitates the formation of stem-loop structures at forces as high as 20pN where rezipping occurs (Fig.9.9A). The stabilizing effect induced by loop formation has been demonstrated in experiments of blocking oligos in nucleic acids hairpins. By hybridizing to the complementary loop region, these oligos prevent the formation of the native stem [58,157]. Stem-loops often contain hairpin-like folds with non-canonical base pairs (each colored structure in Fig.9.9A corresponds to a different number of bases) stabilized by stacking and base pairing interactions. To form the native stem, the two single strands pulled under opposite forces must come close to each other. However, this process facilitates the formation of off-pathway (misfolded) stem-loop structures in the single strands. In Fig.9.9B, we depict the hairpin unzipping at position n (middle) between two consecutive states, $I_p \to I_{p+1}$, which is slowed down by the transient formation of misfolded structures (M_p, left) consisting of stem-loops

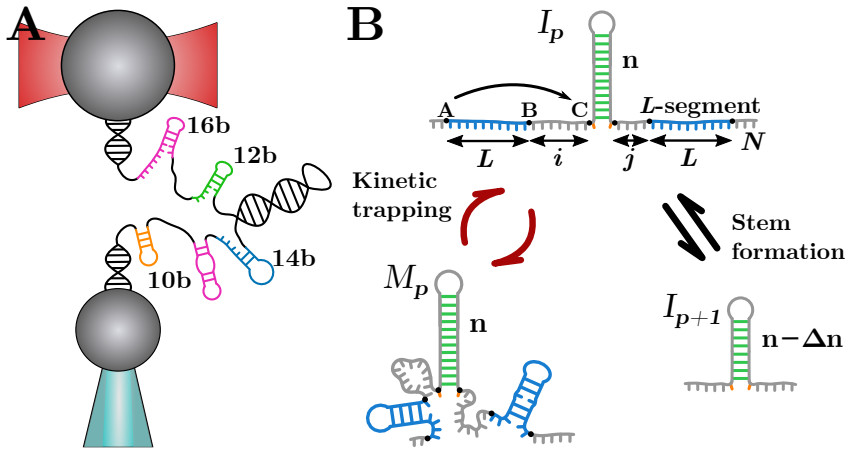


Figure 9.9: Stem-loops formation. **(A)** During the unzipping (rezipping) process, segments of different lengths (represented with different colors) along each RNA single strand form transient stem-loop structures. **(B)** Transition between intermediates I_p and I_{p+1} (black arrows). The formation of the off-pathway (misfolded) structures consisting of stem-loops (M_p) kinetically traps (red arrows) the RNA at I_p slowing down transitions $I_p \to I_{p+1}$ ($I_p \to I_{p-1}$) during unzipping (rezipping).

in the single strands (blue segments). The intermediate I_{p+1} (right) is rescued upon releasing Δn bases forming the stem-loops (M_p) that kinetically trap the hairpin. Notice that kinetic trapping occurs during rezipping for transitions $I_p \to I_{p-1}$. In the reversible regions, the misfolded states have very short lifetimes and are not observed, meaning that kinetic trapping and hysteresis effects are negligible at the experimental pulling speeds.

9.1 THE BARRIER ENERGY LANDSCAPE

The irreversibility of the unzipping-rezipping reaction can be modeled by introducing a many-valley barrier energy landscape (BEL) accounting for the off-pathway competing folds that can form along the two single

strands. The BEL describes the propensity of the hairpin to become kinetically trapped at a particular value of n by off-pathway conformations of high kinetic stability. The complexity of including all possible structures is enormous. Therefore we restricted the analysis to the single stem-loops (loop-BEL) stabilized by stacking and base pairing.

Let us consider all possible segments of L bases along the two unpaired RNA strands (referred to as 1 and 2). Let $\mathcal{S}_L^{(1,2)}$ be the set of all segments of length L contained in each strand of the RNA hairpin, $\mathcal{S}_L^{(1,2)} = \{[b_i, b_{i+L}]; 1 \leq i \leq N' = N - L\}$, where b_i and b_{i+L} stand for the initial and final base of the segment on strands $(1, 2)$ (N being the total number of bases in the hairpin). For a given L-segment, $[b_i, b_{i+L}]$, there are several competing folds, most of them stabilized by short complementary stems plus one or more loops of varying sizes (mostly $3 - 8$ bases). We have searched for the optimal fold of lowest free energy, $\epsilon^0_{L,i}$, by using the *DINAmelt web application* [158, 159] based on Mfold. This yields the optimal set of energies $\{\epsilon^0_{L,i}\}^{(1,2)}$ for $\mathcal{S}_L^{(1,2)}$ at standard conditions.

To construct the loop-BEL, one should consider all possible excitations (i.e. higher energy states) formed by multiple stem-loops folding along the two ssDNA strands at both sides of the junction. In principle, any number of stem-loops can form at arbitrary positions along the two strands. As the complexity of including all possible structures is enormous, in the most straightforward approach, we restricted the analysis to a single stem-loop (loop-BEL) per strand located at an arbitrary position. However, even if energetically favorable, stem-loops situated far away from the junction cannot interfere with the unzipping-rezipping of the hairpin, a reaction occurring precisely at the junction. Thus, for a given L, the loop-BEL at force f and junction position n is defined as

$$\Delta G_L(n, f) = -k_B T \log \sum_{j_1, j_2 = 0}^{N-n} \exp\left(-\frac{\Delta g_L^{(1)}(j_1, f) + \Delta g_L^{(2)}(j_2, f)}{k_B T}\right), \quad (9.9)$$

where $\Delta g_L^{(1,2)}(j,f)$ is the free-energy of forming a single stem-loop in strand $(1,2)$ of length L plus the work at force f to bring it from position j to the junction located at position n. Note that in Eq.(9.9) we assumed that all L-segments at the back of the junction are already hybridized into the native stem and do not contribute to the loop-BEL (green bps in Fig.9.9B). The term $\Delta g_L^{(a,b)}(i,f)$ is given by,

$$\Delta g_L^{(1,2)}(j,f) = \epsilon_{L,j}^{0(1,2)} + \int_0^f x_{L+j}(f')df' \,, \tag{9.10}$$

where $\epsilon_{L,j}^{0(1,2)}$ is the (negative) free energy of formation at zero force of the stem-loop $[b_j, b_{j+L}]$ in strand $(1,2)$. The integral is the energy cost to bring the $L+i$ bases from A to C at force f (Fig.9.9B). This term penalizes stem-loops that form far away from the junction as they cannot kinetically trap the stretched RNA. It has been modeled with the WLC model in Eq.(5.19) with $l_d = 0.68$nm the interphosphate distance [110, 160] and $l_p = 0.805$nm the RNA persistence length [59, 110]. Integration required inverting the WLC as described in Appendix A. Note that Eq.(9.10) equals the free energy difference between structures M_p and I_p in Fig.9.9B.

We computed $\Delta G_L(n,f)$ at the average unzipping force $f \approx 19\,$pN at 500mM NaCl for L-segments in the range $L = [8,28]$, with $L = 8$ the minimum number of bases needed to form stem-loops. In Fig.9.10 we show the native free-energy landscape (FEL), $\Delta G_{\text{Native}}(n,f)$ (relative to the fully unzipped state) as a continuous black line. The contribution by the loop-BEL for $L = 20$ has been added to the native FEL (dashed line) to stress the fact that it kinetically traps off-pathway stem-loop structures at fixed n (red arrows). The dashed line for the loop-BEL emphasizes that this is a kinetic trapping landscape that does not describe transitions between contiguous n values. In Fig.9.11 (bottom) we show the loop-BEL $\Delta G_L(\lambda)$ (dashed black line) for $L = 20$ together with the experimental FDC (top). The position n along the sequence in Eq.(9.9) has

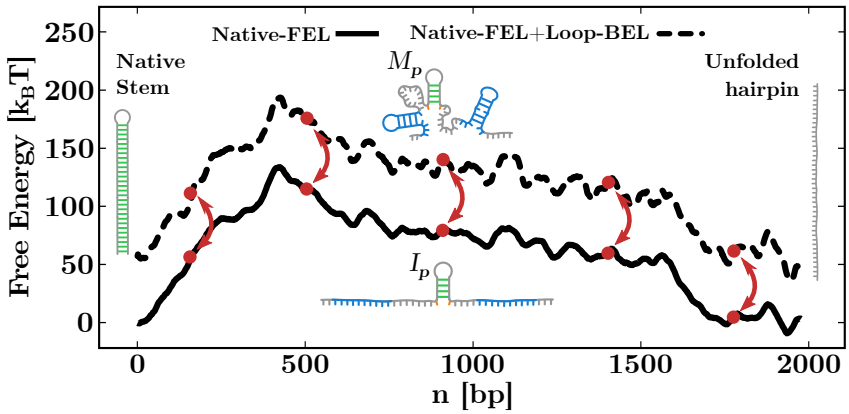


Figure 9.10: Loop-BEL (dashed line) computed with (9.9) for $L = 20$ added to the native-FEL of the hairpin (solid line). For a fixed n, the loop-BEL is the free-energy difference between structures I_p and M_p (equal to the vertical distance between red points). Red arrows depict the kinetic trapping effect induced by the loop-BEL.

been converted to trap-pipette distance λ by using the elastic parameters, $\Delta G_L(\lambda) \equiv \Delta G_L(n, f)$. The position of the loop-BEL minima shows a correlation with the FDC regions of large hysteresis (indicated by rectangles $R1 - R8$). To compare with the DNA case, we computed the loop-BEL for the DNA analogous 2027bp sequence (obtained by replacing uracils with thymines) at the predicted average unzipping force ($\sim$ 16.4pN) at 500mM NaCl [133]. Despite the profiles appearing to be similar, the average barrier energy in DNA ($\sim 47 k_BT$, solid gray line) is lower than in RNA ($\sim 57 k_BT$, solid black line) because of the lower DNA unzipping force (yielding a lower elastic contribution in Eq.(9.10)). We notice that the loop-BEL is overestimated as we have considered a restricted set (single stem-loops) among all possible competing structures: the lower the energy profile of the loop-BEL, the more stable the competing structures and the larger the irreversibility effects. The larger hysteresis in RNA correlates with the higher kinetic stability of the stem-loops for RNA.

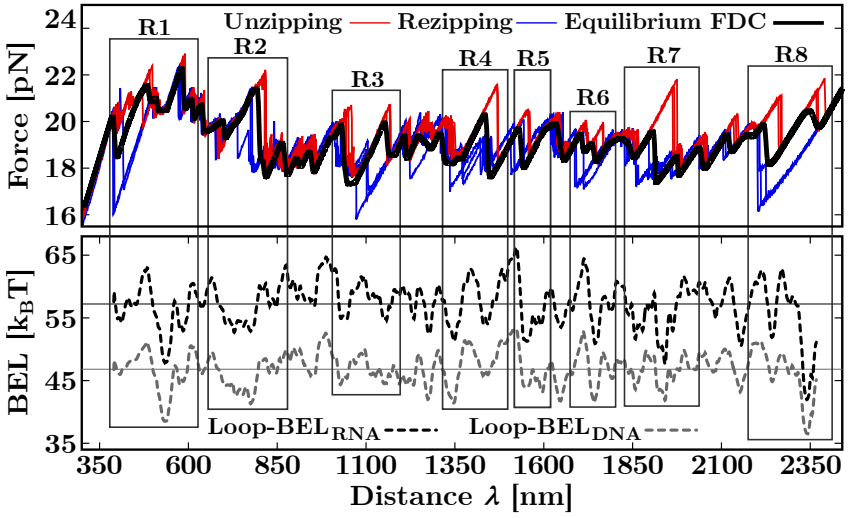


Figure 9.11: Experimental FDCs in 500mM NaCl (top) and loop-BELs at 19pN (bottom) computed for the RNA hairpin (dashed black line) and the equivalent DNA sequence (dashed gray line) for $L = 20$ bases. The mean values of the loop-BEL (solid lines) are also shown. Loop-BEL minima correlate with the hysteresis regions R1-R8.

9.2 LOOP-BEL AND HYSTERESIS

To quantify the correlation between the loop-BEL and the irreversibility, we introduced the following measure of the hysteresis: for a given position λ, the hysteresis is the dissipated work over a given distance $\Delta\lambda$ ($= 3$nm), which is

$$\Delta G^{\text{Hyst}}_{\alpha\beta}(\lambda) = -\int_{\lambda-\frac{\Delta\lambda}{2}}^{\lambda+\frac{\Delta\lambda}{2}} |f_\alpha(\lambda') - f_\beta(\lambda')|\, d\lambda', \tag{9.11}$$

where α, β denote the unfolding (U), refolding (R), and equilibrium (E) FDCs, leading to three distinct profiles $\Delta G^{\text{Hyst}}_{\alpha\beta}(\lambda)$ with $\alpha\beta = \text{UR}, \text{UE}, \text{ER}$. The minus sign in (9.11) has been introduced to positively correlate loop-BEL minima (maxima) with maximal (minimal) hysteresis. Equation

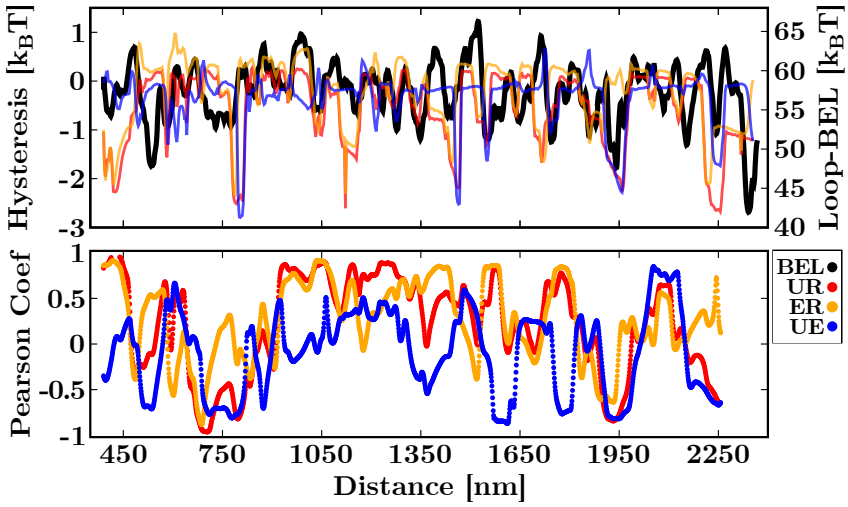


Figure 9.12: Pearson test between the loop-BEL and the hysteresis profiles. **(Top)** Loop-BEL $\Delta G_L(\lambda)$ computed for $L = 20$ (black line) and hysteresis profiles $\Delta G^{\text{Hyst}}_{\alpha\beta}(\lambda)$ with $\alpha\beta =$ UR (red), ER (orange), UE (blue) computed for the 500mM NaCl case. **(Bottom)** Pearson correlation coefficients $r_w(\lambda)$ resulting from the comparison between loop-BEL and the $\Delta G^{\text{Hyst}}_{\alpha\beta}(\lambda)$ over windows of length $w \approx 100$.

(9.11) has been averaged over several cycles and different molecules. Given the loop-BEL, $\Delta G_L(\lambda) \equiv \Delta G_L(n, f = 19\text{pN})$ in Eq.(9.9), and the hysteresis profile, $\Delta G^{\text{Hyst}}_{\alpha}(\lambda)$, we computed the Pearson correlation coefficient $r_w(\lambda) \in [-1, 1]$ over a given spatial window of size w as a function of λ. $r_w(\lambda) = 1$ ($r_w(\lambda) = -1$) indicates fully correlated (anticorrelated) landscapes in that region. Correlation profiles $r_w(\lambda)$ have been calculated for $\Delta G^{\text{Hyst}}_{\alpha\beta}(\lambda)$ with $\alpha\beta \equiv$ UR, UE, ER (Fig.9.12, top). To assess the correlation between the loop-BEL and the hysteresis profile $\alpha\beta$, we defined the average rolling correlation, $\langle r_w \rangle_{\alpha\beta}$, as the average taken over the entire landscape $r_w(\lambda)$. We used a sliding window of size $w \approx 100$nm, the result being insensitive to w as far as it is comparable to the typical number of bases released in a force rip along the FDC ($\sim 50 - 150$ bases) [75, 161].

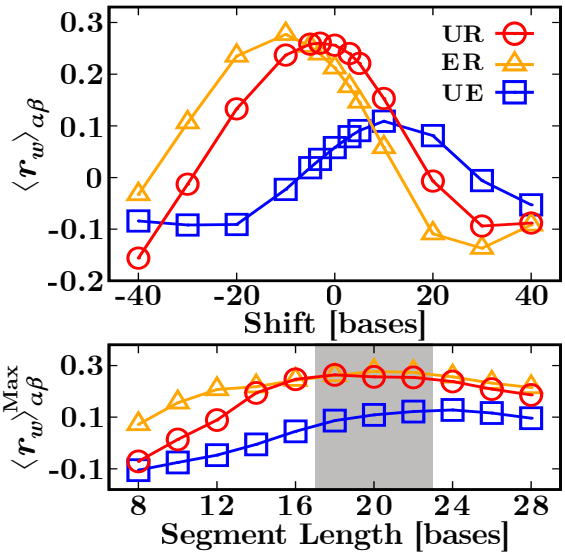


Figure 9.13: Average rolling correlation $\langle r_w \rangle_{\alpha\beta}$ as a function of the shift s between loop-BEL and hysteresis profile for $\alpha\beta \equiv \mathrm{UR}, \mathrm{ER}, \mathrm{UE}$ (top). Maximal average rolling correlation $\langle r_w \rangle^{\mathrm{Max}}_{\alpha\beta}$ (corresponding to $s \approx 0, -10, +10$ for UR, ER, and UE, respectively) for each L (bottom). Hysteresis is maximally correlated with stem-loops of length $L \sim [18, 22]$ bases.

In Fig.9.13 (top), we show $\langle r_w \rangle_{\alpha\beta}$ as a function of the shift s (in bases) of the loop-BEL relative to the hysteresis profiles. $\langle r_w \rangle_{\alpha\beta}$ has been calculated for the L-segment length $L = 20$ at which correlation is maximal (see below). A positive shift $s > 0$ means that we are testing the correlation with the loop-BEL in the rezipped region close to the junction (green bp in Fig.9.9B), whereas a negative shift $s < 0$ implies testing the correlation with the loop-BEL ahead of the junction in the unzipped region (grey and blue bp in Fig.9.9B). Remarkably, maximum correlation is found for $\alpha\beta \equiv \mathrm{UR}$ and $s = 0$ (red circles in Fig.9.13, top), showing that stem-loops formation and hysteresis are highly correlated precisely at the junction. The position of the maximum in $\langle r_w \rangle_{\alpha\beta}$ shifts to $s > 0$ ($s < 0$) for $\alpha\beta \equiv \mathrm{UE}$ (ER) (blue squares and orange triangles respectively, Fig.9.13, top). We

notice that for $\alpha\beta \equiv$ ER the maximum in $\langle r_w \rangle_{\mathrm{ER}}$ is shifted leftwards by $s \approx -10$ bases (orange triangles). Its value almost coincides with the $\alpha\beta \equiv$ UR case ($\langle r_w \rangle_{\mathrm{ER}}^{\mathrm{Max}} \sim \langle r_w \rangle_{\mathrm{UR}}^{\mathrm{Max}} \approx 0.25$, red circles). Therefore, the formation of stem-loops at a distance of ~ 10 bases in the unzipped region slows down the hairpin refolding, leading to the hysteresis observed during the rezipping process. In contrast, the maximum of $\langle r_w \rangle_{\mathrm{UE}}$ (blue squares) is shifted rightwards ($s \approx +10$) with $\langle r_w \rangle_{\mathrm{UE}}^{\mathrm{Max}} \approx 0.1 < \langle r_w \rangle_{\mathrm{UR}}^{\mathrm{Max}} \approx 0.25$ (red circles). The asymmetry between UE and ER demonstrates that the refolding process is the largest source of irreversibility in the unzipping-rezipping experiment. Analogously, the rightwards shift ($\sim +10$ bases) in $\langle r_w \rangle_{\mathrm{UE}}^{\mathrm{Max}}$ is related to breathing of stem-loops and the hysteresis effects observed in the unfolding FDCs. Finally, we analyzed the dependence of $\langle r_w \rangle_{\alpha\beta}^{\mathrm{Max}}$ with the length L of the segments forming the stem-loops (Fig.9.13, bottom). All curves show a broad maximum for $L \approx 18 - 22$, corresponding to the characteristic size of the stem-loops kinetically trapping the RNA. Several tests support the results on control sequences (see Appendix E)

9.3 SEQUENCE DEPENDENCY OF THE HYSTERESIS

The hysteresis is observed in some specific regions of the FDC but not in others (Figs.8.1, 8.2). To explain this behavior, we have searched for specific sequence motifs that promote stacking, hybridization, and stem-loop formation within the RNA single strands.

Here we studied the composition of all segments of length $N \geq 6$ bases containing consecutive purine and consecutive Watson-Crick complementary bases along the two unpaired strands of the RNA hairpin. The aim is to identify differences in purine content (for stacking) and Watson-Crick base pairs (for hybridization) between the two regions, demonstrating that irreversible effects are sequence-dependent. We do not discriminate

between purines G and A for the stacking motifs, being both counted in the same set. For example, irreversible region 1 in Fig.8.1 contains 220bp and a total of 3 segments of 6 consecutive purines (a single GGGGGG and GGAAAG on one strand and AGGGGA on the other strand) and 2 segments of 7 consecutive purines (AGGAGAA, AGAGAAA on one strand). An analogous count is made on segments capable of forming Watson-Crick complementary bases on the two strands by counting the number of segments containing consecutive A, U, or G, C. For example, the same irreversible region 1 has 1 segment of 9 consecutive G, C (CGCGGGGGG) and 1 segment of 10 consecutive G, C (CGCCGCCGCG).

Therefore, for each region, the fraction of bases of a given type (stacking or base-pairing) is defined as $f_N = (M_N N)/\Delta n$, where M_N is the total number of segments of length N and Δn is the total number of bases in that region. Given the values of f_N, for each N, we computed the average fraction of bases (stacking or base-pairing) over all the irreversible (reversible) regions

$$\overline{f_N} = \frac{1}{n_{\text{regions}}} \sum_{\text{regions}} f_N \,, \tag{9.12}$$

where n_{regions} is the number of irreversible (reversible) regions ($n_{\text{regions}} =$ 8 and 4, respectively). Finally, we defined the average segment length $\langle N \rangle$ of a given type (stacking or base-pairing) for the irreversible (reversible) regions as the weighted average over $\overline{f_N}$, which is

$$\langle N \rangle = \frac{\sum_N N \overline{f_N}}{\sum_N \overline{f_N}} \tag{9.13}$$

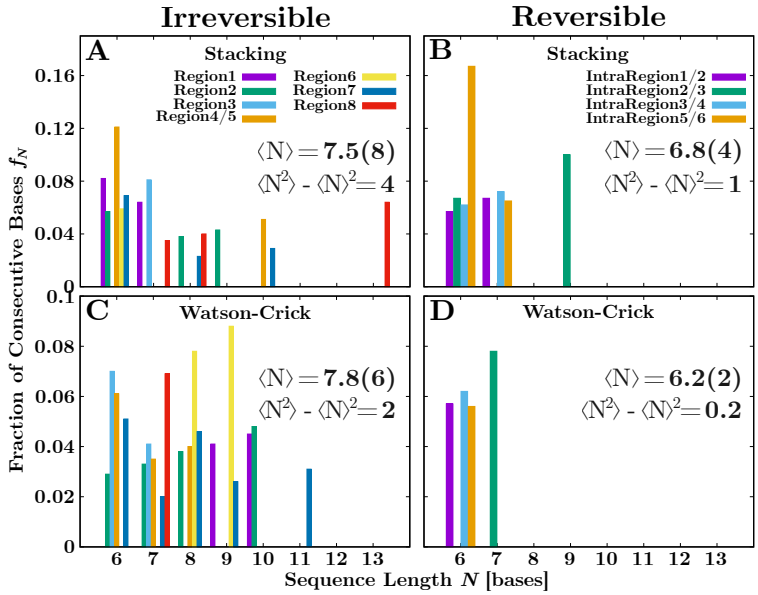


Figure 9.14: Sequence analysis of the irreversible and reversible regions. **(A,B)** Stacking analysis. Fraction of consecutive stacked purines (A, G) as a function of the segment length per each irreversible (left) and reversible (right) region. **(C,D)** Watson-Crick base pairing analysis. Fraction of consecutive (A, U or C, G) as a function of the segment length per each irreversible (left) and reversible (right) region. The analysis shows both a larger average segment length $\langle N \rangle$ and variance of the length of the segments $\langle N^2 \rangle - \langle N \rangle^2$ in the irreversible regions. The error (in brackets) is the statistical uncertainty in the last digit.

Analogously, the variance of the segment length for the irreversible (reversible) regions and the two types of analysis (stacking or base-pairing) has been computed as

$$\langle N^2 \rangle - \langle N \rangle^2 = \frac{\sum_N N^2 \overline{f_N}}{\sum_N \overline{f_N}} - \left(\frac{\sum_N N \overline{f_N}}{\sum_N \overline{f_N}} \right)^2 \tag{9.14}$$

The results are shown in Fig.9.14 and point out that stacking and base-pairing effects are larger in the irreversible regions than in the reversible ones. Overall, we found that stacking and base-pairing contribute to

the observed hysteresis facilitating the formation of stem-loop structures along the single strands.

9.4 THE LOOP-BEL MODEL APPLIED TO DNA

The formation of stem-loops structures may contribute to explaining the broad phenomenology of heterogeneous RNA folding, from misfolding and multiplicity of native states to the formation of complex tertiary structures. In particular, the loop-BEL model can be extended to the formation of non-specific secondary structures observed in pulling experiments of ssDNA and ssRNA. Although this phenomenon has not been investigated in RNA, it has been extensively studied in DNA [58, 60, 162–164]: upon stretching, the ssDNA elastic response deviates from the expected ideal behavior of a polymeric chain (described for example with the WLC) forming a shoulder below $f \sim 10 - 12$pN (see Fig.9.16, right).

To model this phenomenon, let us consider a ssDNA of N bases of a random sequence which, at difference with the previous case, cannot form a native hairpin (i.e. there is no hybridization junction). We consider the set of all possible excitations consisting of multiple stem-loops of a given length L along the sequence. The free energy of such a L-set of excitations equals

$$\Delta G_L(f) = -k_B T \log \sum_{k=0}^{K} \exp\left(-\frac{\Delta g_L(k,f)}{k_B T}\right), \tag{9.15}$$

where $\Delta g_L(k,f)$ is the total free-energy contribution of $k \geq 0$ stem-loops and $K = \lfloor N/L \rfloor$ is the maximum number of stem-loops that can form along the single strand. This is given by

$$\Delta g_L(k,f) = E_L(k) + \left[(N - kL)\Delta G^1_{ss}(f) + k\Delta G_d(f)\right]. \tag{9.16}$$

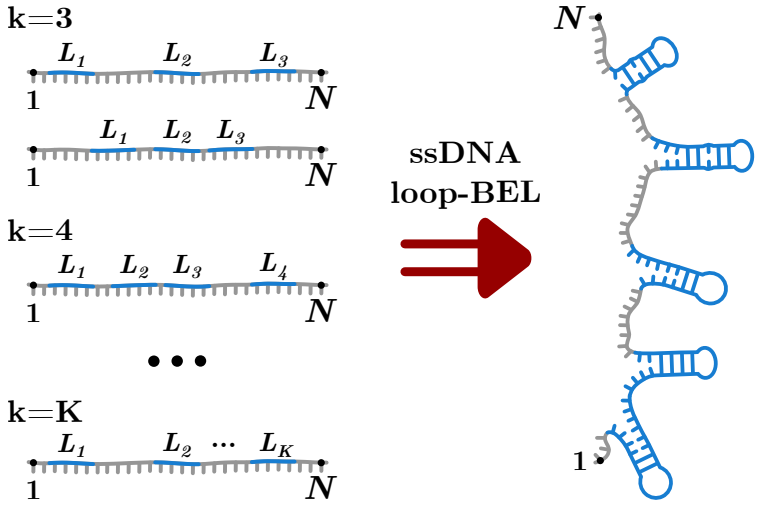


Figure 9.15: Loop-BEL model for ssDNA. All possible configurations of $k = 1, 2, \ldots, K$ stem-loops of length L randomly positioned along the sequence (left) compete for folding against the applied force f. The ssDNA loop-BEL defined in Eq.(9.15) gives the most energetically stable configuration (right).

The term $E_L(k)$ accounts for the most energetically stable configuration of k stem-loops randomly positioned along the sequence (Fig.9.15, left). The term $(N - kL)\Delta G^1_{ss}(f)$ is the energy gain upon stretching the free $N - kL$ bases at force f corrected by the (smaller) energy contribution, $k\Delta G_d(f)$, of orienting k stem-loops along the force axis. Notice that $(N - kL)\Delta G^1_{ss}(f)$ is an extensive quantity, equal to the number of monomers, $N - kL$, times the energy cost to stretch a single monomer, $\Delta G^1_{ss}(f) = -\int_0^f x^1_{ss}(f')df'$, where $x^1_{ss}(f)$ is the extension per monomer and has been modeled according to the WLC in Eq.(5.19). The same consideration holds for the dipole contribution of k stem-loops, $\Delta G_d(f) = -\int_0^f x_d(f')df'$, where $x_d(f)$ is the dipole extension and has been modeled according to Eq.(5.18).

An exact computation of $E_L(k)$ in Eq.(9.16) requires considering non-overlapping stem-loops: if a stem-loop of length L forms at position n

along the sequence, the next stem-loop can only form outside the interval $[n - L : n + L]$. We simplified this unaffordable mathematical task by considering overlapping stem-loops in a mean-field approximation. In this approximation, $E_L(k)$ is taken as the typical total energy of k stem-loops randomly chosen over the ensemble of C_k different realizations without imposing any constraints on these loops (i.e. they can be overlapping or non-overlapping). In contrast, the stretching contribution $(N - kL)\Delta G^1_{ss}(f) + k\Delta G_d(f)$, is taken independent of the k stem-loops realization. Therefore, we have,

$$E_L(k) \approx \min_{C_k} \left\{ \sum_k \epsilon^0_{L,k} \right\} - k_B T \log(C_k) \ , \tag{9.17}$$

where for the typical energy of k stem-loops ($\epsilon^0_{L,k} < 0, \forall k, L$), we took the most stable configuration (i.e. the one of lowest energy) within the ensemble C_k. The second term in the r.h.s. of Eq.(9.17) is an entropic contribution stabilizing stem-loops formation. The total number of configurations is given by the binomial coefficient, $C_k = \binom{K}{k}$, or the number of ways k objects (stem-loops) can be arranged into $K = \lfloor N/L \rfloor$ different positions. For large N, the total number of configurations in C_k is enormous. For example, the maximum number of configurations of a $N = 2000$bp sequence with $L = 20$ bases is $\binom{100}{50} \sim 10^{29}$. Thus, we restricted the sampling to a few hundred configurations (typically 500).

From Eqs.(9.15), (9.16), we can compute the average ssDNA extension for a given L, which is defined as

$$x_{ss,L}(f) = -\frac{\partial \Delta G_L(f)}{\partial f} = \frac{1}{Z_L} \sum_{k=0}^{K} x_L(k, f) \exp\left(-\frac{\Delta g_L(k, f)}{k_B T} \right) , \tag{9.18}$$

where Z_L is the system's partition function for a given L (c.f. Eq.(9.15)),

$$Z_L = \exp\left(-\frac{\Delta G_L(f)}{k_B T}\right) = \sum_{k=0}^{K} \exp\left(-\frac{\Delta g_L(k,f)}{k_B T}\right) \tag{9.19}$$

and $x_L(k,f)$ is the ssDNA extension when k stem-loops are formed,

$$x_L(k,f) = (N - kL)x_{ss}^1(f) + kx_d(f)\,. \tag{9.20}$$

Finally, the thermodynamic free energy and the ssDNA extension averaged over all L-segments are computed as

$$\Delta G(f) = -k_B T \log \sum_L \exp\left(-\frac{\Delta G_L(f)}{k_B T}\right)\,. \tag{9.21}$$

and

$$x_{ss}(f) = \frac{1}{Z}\sum_L x_{ss,L}(f) \exp\left(-\frac{\Delta G_L(f)}{k_B T}\right)\,, \tag{9.22}$$

where $Z = \sum_L Z_L$ is the system's partition function.

In Fig.9.16, we show the ssDNA extension predicted by the loop-BEL model for a random DNA sequence of $N = 2027$ bases at 10mM (left panel) and 1M (middle panel) NaCl salt concentrations. These results are compared with experimental data from pulling experiments of a long DNA hairpin at the same salt conditions (right panel). The expected ssDNA elastic response for $L \in [10, 100]$ computed with Eq.(9.18) (solid lines) is shown along with $x_{ss}(f)$ in Eq.(9.22) (dashed black line). The loop-BEL model reproduces the deviation from the ideal-WLC model (dashed gray line) experimentally observed below $\sim$ 10pN. This behavior results from the competition between the stem-loops of different sizes: the lower the force, the larger the contribution to Eq.(9.22) by larger stem-loops. Therefore, as the ssDNA approaches the random coil state ($f = 0$pN), the energetic gain to stretch large L-segments in Eq.(9.16)

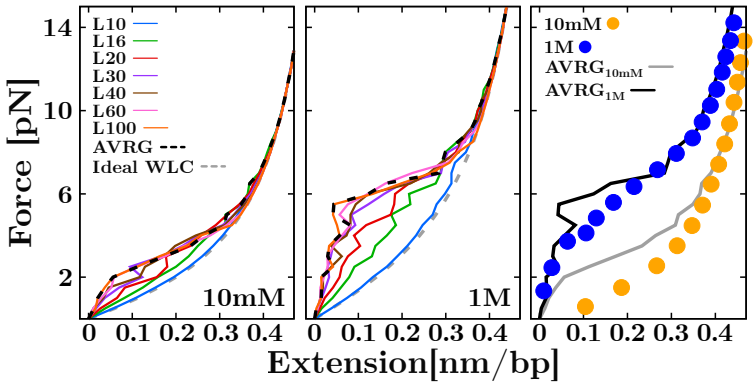


Figure 9.16: Loop-BEL prediction of ssDNA pulling experiments for different L-segments (solid lines) and by averaging over all L (dashed black lines) at 10mM and 1M NaCl (left and middle panels). The comparison between these predictions (solid lines) and experiments (solid dots) shows that the model reproduces the observed formation of (non-specific) secondary structure (right panel).

tends to zero while $E_L(k)$ remains constant, favoring the formation of large stem-loop structures. Remarkably, as the force approaches the dsDNA unzipping force ($f \approx 15$pN), stem-loops of length $L \sim 30 - 40$ bases become the most likely folds, as can be seen by comparing the average loop-BEL (dashed black line) to the loop-BEL at each L (solid lines) in Fig.9.16 A and B. This number is not far from what has been reported for RNA where $L \sim 20$ is the optimal stem-loop size (see above and Ref. [75]). At higher forces, the elastic response collapses to the WLC, as experimentally observed. Let us notice that both at 10mM and 1M NaCl, the predicted extension differs from the experimental data (solid dots in Fig.9.16, right) when $f < 6$pN. This is particularly evident at 10mM NaCl, where our model indicates that stem-loops still form at low force while no secondary structure is observed in the pulling trajectories. This and other potential inconsistencies come from the crude approximations made in Eq.(9.17). This approximation only holds when

$k \ll K$, i.e. when the typical distance between consecutive stem-loops is much larger than L, so the overlapping is negligible. Despite the simplicity of the mean-field approximation, the loop-BEL model helps to study the complex behaviors observed in NA. A more rigorous analytical treatment may lead to a deeper understanding of heterogeneous folding in NA.

9.5 CONCLUSIONS

We have shown that the large irreversibility observed between RNA unzipping and rezipping FDCs is driven by the collective effects of multiple stem-loop structures that kinetically trap the RNA. The effect is stronger in magnesium than in sodium, probably because the two charges of magnesium transiently stabilize nucleotide contacts to a higher extent. To model this mechanism, we used the free-energy landscape formalism to understand protein and NA folding into complex tertiary structures [63, 79]. By defining a stem-loops barrier energy landscape (loop-BEL, Fig.9.9B), we found a correlation between the sequence regions where stem-loops are maximally stable (minima of the loop-BEL) with those where hysteresis along the FDC is large. To support this interpretation, we measured the correlation between the loop-BEL and the hysteresis profiles (Eq.(9.11)). We have found that the hysteresis observed in the FDCs maximally correlates with the stem-loop formation at the hybridization junction. Typical stem-loop sizes of about 20 bases are responsible for the observed hysteresis effects. Interestingly, this number is similar to the number of foldon residues in protein folding [18]. We stress that the loop-BEL as a function of n is not a standard free-energy landscape as neither the trap-pipette distance λ nor n are actual reaction coordinates for the stem-loops. For a given n (λ), the loop-BEL is a kinetic trapping landscape that quantifies off-pathway (misfolded) configurations M_p that

compete with the folding intermediates I_p. Future work should lead to a better understanding of the stabilizing kinetics of these structures and the energy landscape describing transitions between them. We notice that along the reversible regions, the signal-to-noise ratio is very low due to instrumental drift and noise effects, which are detrimental in evaluating the correlation between sequence and hysteresis.

Remarkably, hysteresis is observed in some specific FDC regions but not in others. To explain this, we have searched for specific sequence motifs that promote stacking, hybridization, and stem-loop formation within the RNA single strands (Sec.9.3). We found that the irreversible regions are characterized by a high frequency of purine stacks and Watson-Crick bonds along the unpaired strands, which lead to the multiple peaks observed in the experimental FDCs, even for forces as high as 20pN. Stacking alone could not transiently stabilize stem-loops at such high forces, so the concurrent formation of base pairs within each RNA strand is necessary. It is quite reasonable that such stem-loop structures also exhibit some degree of cooperativity; the more they proliferate, the more they facilitate the formation of additional nearby stem-loops inhibiting native folding. Cooperative folding effects have been also found in DNA [60, 165], RNA [147, 166, 167] and proteins [76, 168, 169]. The intermediates I_p in the unzipping-rezipping experiments are reminiscent of the cooperative foldons hypothesized to drive protein folding [18, 149]. This cannot be otherwise, as the only way to form the native stem is to sequentially form the intermediates, one after the other, starting from the unfolded state. The remarkable effect of force is to increase the lifetime of the intermediates that would be difficult to detect in melting experiments.

Moreover, the loop-BEL model allows for predicting the formation of non-specific secondary structure in ssDNA pulling experiments [60]. This feature appears to be caused by the competition between multiple

stem-loop folds of different sizes that make the extension deviate from the ideal WLC behavior (Fig.9.16).

Stem-loops formation appears as a general mechanism driving RNA hairpins folding and refolding processes and the elastic response of ssDNA sequences. Moreover, it may help to understand the broad phenomenology shown by NAs. However, the predictions obtained by the present model are limited by crude approximations introduced to simplify the enormous complexity of complete stem-loops modeling. Firstly, in the computation of the stem-loops free-energy (Eqs. (9.16) and (9.17)), we disregarded overlapping effects between consecutive L-segments. This implies that partially overlapping L-segments can simultaneously fold into stem-loops. Moreover, we neglected cooperativity effects that favor the nucleation of contiguous stem-loops, a phenomenon reminiscent of domain coarsening in helix-coil models. Ultimately, full loop-BEL modeling would require allowing for the simultaneous formation of stem-loops of different lengths L. In contrast, the current model accounts for this effect through a mean-field approximation (Eqs. (9.21) and (9.22)). Despite the lack of comprehensive modeling of stem-loops formation, the phenomenon might explain many features of heterogeneous NA folding. The development of a more accurate description of the loop-BEL model accounting for the complex phenomenology discussed above is left for future work.

Part IV

RNA COLD MISFOLDING

10

RNA AT LOW TEMPERATURE

Natural evolution has produced living organisms of utmost complexity. It is acknowledged that biological systems evolve under the rules of Darwinian selection, led by the survival of the fittest to environmental pressure. Free energy governs the spontaneous evolution of all thermodynamic transformations determining their stability. The change in the environmental conditions at which a biological process occurs may alter the existing equilibrium promoting the stabilization of previously penalized structures [170]. In recent years, evidence has emerged that RNAs can fold into multiple conformations at sufficiently low temperatures triggering a new phenomenology that we call *RNA cold misfolding*, the tendency of native RNAs to form mechanically compact and kinetically stable disordered structures [75, 147].

RNAs directly impact biological diversity and life [171, 172]. The promiscuity of base pairing and stacking interactions makes RNA a unique biopolymer with many functions, from information carrier to regulatory and enzymatic activity. As a result, RNAs can fold into multiple configurations stabilized by secondary and tertiary structures [173, 174], multivalent cations, and ligands [175–178]. RNA exhibits significant heterogeneity at the sequence and conformational level [179–182]. Upon folding, RNA can form native and non-native structures (such as misfolded and intermediates) [147], with critical roles at the level of genomic

maintenance and the cellular function [183, 184], therapeutics [185, 186], and diseases [31–33]. Although new RNAs with new functionalities and structures are being discovered, the role of many RNAs remains unknown. Besides the much-studied tRNA, rRNA, microRNA, riboswitches, ribozymes, and artificially evolved RNAs, novel behaviors have been observed in response to environmental cues such as temperature (e.g., RNA cold denaturation [187, 188], RNA thermometers [124, 189]), and in concerted action with proteins (catalytic complexes, chaperones, packaging, condensation, etc.). Moreover, a knowledge gap has appeared not only at the level of RNA transcriptomics but also at the level of non-coding RNAs (ncRNAs), and their remarkable variety of functions in concert with ligands and proteins [184, 190–194].

Despite the enormous progress in next-generation sequencing and big data analysis, our current knowledge of RNA diversity is compromised by the limited accuracy, sensitivity, and specificity of available methods to detect different RNA conformations across RNA populations. Moreover, determining the folding pathways and the energetics of the various RNA structures is essential to understanding RNA function. Single-molecule techniques have represented a big step in addressing RNA complexity [37, 195]. Their great sensitivity and accuracy permit us to detect and measure the folding energies of rarely occurring conformations that escape detection by the standard bulk methods. Powerful techniques such as single-molecule FRET [196, 197] and force spectroscopy [168, 198] can monitor RNA conformational transitions in real-time. More recently, solid-state nanopore microscopy for RNA target detection can analyze thousands of single RNAs without amplification offering exciting prospects [199, 200].

We studied short RNA duplexes at low temperatures by calorimetric unzipping experiments with a temperature-jump LOT. These experiments showed that RNA hairpins at low temperatures systematically misfold.

The stability of these structures increases upon lowering the temperature and overtakes the native state at $T \sim 7°C$. We characterized the change with the temperature of the free energy of formation and the elastic properties of the native and misfolded states. These results point out that the misfolded state has a more compact conformation suggesting that the molecule folds into a complex tertiary structure. RNA misfolding is not predicted by current secondary structure prediction folding models such as Mfold [24], Vienna package [201], McGenus [202] and pKiss [203]. RNAs' ability to adapt to changing environments may have consequences for the reservoir of microbial life in the arctic soil and the permafrost where ancient frozen RNAs that have remained intact and protected for eons may come to life due to climate change [204, 205].

10.1 UNZIPPING OF SHORT RNA HAIRPINS

We used a temperature-jump LOT (see Sec.3.2.2) to unzip six fully-complementary Watson-Crick RNA hairpins featuring two 20bp stem sequences (H1 and H2) and loops of different lengths ($L = 4, 8, 10, 12$ nucleotides) and compositions (poly-A or poly-U). For the pulling experiments, the RNA hairpin is flanked by long hybrid DNA/RNA handles ($\sim$ 500bp). Further details about the sequences and the synthesis protocol are given in Appendix D.2. Each molecule has been named after its stem and loop type. For example, hairpin H1L4A is made by hairpin $H1$ and a poly-A loop of 4 bases. The unzipping experiments were carried out at temperatures in the range $42 - 7°C$ at 4mM $MgCl_2$ (divalent salt) and 1M NaCl (monovalent salt). Both buffers have been prepared by adding the salt (divalent or monovalent) to a solution of 100mM Tris-HCl (pH 8.1), 1mM EDTA. The pulling protocols have been carried out at a constant pulling speed, $v = 100$nm/s. We sampled $5 - 6$ different

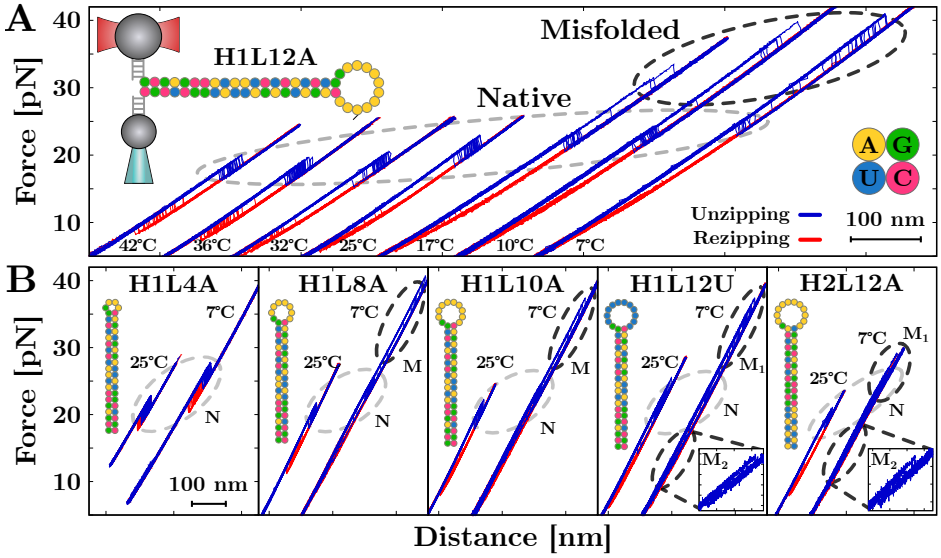


Figure 10.1: Unzipping of different RNA hairpins at different temperatures. Unfolding traces are shown in blue, and refolding traces are in red. (**A**) The H1L12A hairpin in the range $[7, 42]^\circ$C. At $T \geq 25^\circ$C, only the native state is present. Below 25°C, a misfolded structure appears with an increasing frequency upon lowering T. At 7°C, the misfolded is predominant on the native structure. (**B**) FDCs at 25°C and 7°C for different molecules. Hairpins H1L4A, H1L8A, and H1L10A show the effect of increasing the loop size ($L = 4, 8, 10$ bases, respectively) on misfolding. Hairpin H1L12U features stem H1 and the 12 bases poly-U loop, while hairpin H2L12A features stem H2 and the 12 bases poly-A loop. All hairpins exhibit a misfolded structure at low temperatures competing with the native state, except H1L4A. Molecules H1L12U and H2L12A also show a second misfolded state at low forces (insets). The different hairpin sequences are shown in the corresponding panels.

molecules at least for each hairpin and at each temperature, collecting ~ 200 unfolding-folding trajectories per molecule.

Fig.10.1A shows the temperature-dependence of the force-distance curves (FDCs) for the H1L12A sequence at 4mM magnesium. Above room temperature ($T = 25^\circ$C), the hairpin unfolds around 20pN (blue force rips): the rupture force distribution is unimodal. However, upon

lowering the temperature, for $T \leq 17°C$, high-force unfolding events are observed ($\sim 30 - 40$pN, dashed black ellipse) that are above the native unfolding events ($\sim 20 - 25$pN, dashed grey ellipse). The latter events reveal the formation of misfolded structures different from the native. Analogous results have been obtained for the sodium experiments. The misfolded state prevails over the native below $T = 10°C$ with occupancy larger than 50%. Moreover, refolding forces (red curves in Fig.10.1A) decrease with temperature with larger hysteresis at low T, a signature of the increasing competition between native and misfolded structures. This is in contrast to DNA unzipping, where no misfolding is observed and refolding forces increase with lowering T [206, 207].

Remarkably, RNA cold misfolding is not unique to the H1L12A hairpin but appears to be a general phenomenon for RNA. Our results show that it strongly depends on the molecular length and composition (Fig.10.1B). Below $T = 25°C$, nearly all the studied sequences feature misfolding. The only exception is the H1L4A hairpin (loop of $L = 4$ bases) that only folds into the native state, independently of temperature. Misfolding appears for $L > 8$ bases (H1L8A hairpin), and its occurrence increases with increasing L. At 7°C, the H1L12A hairpin is the only case of misfolding prevailing over the native state.

Loop composition also affects misfolding: the unzipping of hairpin H1L12U, featuring the same stem as H1L12A but with a 12 bases poly-U loop, shows two misfolded structures forming at 7°C (see inset 10.1B). Finally, stem sequence modulates misfolding, as we observe H2L12A to feature two distinct misfolded states at low temperatures. Although the appearance of misfolding does not primarily depend on the RNA sequence, the exact feature of each misfolded state does. For instance, the misfolded structures formed by H1L12U are different from those formed by H2L12A.

To better understand the molecular characteristics that promote and modulate the low-T misfolding, we first studied the temperature dependence of the ssRNA elastic response, focusing on the effects of stacking interactions and secondary structure formation.

10.2 TEMPERATURE-DEPENDENCE OF RNA ELASTIC PROPERTIES

In the unfolded state, the RNA molecule is a stretched polymeric chain of monomers (the nucleotides). The mechanical properties of the single-stranded RNA (ssRNA) drive the folding dynamics as they play a crucial role in determining the specific and non-specific secondary structure of the molecule upon refolding. In Fig.10.2, we studied the ssRNA elastic properties of the H1L12A hairpin in the temperature range $[7, 42]^{\circ}$C.

We used two different and independent methods to measure the experimental RNA force-extension curves (FECs): the force-jump and the two-branches method [108, 208] (triangles and circles in Fig. 10.2A, respectively – see Appendix F). The results of both methods have been put together for an accurate derivation of the ssRNA elastic properties at each T. Notice that the force range accessible for studying the ssRNA increases with lowering T. This is a consequence of the increasing hysteresis (Fig.10.1A) between unzipping and rezipping FDCs: the larger the hysteresis is, the longer the ssRNA unfolded branch (red traces on Fig.10.1) extends to low forces.

The RNA elastic response has been modeled according to the worm-like chain (WLC) model (see Eq.5.19 in Sec.5.3). To accurately derive the ssRNA WLC elastic parameters (persistence length, l_p, and interphosphate distance, d_b) at all temperatures, we first fit the FECs at each T to a WLC independently (red lines in Fig.10.2A). This analysis revealed a linear dependency of l_p and d_b with T. The measured elastic parameters at each T are shown in Fig.10.3 (red dots) and are reported in Table 10.1.

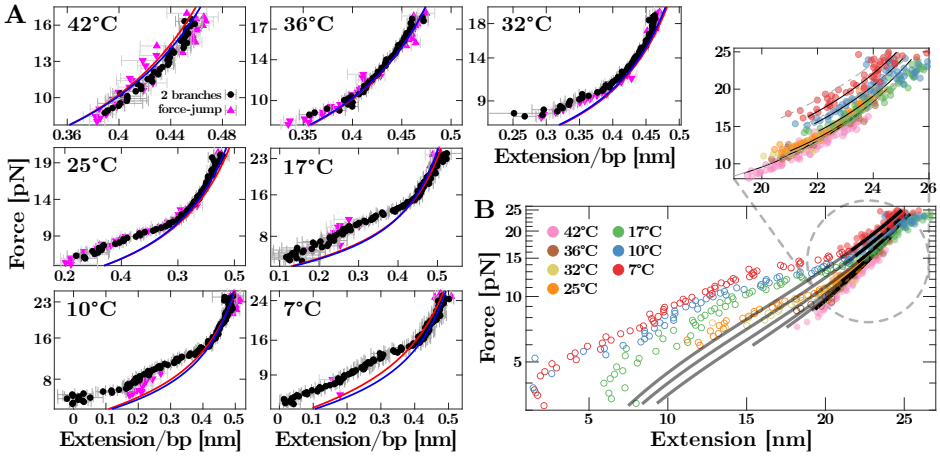


Figure 10.2: Temperature dependence of the H1L12A hairpin elastic properties. (**A**) ssRNA elastic response at different T measured with the two branches (magenta triangles) and the force-jump (black circles) methods (see text). The WLC ideal elastic response is also shown for both the single-T fit (red line) and the multi-T fit (blue line). Notice that the accessible force range increases with lowering T as the rezipping force also decreases (larger irreversibility). (**B**) The multi-T fit of the WLC to the experimental FECs. The data featuring non-specific secondary structure (empty dots) have not been considered for the WLC fit (grey lines).

Thus, by imposing the linear T dependence of the elastic parameters, we do a simultaneous fit of the FECs at all temperatures, which allows for the accurate measurement of the T dependence of the elastic parameters (black lines in Fig.10.3).

The results of this multi-T fit are shown in Fig.10.2A (blue lines) and Fig.10.2B. Excepted at 42°C, the FECs exhibit an increasing deviation from the ideal WLC behavior at low forces. This is due to the temperature-dependent staking interactions along the ssRNA, a phenomenon already observed in pulling experiments of ssDNA and ssRNA [58, 60, 162, 209]. Therefore, only the ssRNA high-force behavior has been fitted (solid dots and solid lines in Fig.10.2B) down to the point at which this deviation starts (empty dots and grey lines in Fig.10.2B).

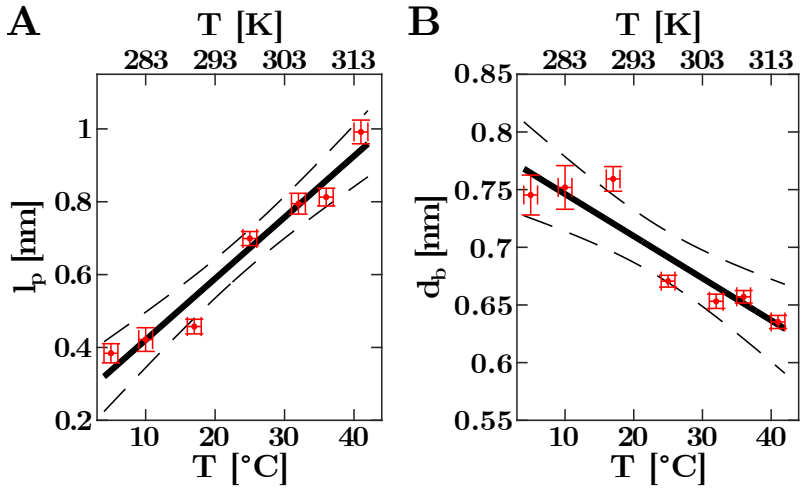


Figure 10.3: Temperature dependence of the persistence length, l_p, (left), and the interphosphate distance, d_b, (right) in Celsius (bottom label) and Kelvin (top label) degrees. The black line represents a linear fit to the data points; the dashed lines mark the prediction error.

The persistence length, l_p, (Fig.10.3A) exhibits a strong linear dependence with T, increasing by ~ 2.5 times over the studied temperature range. A fit-to-data gives a slope of $\partial l_p/\partial T = 0.17 \pm 0.02$ Å/K. In contrast, the interphosphate distance, d_b, (Fig.10.3B) decreases with T with a linear dependence of slope $\partial d_b/\partial T = -0.04 \pm 0.01$ Å/K. Notice that the T-dependence of d_b (changing $\sim -20\%$ over the explored T range) is much weaker than the l_p one ($\sim +250\%$). As a consequence, the ssRNA becomes stiffer with decreasing T. Thus, it progressively becomes more and more difficult for the ssRNA to bend on itself upon refolding and aligning the complementary ssRNA unpaired strands for hybridization. This hinders the refolding of the molecule into the native state, promoting mismatching of the ssRNA strands and the formation of misfolded structures.

T [°C]	T [K]	l_p [nm]	d_b [nm]
42	315	0.99 (3)	0.64 (1)
36	309	0.81 (2)	0.66 (1)
32	305	0.79 (3)	0.65 (1)
25	298	0.70 (2)	0.67 (1)
17	290	0.46 (2)	0.76 (1)
10	283	0.42 (3)	0.75 (2)
7	280	0.38 (3)	0.75 (2)

Table 10.1: Temperature dependence of the persistence length (l_p) and the interphosphate distance (d_b) measured from the ssRNA elastic response of the H1L12A hairpin. The error (in brackets) refers to the last digit.

These effects are sequence-dependent, as we deduce by studying the elastic properties of the other RNA sequences (Fig.10.4). In fact, while changing the loop size does not significantly affect the ssRNA behavior, a different loop sequence (H1L12U hairpin) or stem (H2L12A hairpin) produces a different elastic response. In the first case, the poly-U loop features weaker stacking interactions than its poly-A counterpart, allowing the ssRNA to elongate more under an applied force. In contrast, the H2L12A stem features stronger non-specific secondary structure interactions making the ssRNA stiffer and less stretchable.

10.3 LOW-TEMPERATURE MISFOLDING

The FDC rupture force upon unfolding and the corresponding released extension characterize a molecular state (native or misfolded) as they depend on the folded configuration's secondary structure and its free energy of formation. Accurate characterization of the energetics and kinetics of molecular folding is crucial to study the low-T misfolding, Therefore, we first separated the experimental trajectories generated by unzipping from the native state from those due to unzipping from

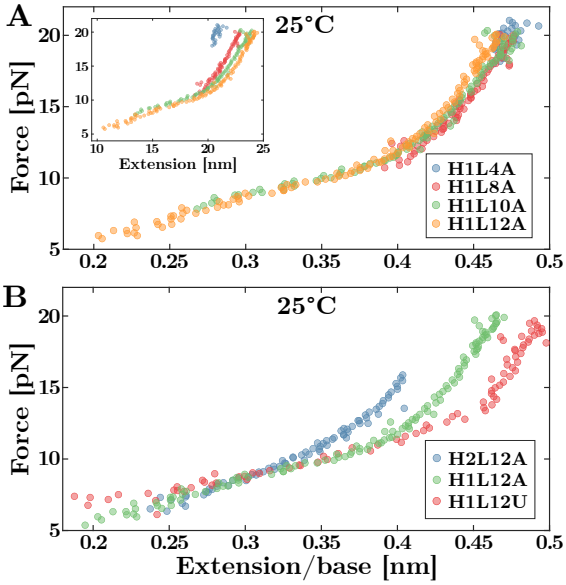


Figure 10.4: Comparison of the hairpins ssRNA FECs (normalized over the total number of bases of each hairpin) at $T = 25°C$. (**A**) ssRNA FECs of the H1 hairpin family. Despite the extension being dependent on the total number of bases (inset), the normalized extension (main) is the same for all hairpins. (**B**) ssRNA FECs for the hairpins with different sequences. The normalized extensions do not collapse as different sequences have diverse stacking interactions: the molecule with the strongest stacking (H2L12A) is the shortest, and that with the less stacking (H1L12U) is the longest.

the misfolded state. To do this, we developed a Bayesian classification algorithm using the rupture force and the released extension to assess native and misfolded FDCs (see Appendix G). The H1L12A hairpin results are shown in Fig.10.5A.

Misfolding (red circles) appears below 25°C and becomes more frequent (top panels in Fig.10.5A) by lowering the temperature. Finally, at $T \sim 7°C$, misfolding is predominant on the native, featuring more than 50% of the total number of trajectories. Moreover, the misfolded state unfolds in a broader range of forces than the native. It is characterized by

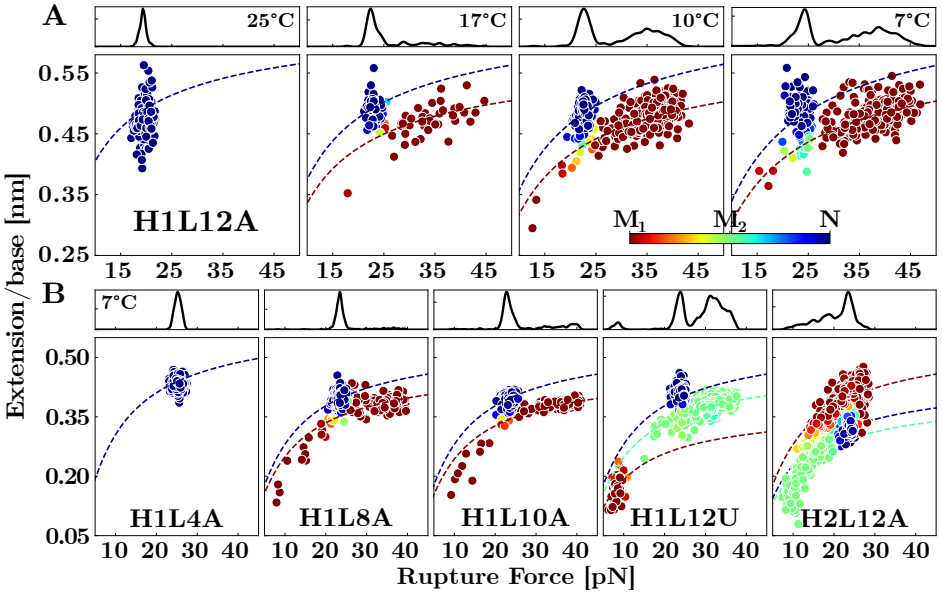


Figure 10.5: Misfolding characterization for the six studied hairpins. (**A**) Bayesian clustering applied to the H1L12A data: misfolded (red points) and native (blue points) states below 25°C. Misfolding increases upon lowering T. The predicted elastic response corresponding to each state is also shown (dashed lines). The rupture forces distributions (top panels) show the increasing misfolded/native ratio. (**B**) Bayesian clustering for the different hairpins at $T = 7$°C. H1L4A, H1L8A, and H1L10A show the misfolding dependence with the loop size. H1L12U and H2L12A exhibit two different misfolded states, M_1 and M_2. The dashed lines show the expected elastic response for each state.

a shorter released extension, pointing out that the number of hybridized base pairs of the misfolded structure is smaller than in the native. In contrast, the native state (blue circles) features a narrow rupture force distribution whose average value moves from $\sim$ 20pN to $\sim$ 25pN with decreasing T.

Low-T misfolding exhibits similar behavior for all temperatures and sequence composition (Fig.10.5B). Although the loop size modulates the percentage of misfolding (compare hairpins from H1L4A to H1L12A),

it always unfolds in the same force range with a characteristic released extension. Moreover, the clustering shows the sequence dependency of the number of misfolded states, with the H1L12U and the H2L12A hairpins featuring an additional low-force misfolded state (M_2). Let us remark that, despite both featuring similar unfolding distributions, different stem sequences have different misfolded structures, i.e. M_2 state of the H1L12U hairpin is different from the M_2 state of the H2L12A hairpin.

In what follows, we used the results from the Bayesian clustering to study the temperature dependence of the folding free energy of the native and misfolded states for the H1L12A hairpin.

10.3.1 *Derivation of Folding Free Energies*

The stability of a molecular state (native or misfolded) is measured by its free energy of formation, $\Delta G_0 = G_U - G_F$, which is the (positive) free energy difference between the unfolded and the folded configurations of the molecule when no external force is applied ($f = 0$). We used the fluctuation theorem (FT) [210, 211] to extract ΔG_0 from irreversible work (W) measurements (see Chapter 4). The work is computed by integrating the FDCs between a minimum (λ_{min}) and a maximum (λ_{max}) optical-trap positions where the molecule is folded and unfolded, respectively (inset in Fig.10.6A). The crossing point between the forward (F) work distribution, $P_F(W)$, and the reverse (R) work distribution, $P_R(-W)$, gives the free energy of formation, ΔG. Finally, ΔG_0 is obtained by subtracting to ΔG the ssRNA elastic contribution, which is the energy needed to stretch the polymer between λ_{min} and λ_{max}.

In Fig.10.6 we show $P_F(W)$ ($P_R(-W)$) for the H1L12A hairpin at the different experimental temperatures. At high-T (Fig.10.6A), the work distributions progressively separate with lowering T due to the increasing

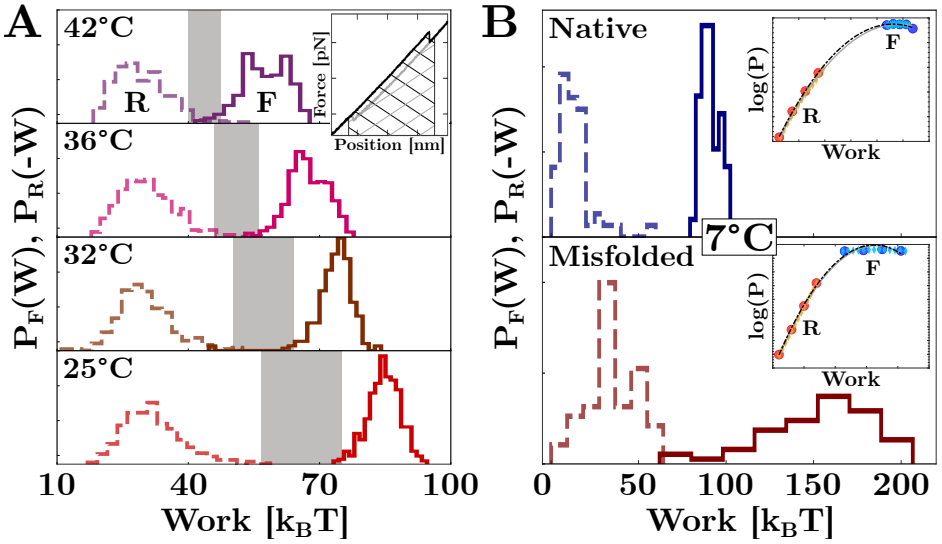


Figure 10.6: Study of the H1L12A hairpin. (**A**) Forward (dashed line) and reverse (solid line) work distributions computed from the experimental FDC (inset) above 25°C. The free energy of formation (see text) increases (gray band) with decreasing T. Notice that also irreversibility increases, i.e. $P_F(W)$ and $P_F(-W)$ get further apart. The results include work measurements of several H1L12A molecules ($\sim 5-6$ molecules at each T). (**B**) $P_F(W)$ (dashed line) and $P_R(-W)$ (solid line) for the native (blue) and misfolded (red) state at 7°C. The insets show the results of the matching method applied to native and misfolded states (see text).

hysteresis between the unzipping and rezipping FDC (gray bands). As a consequence, $P_F(W)$ and $P_R(-W)$ do not cross. In the absence of crossing, we used the matching method [63, 210], the Bennett acceptance ratio (BAR) method [83], and the continuous effective barrier analysis (CEBA) [116, 117, 119] to measure the free energy, ΔG (see Sec.4.2 and Sec.6.2). The large molecule-to-molecule statistical variability and the hysteresis made it essential to cross-validate results by different methods to extract reliable free energy values. At 7°C (Fig.10.6B), the Bayesian clustering has been used to assess the FDCs to the correspondent state (native or misfolded) before computing the forward and reverse work

distributions. In analogy with the high-T regime, the native state (in blue) features skewed work distributions, although the large hysteresis and the low refolding forces make $P_{\rm R}(-W)$ asymmetric. In contrast, misfolding (red) exhibits spread work distributions suggesting a large variability in the free energy of the formed structures. In the matching method, the crossing point (and then ΔG) between the work distributions is determined by matching $P_{\rm F}(W)$ and $P_{\rm R}(-W)\exp\left((W-\Delta G)/k_{\rm B}T\right)$. In practice, this is done by fitting the latter functions to a single Gaussian distribution [63] (dashed black line in Fig.10.6, insets). At 7°C, this method can be successfully applied to the FDCs of the native state but not to the FDCs of the misfolded state as the work distributions cannot be matched to a single Gaussian distribution (inset of Fig.10.6B, bottom). This can be explained by considering misfolding due to different structures with similar features. As a single function cannot characterize more the one structure, the matching method fails.

Given ΔG and knowing the ssRNA elastic response with T, we measured the free energy of formation (ΔG_0) for the H1L12A at all T for native and misfolded states (Fig.10.7). Above 25°C, the native state (right panel of Fig.10.7 and Table 10.2) becomes more stable with decreasing T as the base paring interactions become stronger. Our results (red and blue dots) agree with the predictions by Mfold [24] (black triangles) only at 36°C, whereas they exhibit a progressive deviation from the expected behavior at higher (lower) T This difference can be quantitatively assessed by measuring the enthalpy (ΔH_0) and entropy (ΔS_0) of the native state. By fitting the T-dependence of ΔG_0 according to the linear relation $\Delta G_0 = \Delta H_0 + T\Delta S_0$, we derived $\Delta H_0^{\rm BAR} = 105 \pm 33 {\rm kcal\,mol^{-1}}$, $\Delta S_0^{\rm BAR} = 238 \pm 11 {\rm cal\,mol^{-1}K^{-1}}$ (for the BAR results) and $\Delta H_0^{\rm CEBA} = 104 \pm 25 {\rm kcal\,mol^{-1}}$, $\Delta S_0^{\rm CEBA} = 231 \pm 8 {\rm cal\,mol^{-1}K^{-1}}$ (for the CEBA results). In contrast, we get $\Delta H_0^{\rm Mfold} = 196 {\rm kcal\,mol^{-1}}$, $\Delta S_0^{\rm Mfold} = 533 {\rm cal\,mol^{-1}K^{-1}}$ for the Mfold values, which almost double

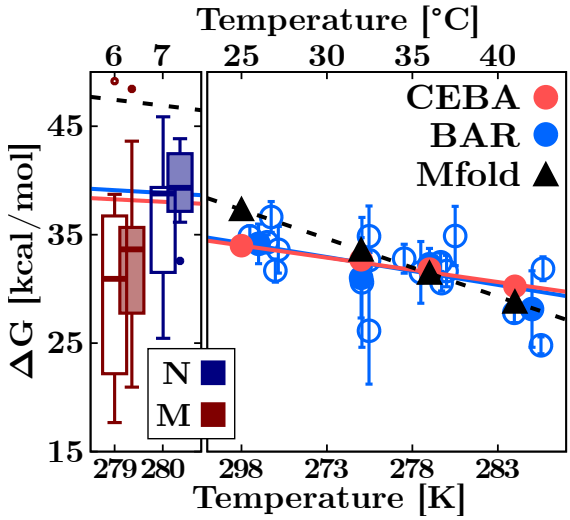


Figure 10.7: Temperature dependence of the folding free energy in K (bottom label) and °C (top label). **Left.** Native (blue) and misfolded (red) free energies at low T. The large dispersion of the results indicates that there are many different misfolded structures. Results in magnesium (empty bars) and in sodium (solid bars) are plotted together. For the comparison, the 100/1 equivalence rule between monovalent and divalent salt concentrations (see Sec.8.4) has been applied to the sodium results. **Right.** At high T, only the native state is present. Results obtained with BAR (red) and CEBA (blue) methods are compared with the prediction by Mfold (black). The experimental variability (empty dots) is also shown along with the average values (solid dots) for the BAR case.

the experimentally measured enthalpy and entropy. The current characterization of the RNA energetics at the single base pair level does not account for the temperature dependence of entropies and enthalpies upon forming the stem. As shown by calorimetric experiments, this effect is not negligible [124,212–214] leading to large deviations of the predicted RNA hybridization free energies from the experimental values measured far from the physiological (reference) temperature of 37°C.

At low-T (left panel of Fig.10.7 and Table 10.3), we reported the hybridization free energies measured from magnesium (empty boxes) and

		ΔG_0 [kcal/mol]		
T [°C]	**T [K]**	**BAR**	**CEBA**	**Mfold**
42	315	28 (3)	30 (3)	28
36	309	32 (1)	33 (2)	31
32	305	31 (4)	31 (4)	34
25	298	34 (2)	35 (1)	37

Table 10.2: Free energies of the H1L12A hairpin native state in the high-T regime measured with BAR and CEBA methods, and predicted by Mfold. The error (in brackets) refers to the last digit.

sodium (full boxes) experiments. A divalent salt increases the instability of the pulling experiments, lowering the sampled statistics per molecule and making results in sodium more accurate. Remarkably, results for both salt conditions show that misfolding (red) has a much larger variability than the native (blue). These results support our hypothesis of misfolding being due to the formation of different structures of similar size but involving various secondary and tertiary folding.

Moreover, misfolded states appear to be less stable than the native (lower free energy), which points to the kinetic nature of misfolding. However, even at 7°C, the native state is more stable than the misfolded one, implying that the hairpin should mostly fold into the native. Therefore, kinetic effects contribute to the formation of the misfolded structures as they overtake the native at low T. We note that above 7°C (at 10°C and 17°C), the sparse misfolding statistics with respect to the native do not allow for a reliable sampling of the work distributions and, ultimately, for the extraction of ΔG_0.

			ΔG_0 [kcal/mol]	
T [°C]	**T [K]**	**State**	**Mg^{2+}**	**Na^{+}**
7	280	N	38 (9)	37 (3)
		M	31 (10)	31 (8)

Table 10.3: Free energies of the H1L12A hairpin native (N) and misfolded (M) states at $T = 7°C$ (derived with the BAR method). To be compared, the 100/1 equivalence rule between monovalent and divalent salt concentrations has been applied to the sodium results. The error (in brackets) refers to the last digit.

10.4 UNIVERSALITY OF THE RNA COLD MISFOLDING

Cold misfolding appears as a general phenomenon involving the formation of one or more multiple structures that compete with the native fold. On the one hand, the appearance of these structures does not appear to be related to the salt type (monovalent or divalent salt). On the other hand, misfolding is sequence-dependent, meaning that different RNA molecules fold into misfolded structures with different features. However, misfolding shows common traits for all the studied hairpins.

The unfolding (folding) transition can be described as an activated process along the molecular free energy landscape (FEL) [79, 215, 216]. By varying the reaction coordinate (the end-to-end molecular extension), the system switches from the folded (F) to the unfolded (U) state by crossing a transition state (TS) of height B. The folded structure and its behavior upon stretching are regulated by the force-dependent position of the TS ($x^{\ddagger}$) mediating the transition [217–219]. We used the Bell-Evans (BE) model [114, 115] to extract information about the FEL from the unzipping data. The BE model relates the standard deviation of the rupture forces distribution, σ_f, and the distance to the kinetic barrier, $x^{\ddagger}$ (see Sec.6.1).

As T decreases, the mean of the unzipping rupture force (squares in Fig.10.8A) increases for both native (bottom) and misfolded states (top)

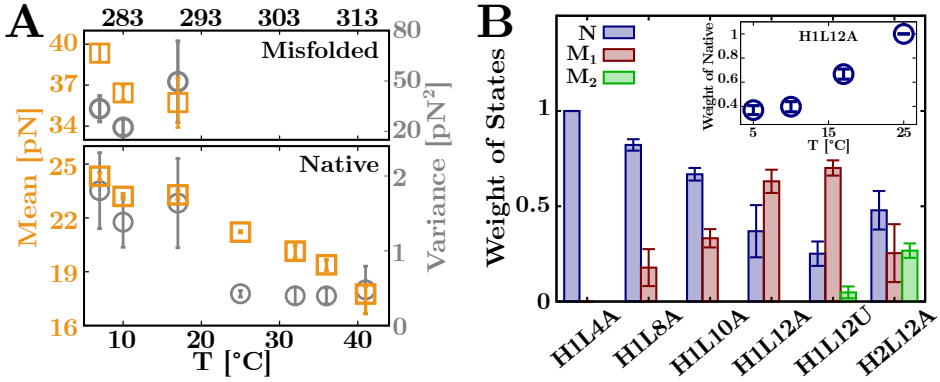


Figure 10.8: RNA unfolding kinetics. (**A**) Mean (orange squares) and variance (grey circles) of the rupture forces distribution for native (bottom) and misfolded (top) at each experimental T (in Celsius and Kelvin degrees).(**B**) Percentage of native and misfolded trajectories for the studied RNA hairpins at 7°C. The H1L12U and the H2L12A molecules exhibit a second misfolded state (green). The inset shows the T-dependence of the weight of the native for the H1L12A hairpin.

according to the increasing hysteresis and hybridization free energy of the folded state. At high forces, the TS moves closer to the F state, increasing the instability of the native state [217] and favoring misfolding. Moreover, the variance of the unzipping rupture force (grey circles in Fig.10.8A) associated with misfolding events is much larger than the native one ($\sigma^2_{f,M} \sim 20 - 50\,\sigma^2_{f,N}$). This gives a $x^{\ddagger}_M = 0.9 \pm 0.1$ nm for misfolding, while $x^{\ddagger}_N = 3.7 \pm 0.4$ nm for the native pointing out that the misfolded structures are compact, kinetically very stable and brittle. Brittleness is related to the deformability of the folded structure upon stretching, a high brittleness implying that the molecule folds into a compact non-deformable configuration. The closer the TS is to the folded state, the more brittle the molecular structure.

Despite low-T misfolding being a universal RNA feature, its recurrence is strongly sequence-dependent. Fig.10.8B shows the fraction of native (blue) and misfolded (red and green) trajectories for each studied hairpin.

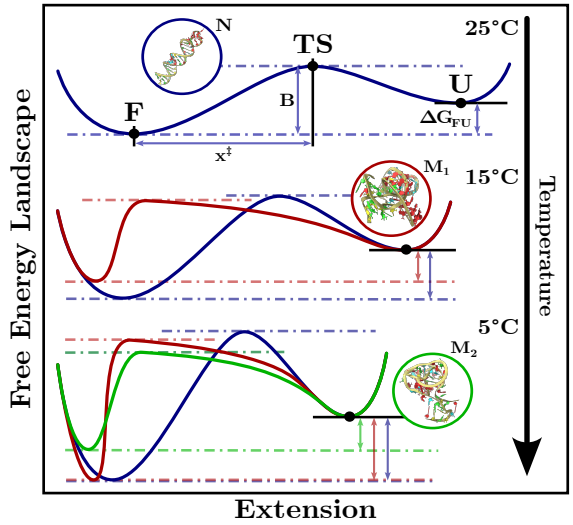


Figure 10.9: RNA kinetics of folding along the FEL at different T. The transition from the folded (F) state of energy ΔG_{FU} to the unfolded (U) state is mediated by the transition state (TS) of height B at a distance $x^{\ddagger}$ from F. In the high-T regime, the molecule can only explore the native (N) FEL (blue line) folding to the native hairpin (inset). Upon lowering T, misfolding increases: multiple FELs associated with different misfolded (M) structures (red and green lines) compete with the native one. We notice that the M_1 and M_2 insets illustrate plausible compact molecular states and are not real misfolded structures.

For a loop of $L = 4$ bases (H1L4A), only the native fold is present. Upon increasing L, the misfolded to native ratio increases as the total ssRNA length increases accordingly. These additional degrees of freedom allow the molecule to arrange into structures other than the native. Notice that only for $L = 12$ bases (H1L12A) at 7°C (Fig.10.8B, inset), the fraction of misfolded is higher than the native. As pointed out in Sec.10.2, the ssRNA elastic response depends on the stacking interaction between the loop bases. When stacking is weaker (poly-U loop), the ssRNA is more flexible, and a second misfolded state appears (H1L12U). In addition, base-pairing interactions also affect the misfolded phenomenology, H2L12A showing two misfolded states. The H2L12A stem features six consecutive A, U

NNBP motifs (AU/UA or UA/AU), which have lower hybridization free energy than G, C motifs. Consequently, the hairpin does not unfold in a single cooperative event: the sequential opening of the A/U bp, one after the other (the so-called *fraying*), precedes the cooperative melting of the second part of the stem. This can be seen as a slope change in the measured FDCs preceding the force rip (see Fig.10.1B). Upon refolding, the weaker stability of the first base pairs of the H2L12A stem allows for forming a second misfolded state (M_2) at low forces.

The study of the unfolding (folding) kinetics allows for the characterization of the FELs associated with the native and misfolding transitions (Fig.10.9). At $T \geq 25°C$, only the native fold is observed. The molecule can only explore the native FEL (blue line) as the kinetic barrier to misfolding is too large to compete with the native one. At $\sim 15°C$, the native is still the prevalent fold but misfolding starts to appear. By decreasing T, both the native B and ΔG_{FU} increase, i.e. more energy is needed for the system to switch from U to F (and vice versa) along the native FEL. Thus, the misfolded FEL (red line) becomes accessible, and misfolding competes with the native transition. The probability of misfolding and the number of misfolded configurations increases as T is further lowered. At $\sim 5°C$, multiple misfolded FELs (red and green lines) compete with the native FEL: misfolding becomes the prevalent transition in the unfolding (folding) reaction.

10.5 CONCLUSIONS

We reported evidence of RNA cold misfolding in unzipping experiments of short RNA hairpins (44 to 52 bases) with a temperature-jump LOT. To study this phenomenon, we unzipped six RNA hairpins made of a 20bp fully complementary Watson-Crick stem (H1 or H2) ending in loops of different sizes ($L = 4, 8, 10, 12$ bases) and compositions (poly-A and

poly-U). Although these hairpins have been designed to only fold into the native (hairpin-like) state (as supported by the available secondary structure prediction tools, such as Mfold and Vienna package, among others), they exhibit one or more misfolded states at low temperatures (Fig.10.1). Above $T = 25°C$, the experimental observations only show the formation of the native state in agreement with the expected behavior. In contrast, misfolded states are formed upon lowering T. At $T \leq 17°C$, misfolding appears with increasing prevalence (Fig.10.5). Although exhibiting sequence-dependent features (number of misfolded states, misfolded/native ratio, etc.), RNA cold misfolding appears as a general RNA phenomenon with common traits: the formed structures are compact and brittle (distance to the TS, $x_M^\ddagger \sim 1/4\, x_N^\ddagger$) and unfold in a broad range of forces ($10 - 40$pN).

The large molecule-to-molecule variability of the measured hybridization free energies (Fig.10.7) indicates that a single misfolded state may be due to the formation of different folds, characterized by different secondary and tertiary structures. This suggests that RNA cold misfolding may have a more rich and complex phenomenology than previously anticipated. However, the accurate characterization of these structures may exceed the instrumental precision, making it hard to distinguish them using the currently available experimental techniques.

The discovery of the RNA cold misfolding enlarges the already rich RNA phenomenology. RNAs' ability to adapt to changing environments may help in understanding the diversity of behaviors observed in RNA, shading light to the underlying mechanisms driving RNA folding.

Part V

APPENDIXES

A

THE WORM-LIKE CHAIN INVERSION

The inextensible WLC model described in (5.19) gives a very direct way to compute $f = f(x)$. On the contrary, the computation of $x = x(f)$ is not straightforward. Although numerical inversion is possible [220], it is helpful to retrieve the explicit inversion formula [102]. Therefore, we show that (5.19) can be inverted to express $z := x/L_c$ as a function of f.

Let us first define the normalized quantity $\tilde{f} = (4P/k_bT)f$. Thus, Eq.(5.19) can be rewritten as $\tilde{f} = (1-z)^{-2} - 1 + 4z$. By multiplying both sides of the previous by $(1-z)^2$ and by moving all terms to the r.h.s., one gets

$$\begin{aligned} &0 = z^3 + a_2 z^2 + a_1 z + a_0 \\ &\text{with} \qquad a_2 = -\frac{9}{4} - \frac{\tilde{f}}{4}, \quad a_1 = \frac{3}{2} + \frac{\tilde{f}}{2}, \quad a_0 = -\frac{\tilde{f}}{4}. \end{aligned} \tag{A.1}$$

Thus, obtaining z as a function of f maps to find a cubic polynomial's roots. The approach taken here is the canonical one [221, 222]. Let us start defining the following intermediate quantities:

$$R := \frac{9a_1a_2 - 27a_0 - 2a_2^3}{54} \qquad Q := \frac{3a_1 - a_2^2}{9}, \tag{A.2}$$

from which we obtain the standard determinant D for cubic equations:

$$D := Q^3 + R^2 \tag{A.3}$$

If $D > 0$, there is only one real solution to Eq.(A.1). To express the solution, let us define

$$T := \sqrt[3]{R + \sqrt{D}} \qquad S := \sqrt[3]{R - \sqrt{D}} \tag{A.4}$$

(since $D > 0$, we also have that $\sqrt{D}$ is real, and thus there is indeed at least one real cubic root for T and S). The desired inverse value $z^* = z(f)$ is then finally obtained as:

$$z^* = -\frac{1}{3}a_2 + S + T\,. \tag{A.5}$$

If $D < 0$, the cubic equation has three real roots, which can be obtained by introducing the following intermediate quantity:

$$\theta := \arccos\left(\frac{R}{\sqrt{-Q^3}}\right), \tag{A.6}$$

from which the three real roots z_1, z_2, z_3 can be obtained directly as:

$$\begin{gathered} z_i = 2\sqrt{-Q}\cos\left(\frac{\theta + \theta_i}{3}\right) - \frac{1}{3}a_2 \\ \text{with} \qquad \theta_1 = 0, \quad \theta_2 = 2\pi, \quad \theta_3 = 4\pi\,. \end{gathered} \tag{A.7}$$

Notice that the roots of Eq.(A.6) can be obtained by re-using the quantities S and T defined above, but doing so is a more complex task that requires using complex number algebra.

The root of interest is the one lying in the interval $[0, 1]$ since $z = x/L_c$, and the inextensible WLC fulfills the property that the extension

x is always smaller than the contour length L_c. Using the standard trigonometric formula and the fact that $2\sqrt{-Q} > 0$, it is pretty easy to verify that $z_1 - z_2 > 0$ and $z_3 - z_2 \geq 0$ for the given range of θ (which must belong to $[0, \pi]$ by definition of the arccosine), which implies that z_2 is the smallest of all the roots. Moreover, notice that all the roots must be positive, since $\forall z < 0$, $f(z) < 0$ in Eq.(5.19) and it is strictly monotonically decreasing. As all the roots are positive and z_2 is the smallest one in $[0, 1]$, $z_2 = z^* = z(f)$ when $D < 0$. This result also extends to the $D = 0$ case as from Eq.(A.6) we get that $\theta = 0$ and then $z_3 = z_2$.

Let us finally note that in the *extensible* WLC case, the main difference with the inextensible case consists in the replacement $L_c \to L_c(1 + f/Y)$, where Y the Young Modulus, i.e. the contour length is now force dependent. It can be shown that the following relationship exists between the two models:

$$x^{ext}_{\text{WLC}}(f) = x^{inext}_{\text{WLC}}(f)\,(1 + f/Y) \tag{A.8}$$

so that the explicit inversion of the inextensible WLC model directly yields the explicit formula for the extensible model.

B

STOCHASTIC GRADIENT DESCENT IN A NUTSHELL

The basic principle behind stochastic approximation can be traced back to the Robbins–Monro algorithm of the 1950*s* [223]. Since then, stochastic gradient descent (SGD) methods have become one of the most widely used optimization methods [224–227].

SGD is an iterative method for optimizing an objective function, $J(w)$, with suitable smoothness properties (e.g., differentiable or subdifferentiable). The set of parameters, w^*, minimizing $J(w)$, is iteratively approximated according to an update algorithm proportional to the anti-gradient of the objective function, $-\nabla_w J(w)$. Starting from an initial guess of w, at each step of the algorithm, the parameters are updated according to

$$\begin{cases} w_{t+1} & = w_t + v_{t+1} \\ v_{t+1} & = \beta v_t - \eta \nabla_{w_t} J(w)\,, \end{cases} \tag{B.1}$$

where v_t is the *velocity* of the optimization and $\eta \geq 0$ is the step size (called *learing rate*). The parameter β (the so-called *momentum coefficient*) accounts for a fraction of the previous step in the current update. The critical difference between SGD and standard gradient descent algorithms is that information (total entropy and coefficients) from only one segment at a time is used to calculate the step, and the segment is picked randomly at each step.

The SGD convergence rate can be improved by considering Nesterov's Accelerated Gradient (NAG), introduced in 1983 [228, 229]. According to NAG, the update equations are

$$\begin{cases} w_{t+1} &= w_t + v_{t+1} \\ v_{t+1} &= \beta v_t - \eta \nabla_{w_t + \beta v_t} J(w) \,. \end{cases} \tag{B.2}$$

While the classic momentum (CM) algorithm updates the velocity vector by computing the gradient at w_t, the NAG algorithm computes the gradient at $w_t + \beta v_t$. To make an analogy, while CM faithfully trusts the gradient at the current iterate, NAG puts less faith in it and looks ahead in the direction suggested by the velocity vector; it then moves in the direction of the gradient at the look-ahead point. If $\nabla_{w_t+\beta v_t} J(w) \approx \nabla_{w_t} J(w)$, then the two updates are similar. The advantage of using NAG is that it converges at a rate of $\mathcal{O}(1/t^2)$, while CM converges at a rate of $\mathcal{O}(1/t)$.

To derive the DNA NNBP entropies from unzipping experiments of a 3.6kbp DNA hairpin (see Sec.7.3), we used an SGD minimization implementing NAG update equations. The application to our case is straightforward. Let us rewrite Eq.(7.5) as $\mathbf{\Delta S}^{\text{exp}} = C\mathbf{\Delta s}$, where $\mathbf{\Delta S}^{\text{exp}}$ is the K elements vector of entropies measure for all segments with the Clausius-Clapeuron equation, $\mathbf{\Delta s}$ is the vector of the $I = 9$ entropy parameters (the 8 independent NNBP entropies *plus* the ssDNA stretching entropy), and C is the $K \times I$ matrix of the coefficients, $c_{k,i}$. We notice that the last parameter, Δs_{ss}, is kept constant and only the remaining 8 NNBP parameters are minimized by the algorithm.

Thus, for a given loss function (ex., least squares), the algorithm has to minimize

$$J(w) = \sum_{k=1}^{K} (\hat{w}_k - w_k)^2 = \sum_{k=1}^{K} (\Delta S_k^{\text{exp}} - C_k \mathbf{\Delta s})^2 \,. \tag{B.3}$$

By using this method, we measured the DNA entropies at the single base-pair level for each experimental temperature in the range $[280, 315]$ K (see results in Fig.7.8 of Sec.7.3.2 and Table C.2 of Appendix C).

C

NNBP DNA THERMODYNAMICS: TABLES OF RESULTS

In the tables that follow, we report the DNA NNBP free energies (Δg_i – Table C.1), entropies (Δs_i – Table C.2), and enthalpies (Δh_i – Table C.3) studied as a function of temperature in the range $[7, 42]^\circ$C ($[280, 315]$ K). These results have been obtained from unzipping experiments of a 3.6kbp DNA hairpin with the temperature-controlled LOT.

A detailed discussion of the experimental technique and the method developed to derive these results can be found in Chapter 7. In particular, the derivation of the NNBP free energies is reported in Sec.7.2. At the same time, the extraction of the NNBP entropies and enthalpies are discussed in Sec.7.3.2 and Sec.7.3.3, respectively.

Ultimately, we tested the NNBP parameters measured from unzipping experiments by computing the melting temperatures for the 92 oligos of Ref. [138] (experimentally measured in bulk assays). Our predictions are reported in Table C.4.

Table C.1: NNBP Δg_i [kcal mol^{-1}] at different temperatures.

Temperature ± 1 [K]	**280**	**283**	**285**	**288**	**291**	**295**	**298**	**303**	**308**	**315**
AA/TT	−1.57(2)	−1.49(2)	−1.55(2)	−1.47(2)	−1.43(1)	−1.43(1)	−1.30(1)	−1.27(1)	−1.19(1)	−1.15(1)
AC/TG	−1.52(2)	−1.38(1)	−1.53(2)	−1.35(1)	−1.47(2)	−1.34(1)	−1.43(1)	−1.36(1)	−1.17(1)	−1.21(1)
AG/TC	−1.73(2)	−1.66(2)	−1.49(2)	−1.60(2)	−1.54(2)	−1.42(1)	−1.41(1)	−1.25(1)	−1.35(1)	−1.25(1)
AT/TA	−1.37(1)	−1.43(1)	−1.27(1)	−1.32(1)	−1.25(1)	−1.28(1)	−1.17(1)	−1.00(1)	−1.19(1)	−1.00(1)
CA/GT	−1.86(2)	−1.92(2)	−1.82(2)	−1.88(2)	−1.82(2)	−1.91(2)	−1.65(2)	−1.68(2)	−1.70(2)	−1.52(2)
CC/GG	−2.03(2)	−1.86(2)	−2.04(2)	−1.91(2)	−2.00(2)	−1.88(2)	−1.91(2)	−1.86(2)	−1.56(2)	−1.70(2)
CG/GC	−2.51(3)	−2.52(3)	−2.39(2)	−2.46(3)	−2.35(2)	−2.24(2)	−2.43(2)	−2.30(2)	−2.03(2)	−1.93(2)
GA/CT	−1.52(2)	−1.59(2)	−1.55(2)	−1.48(2)	−1.49(2)	−1.42(1)	−1.52(2)	−1.46(2)	−1.26(1)	−1.20(1)
GC/CG	−2.79(3)	−2.83(3)	−2.51(3)	−2.78(3)	−2.55(3)	−2.53(3)	−2.49(3)	−2.36(2)	−2.34(2)	−2.11(2)
TA/AT	−1.31(1)	−1.19(1)	−1.10(1)	−1.11(1)	−1.10(1)	−1.00(1)	−1.00(1)	−0.74(1)	−0.97(1)	−0.88(1)

The 10 DNA free-energies measured from unzipping a 3.6kbp hairpin in the temperature range $[280, 315]$ K (see text). The entropy of the last two motifs (GC/CG and TA/AT) has been computed by applying circular symmetry relations. The error (in brackets) refers to the last digit.

Table C.2: NNBP Δs_i [cal mol^{-1}K^{-1}] at different temperatures.

Temperature ± 1 [K]	**280**	**283**	**285**	**288**	**291**	**295**	**298**	**303**	**308**	**315**
AA/TT	−11.0(4)	−11.8(5)	−13.4(6)	−12.8(4)	−13.6(4)	−14.9(4)	−15.1(5)	−15.4(4)	−16.3(4)	−16.9(4)
AC/TG	−14.8(1)	−15.1(1)	−15.7(1)	−16.0(1)	−16.3(1)	−16.8(1)	−17.1(1)	−17.6(1)	−18.2(1)	−19.0(2)
AG/TC	−11.9(3)	−12.3(3)	−12.3(3)	−13.3(3)	−13.4(3)	−13.5(3)	−14.2(3)	−14.9(3)	−15.2(3)	−16.2(2)
AT/TA	−14.6(1)	−14.8(1)	−15.4(1)	−15.4(1)	−16.0(1)	−16.6(2)	−16.3(1)	−16.8(1)	−17.3(1)	−17.3(1)
CA/GT	−12.5(2)	−13.1(2)	−13.7(3)	−14.1(2)	−14.6(2)	−15.6(3)	−15.5(2)	−16.8(3)	−18.2(5)	−17.9(2)
CC/GG	−12.3(3)	−12.9(3)	−12.5(4)	−13.8(3)	−13.3(4)	−13.4(4)	−14.2(3)	−14.7(4)	−14.9(4)	−16.2(2)
CG/GC	−22.2(3)	−22.1(4)	−21.7(5)	−23.2(3)	−22.6(4)	−22.5(4)	−23.2(4)	−23.6(4)	−23.3(4)	−24.4(3)
GA/CT	−12.3(3)	−12.4(4)	−12.8(4)	−13.6(4)	−14.1(2)	−14.2(3)	−14.6(4)	−15.2(4)	−15.8(4)	−16.8(3)
GC/CG	−20.8(3)	−21.1(3)	−20.4(4)	−22.1(3)	−21.4(4)	−21.5(3)	−22.2(3)	−23.1(2)	−23.0(3)	−23.6(3)
TA/AT	−15.5(1)	−15.8(1)	−16.1(1)	−16.2(1)	−16.4(1)	−16.8(1)	−16.9(1)	−17.1(1)	−17.0(2)	−17.5(1)

The 10 DNA entropies measured from unzipping a 3.6kbp hairpin in the temperature range $[280, 315]$ K (see text). The entropy of the last two motifs (GC/CG and TA/AT) has been computed by applying circular symmetry relations. The error (in brackets) refers to the last digit.

Table C.3: NNBP Δh_i [kcal mol^{-1}] at different temperatures.

Temperature ± 1 [K]	**280**	**283**	**285**	**288**	**291**	**295**	**298**	**303**	**308**	**315**
AA/TT	−4.66(11)	−4.84(14)	−5.38(16)	−5.17(12)	−5.41(12)	−5.83(13)	−5.79(15)	−5.93(12)	−6.23(12)	−6.49(12)
AC/TG	−5.67(4)	−5.66(4)	−6.01(4)	−5.98(4)	−6.22(3)	−6.32(4)	−6.53(4)	−6.68(3)	−6.78(4)	−7.18(6)
AG/TC	−5.07(8)	−5.13(8)	−5.00(10)	−5.44(9)	−5.44(8)	−5.41(10)	−5.63(9)	−5.77(9)	−6.03(10)	−6.37(8)
AT/TA	−5.46(3)	−5.62(3)	−5.68(3)	−5.76(3)	−5.91(4)	−6.19(5)	−6.03(3)	−6.10(4)	−6.53(4)	−6.45(3)
CA/GT	−5.35(6)	−5.62(7)	−5.74(8)	−5.97(6)	−6.08(7)	−6.52(9)	−6.28(6)	−6.77(10)	−7.30(16)	−7.17(7)
CC/GG	−5.48(9)	−5.51(8)	−5.61(11)	−5.91(8)	−5.87(11)	−5.82(11)	−6.14(10)	−6.31(11)	−6.16(12)	−6.81(7)
CG/GC	−8.73(9)	−8.79(12)	−8.59(14)	−9.15(9)	−8.96(11)	−8.88(12)	−9.33(11)	−9.46(11)	−9.24(14)	−9.63(10)
GA/CT	−4.97(8)	−5.09(11)	−5.20(11)	−5.40(11)	−5.62(7)	−5.63(9)	−5.86(11)	−6.06(13)	−6.13(11)	−6.49(10)
GC/CG	−8.63(9)	−8.79(10)	−8.35(13)	−9.17(8)	−8.80(11)	−8.88(10)	−9.09(10)	−9.35(8)	−9.45(10)	−9.56(10)
TA/AT	−5.67(3)	−5.66(3)	−5.71(3)	−5.79(3)	−5.89(3)	−5.98(3)	−6.04(4)	−5.91(3)	−6.22(5)	−6.40(4)

The 10 DNA enthalpies measured from unzipping a 3.6kbp hairpin in the temperature range $[280, 315]$ K (see text). The entropy of the last two motifs (GC/CG and TA/AT) has been computed by applying circular symmetry relations. The error (in brackets) refers to the last digit.

Table C.4: DNA Oligos Melting Temperatures [°C]

Sequence ($5' \to 3'$)	$T^{exp} \pm 1.6°$	$T^{unz} \pm 2°$	T^{UO}	T^{Hug}
ATCAATCATA	33.6	32	34.0	33.4
TTGTAGTCAT	36.0	34	36.7	33.5
GAAATGAAAG	34.4	36	34.6	32.6
CCAACTTCTT	40.6	39	40.4	36.9
ATCGTCTGGA	44.9	42	46.2	43.3
AGCGTAAGTC	40.3	42	45.1	41.8
CGATCTGCGA	49.1	49	50.5	48.2
TGGCGAGCAC	55.3	54	56.3	51.8
GATGCGCTCG	53.5	53	54.0	51.7
GGGACCGCCT	57.0	53	58.6	53.3
CGTACACATGC	49.9	49	51.2	48.9
CCATTGCTACC	48.9	50	49.6	46.6
TACTAACATTAACTA	51.1	51	51.9	50.8
ATACTTACTGATTAG	51.5	52	49.7	49.4
GTACACTGTCTTATA	54.8	53	54.8	53.7
GTATGAGAGACTTTA	55.4	54	54.8	53.9
TTCTACCTATGTGAT	53.7	56	55.1	54.3
AGTAGTAATCACACC	57.1	56	56.9	55.8
ATCGTCTCGGTATAA	58.6	58	58.9	58.5
ACGACAGGTTTACCA	61.3	62	63.6	61.5
CTTTCATGTCCGCAT	62.8	64	63.0	62.4
TGGATGTGTGAACAC	60.4	61	62.3	60.4
ACCCCGCAATACATG	62.9	66	65.6	63.7
GCAGTGGATGTGAGA	63.3	64	64.6	62.6
GGTCCTTACTTGGTG	60.3	61	62.0	59.4
CGCCTCATGCTCATC	65.8	67	66.5	65.1
AAATAGCCGGGCCGC	70.4	73	72.7	69.1
CCAGCCAGTCTCTCC	66.7	67	67.7	64.7
GACGACAAGACCGCG	68.6	67	69.7	68.2
CAGCCTCGTCGCAGC	72.0	72	73.0	70.6
CTCGCGGTCGAAGCG	70.7	71	72.9	71.3
GCGTCGGTCCGGGCT	74.1	73	77.8	74.2
TATGTATATTTTGTAATCAG	61.2	62	58.6	59.6
TTCAAGTTAAACATTCTATC	61.5	65	60.6	60.9
TGATTCTACCTATGTGATTT	64.4	67	63.7	64.2
GAGATTGTTTCCCTTTCAAA	65.3	70	66.3	65.4
ATGCAATGCTACATATTCGC	68.9	73	69.2	69.7
CCACTATACCATCTATGTAC	64.4	65	63.9	63.9
CCATCATTGTGTCTACCTCA	68.5	70	69.4	68.7
CGGGACCAACTAAAGGAAAT	68.5	72	70.3	69.3

Table C.4: DNA Oligos Melting Temperatures [°C]

Sequence ($5' \to 3'$)	$\mathbf{T^{exp}} \pm 1.6°$	$\mathbf{T^{unz}} \pm 2°$	$\mathbf{T^{UO}}$	$\mathbf{T^{Hug}}$
TAGTGGCGATTAGATTCTGC	71.2	72	71.1	70.6
AGCTGCAGTGGATGTGAGAA	73.1	76	74.5	73.8
TACTTCCAGTGCTCAGCGTA	73.6	74	76.0	74.8
CAGTGAGACAGCAATGGTCG	72.5	74	73.5	73.1
CGAGCTTATCCCTATCCCTC	70.3	73	71.3	70.1
CGTACTAGCGTTGGTCATGG	71.1	72	72.9	72.2
AAGGCGAGTCAGGCTCAGTG	76.3	77	77.2	76.3
ACCGACGACGCTGATCCGAT	77.3	76	78.7	79.4
AGCAGTCCGCCACACCCTGA	78.5	80	81.6	78.8
CAGCCTCGTTCGCACAGCCC	78.1	80	81.1	78.6
GTGGTGGGCCGTGCGCTCTG	81.0	81	83.6	80.2
GTCCACGCCCGGTGCGACGG	81.1	81	85.4	82.0
GATATAGCAAAATTCTAAGTTAATA	66.1	70	64.2	65.6
ATAACTTTACGTGTGTGACCTATTA	71.8	72	71.2	72.3
GTTCTATACTCTTGAAGTTGATTAC	67.7	70	67.3	68.6
CCCTGCACTTTAACTGAATTGTTTA	72.5	76	73.4	73.0
TAACCATACTGAATACCTTTTGACG	71.3	74	72.2	72.5
TCCACACGGTAGTAAAATTAGGCTT	73.8	76	74.6	74.2
TTCCAAAAGGAGTTATGAGTTGCGA	73.8	78	75.2	74.6
AATATCTCTCATGCGCCAAGCTACA	76.5	80	76.7	77.0
TAGTATATCGCAGCATCATACAGGC	75.0	77	75.5	75.8
TGGATTCTACTCAACCTTAGTCTGG	73.6	76	73.9	73.6
CGGAATCCATGTTACTTCGGCTATC	74.8	77	75.5	75.6
CTGGTCTGGATCTGAGAACTTCAGG	75.6	78	76.6	76.7
ACAGCGAATGGACCTACGTGGCCTT	81.0	82	82.7	82.2
AGCAAGTCGAGCAGGGCCTACGTTT	81.5	83	82.8	82.4
GCGAGCGACAGGTTACTTGGCTGAT	80.1	81	81.3	81.1
AAAGGTGTCGCGGAGAGTCGTGCTG	82.4	82	83.0	83.2
ATGGGTGGGAGCCTCGGTAGCAGCC	83.4	86	86.6	83.6
CAGTGGGCTCCTGGGCGTGCTGGTC	83.4	86	87.5	84.3
GCCAACTCCGTCGCCGTTCGTGCGC	84.6	85	88.1	86.3
ACGGGTCCCCGCACCGCACCGCCAG	88.3	89	93.0	88.9
TTATGTATTAAGTTATATAGTAGTAGTAGT	66.6	70	65.8	68.1
ATTGATATCCTTTTCTATTCATCTTTCATT	70.4	76	70.3	72.9
AAAGTACATCAACATAGAGAATTGCATTTC	73.2	77	73.0	75.0
CTTAAGATATGAGAACTTCAACTAATGTGT	71.8	76	71.8	74.1
CTCAACTTGCGGTAAATAAATCGCTTAATC	75.5	79	75.2	76.2
TATTGAGAACAAGTGTCCGATTAGCAGAAA	76.4	80	77.5	78.3
GTCATACGACTGAGTGCAACATTGTTCAAA	76.9	79	78.0	79.0
AACCTGCAACATGGAGTTTTTGTCTCATGC	78.7	82	80.3	80.6

Table C.4: DNA Oligos Melting Temperatures [°C]

Sequence ($5' \to 3'$)	$\mathbf{T^{exp}} \pm 1.6°$	$\mathbf{T^{unz}} \pm 2°$	$\mathbf{T^{UO}}$	$\mathbf{T^{Hug}}$
CCGTGCGGTGTGTACGTTTTATTCATCATA	77.6	80	80.0	80.5
GTTCACGTCCGAAAGCTCGAAAAAGGATAC	78.7	80	79.5	80.3
AGTCTGGTCTGGATCTGAGAACTTCAGGCT	80.6	83	82.2	83.5
TCGGAGAAATCACTGAGCTGCCTGAGAAGA	80.9	84	82.5	82.6
CTTCAACGGATCAGGTAGGACTGTGGTGGG	80.1	83	83.3	82.5
ACGCCCACAGGATTAGGCTGGCCCACATTG	84.0	87	87.5	85.4
GTTATTCCGCAGTCCGATGGCAGCAGGCTC	84.1	86	85.9	84.8
TCAGTAGGCGTGACGCAGAGCTGGCGATGG	84.6	87	88.1	87.1
CGCGCCACGTGTGATCTACAGCCGTTCGGC	84.5	87	89.0	88.2
GACCTGACGTGGACCGCTCCTGGGCGTGGT	86.4	88	91.2	89.5
GCCCCTCCACTGGCCGACGGCAGCAGGCTC	87.7	92	93.8	89.9
CGCCGCTGCCGACTGGAGGAGCGCGGGACG	88.6	92	94.8	92.2

Melting temperature (T^{unz}) prediction at 1 M NaCl for the 92 oligos of Ref. [138] obtained with our parameters. T^{exp} are the experimentally measured temperatures by Owczarzy *et al.* in Ref. [138]. T^{UO} and T^{Hug} are the predictions obtained by using the unified oligonucleotide [22] and the Huguet *et al.* [133] parameters, respectively. All temperatures are given in Celsius degrees.

D

SYNTHESIS OF RNA MOLECULES

LOT pulling experiments require using NAs molecular constructs suitable for the setup (see Sec.3.3) and the study that has to be carried out. In what follows, we report the protocols for synthesizing the RNA molecular constructs used in this thesis. In Sec.D.1, we describe the protocol developed to synthesize the 2kbp long RNA hairpin studied in Part 8. In Sec.D.2, we report the synthesis of the different short RNA hairpins studied in Part 10.

D.1 LONG RNA HAIRPINS

We synthesized an RNA hairpin made of a stem of 2027 equally represented canonical Watson-Crick base pairs (Table D.1), ending in a GAAA tetraloop and inserted between short hybrid DNA/RNA handles (29bp). The realization of this molecule has been challenging (see main text), requiring the development of a tailored synthesis protocol. This protocol (schematically depicted in Fig.D.1) can be split into 7 main steps.

1. *PCR amplification of target sequence*

The target sequence was selected inside a λ-DNA (J02459) region (30286 – 38650) that was previously shown to be efficiently transcribed in both

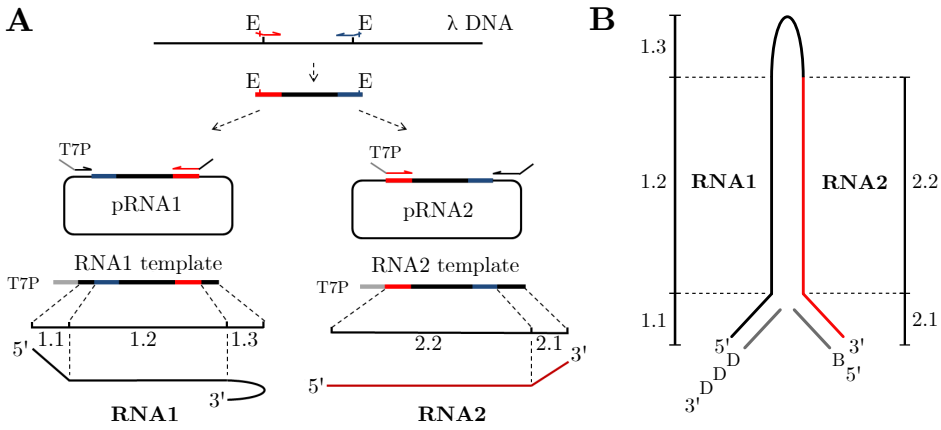


Figure D.1: RNA hairpin synthesis protocol. **(A)** A PCR amplicon of 2027bp in length obtained from λ-DNA and containing EcoRI sites (E) at its termini was cloned into pBR322 plasmid in both orientations, generating the pRNA1 and pRNA2 constructs. The pRNA1 and pRNA2 constructs were then used as templates for PCR reactions. Both PCR products were used as templates for *in vitro* transcription reactions to synthesize RNA1 and RNA2 molecules. Regions 1.1 and 2.1 are derived from the pBR322 sequence, regions 1.2 and 2.2 from the λ-DNA sequence, and region 1.3 from the RNA1 reverse primer. **(B)** RNA hairpin structure and assembly. The hairpin is formed by annealing molecules RNA1 and RNA2, a digoxigenin (DIG)-labeled, and biotin (BIO)-labeled oligonucleotide. RNA1 molecule contains three regions: region 1.1 anneals with DIG-labeled oligonucleotide to form RNA1 handle. Region 1.2 anneals with RNA2 and together with region 2.2 from RNA2 forms most of the hairpin stem. Finally, region 1.3 forms the hairpin loop and the upper part of the stem. Apart from region 2.2, the RNA2 molecule also contains a 3′ portion (region 2.1) that anneals with BIO-labeled oligonucleotide to form the RNA2 handle.

strands by T7 RNA polymerase, the same enzyme used in our experiments [111]. A PCR reaction was performed to obtain an amplicon of 2027bp in length using 1μM of Univ_hairpin_F and 1μM of EcoRI_2.0kb_R primers (Table D.2), 25ng of λ-DNA (Dam-) as DNA template, 1.5mM $MgCl_2$, 1X Opti and 1X HiSpec buffers, dNTPs 0.2mM each, and 4U of Eco Taq Plus DNA Polymerase (Ecogen). Primer sequences were selected using Primer3Plus software [230]. EcoRI sites were added at 5′ termini of both primers (Table D.2, sequence in bold). Cycling parameters were as follow: initial denaturation step (94°C) for 1 min 30 sec, enzyme addition (hot start), 30 cycles of denaturation at 94°C for 45 sec, annealing at 60°C for 1 min and extension at 72°C for 6 min, with a final extension step at 72°C for 7 min.

2. *Synthesis of pRNA*1 *and pRNA*2 *constructs*

The 2kbp PCR amplicon was digested with EcoRI (NEB, New England Biolabs) and cloned into vector pBR322/EcoRI [231, 232]. Plasmid DNA was purified using Illustra PlasmidPrep Mini Spin Kit (GE Healthcare). The insert orientation was evaluated by digesting plasmids with EcoRV, and according to it, constructs were defined as pRNA1 or pRNA2 (Fig.D.1). These constructs were used as templates for PCR reactions using primers RNA1_T7Forw and RNA1_Rev (pRNA1) or RNA2_T7Forw and RNA2_Rev (pRNA2) (Table D.2).

3. *PCR amplification of templates for in vitro transcription*

The pRNA1 and pRNA2 constructs were used as templates for PCR reactions using primers RNA1_T7Forw and RNA1_Rev (pRNA1) or RNA2_T7Forw and RNA2_Rev (pRNA2) (see Table D.2). RNA_T7Forw primer contains a cytidilate nucleotide (in bold) upstream the minimal T7 RNA Polymerase Promoter (5′ – CTAATACGACTCACTATAGGGA – 3′) to improve transcription efficiency [233], followed by a pBR322-

NNBP	Frequency (%)
AA/UU	17.1
CA/GU	14.7
GA/CU	12.7
AU/UA	9.6
GU/CA	10.0
CC/GG	8.7
CG/GC	4.1
AG/UC	11.0
GC/CG	5.6
UA/AU	6.3

Table D.1: Occurrence of NN motifs in the RNA sequence.

annealing sequence ($5' - ATAAAAATAGGCGTATCACGAG - 3'$). This sequence codes for part of the RNA1 handle. Primer RNA1_Rev anneals at its $3'$ termini ($5'$-GAAAAACGCCTCGAGTGAAG-$3'$) with the Univ_hairpin_F binding site located at the end of pRNA1 insert opposite to RNA1_T7Forw binding site (see Fig.D.1). The $5'$ termini of RNA1_Rev ($5' - CTCATCTGTTTCCAGATGAG - 3'$) codes for the last 8bp of the RNA hairpin near the loop and the GAAA tetraloop itself (in bold, reverse complement). The sequence $5' - GGGA - 3'$ was introduced between $5'$ and $3'$ portions of RNA1_Rev in order to introduce the sequence $5' - UCCC - 3'$ into RNA1. This tetranucleotide RNA sequence is located between the hairpin stem portion formed by sequences from the 2kbp insert and the last 8bp-stem and loop regions coded by RNA1_Rev sequence and serves to base pair the first four nucleotides ($5' - GGGA - 3'$) at the $5'$ end of RNA2 molecule. RNA2_T7Forw anneals with same sequence that pairs with RNA1_Rev primer ($5' - GAAAAACGCCTCGAGTGAAG - 3'$), but it is used in PCR reactions with pRNA2 construct. As in the case of RNA1_T7Forw, its $5'$ termini contain an optimized T7 promoter containing an upstream

Oligonucleotides	Sequence
Univ_hairpin_F	5′ – AC**GAATTC**GAAAAACGCCTCGAGTGAAG – 3′
EcoRI_2.0kb_R	5′ – AC**GAATTC**TTGGGGTGTGTGATACGAAA – 3′
RNA1_T7Forw	5′ – CTAATACGACTCACTATAGGGAATAAAAATAGGCGTATCACGAG – 3′
RNA1_Rev	5′ – CTCATCTGTTTCCAGATGAGGGGAGAAAAACGCCTCGAGTGAAG – 3′
RNA2_T7Forw	5′ – CTAATACGACTCACTATAGGGAGAAAAACGCCTCGAGTGAAG – 3′
RNA2_Rev	5′ – GAACATACGAAACGGATGATAAGCTGTCAAACA – 3′
S Handle A	5′ – ACGAAAGGGCCTCGTGATACGCCTATTTTT – 3′
S Handle B2	5′ – Bio – GAACATACGAAACGGATGATAAGCTGTCAA – 3′

Table D.2: List of the oligonucleotides used in the synthesis of the 2027bp RNA hairpin.

C nucleotide (5′ – CTAATACGACTCACTATAGGGA – 3′). RNA2_Rev primer contains a pBR322-annealing region and codes for the RNA2 handle. PCR reactions were performed using the same conditions previously described. Amplification products were purified from PCR mixtures using the GFX PCR DNA and Gel Band Purification Kit (GE Healthcare).

4. *In vitro transcription of RNA*1 *and RNA*2 *molecules*

in vitro transcription reactions were performed using the T7 MEGAscript-High Yield transcription Kit (ThermoFisher Scientific/Ambion) according to the manufacturer's recommendations. Samples were incubated with 3μL of TURBO DNase (2U/μL) for 15 min at 37°C, and synthesized RNA was precipitated by adding 90μL of LiCl Precipitation solution (7.5M lithium chloride, 50mM EDTA). Reactions were incubated overnight at −20°C, centrifuged for 15 min at 13,000 rpm, washed twice with 70% ethanol, and resuspended in 15μL of nuclease-free water.

5. *Treatment of RNA*2 *molecules with Calf Intestinal Phosphatase (CIAP) and Polynucleotide Kinase (PNK)*

in vitro transcribed RNA2 molecules were treated with 1U of CIAP (Roche) for 1h at 50°C to remove their 5′ triphosphate ends. Dephosphorylated RNA2 molecules were treated with Polynucleotide Kinase (PNK) to

produce RNA molecules containing 5′ monophosphate termini according to the manufacturer's recommendations. The reactions were heat-inactivated by incubating for 20 min at 65°C and precipitated with LiCl, as described in the following section.

6. *Digoxigenin 3′ tailing of S Handle A oligonucleotide*

S Handle A (Table D.2) tailing with digoxigenins was performed by using the DIG oligonucleotide Tailing Kit 2nd Generation (Roche), according to the manufacturer's recommendations. DIG-labeled S Handle A was purified using the Qiaquick Nucleotide Removal kit (Qiagen).

7. *Assembling RNA*1 *and RNA*2 *molecules to form the* 2*kbp RNA hairpin*

The assembly of RNA hairpin was performed in one annealing step, where RNA1, CIAP, and PNK-treated RNA2, S Handle A and biotin-labeled S Handle B2 oligonucleotides (Table D.2) were incubated together. A total of 20μg of RNA1 and 20μg of CIAP/PNK-treated RNA2 were incubated with 5μL of DIG-tailed S Handle A (2μM), 5μL of 5′-Bio-S Handle B2 (2μM), 2μL of Tris 1M, pH 7.0, 2μL of NaCl 5M and water to a final volume of 80μL. Reactions were incubated for 1 h at 65°C and cooled to 10°C at a rate of 0.5°C/min using a thermocycler. After a final cooling step at 10°C for 1 h and 30 min, the samples were subjected to microdialysis. The annealing reaction was pipetted over a 0.05μm Millipore membrane which was put in a plate containing 50mL of 20mM Tris.HCl, 5mM NaCl, pH 7.5, and allowed to stay for 1 h. Microdialyzed, annealed molecules were then incubated with 1μL of T4 RNA ligase 2 (RNL2 1U/μL) (NEB) and 1X RNL2 Reaction Buffer for 2 h at 37°C to covalently join RNA1 and RNA2 molecules. Ligated RNA hairpin molecules were microdialyzed against Tris.HCL 100mM, EDTA 14mM as described above and stored at −20°C or directly used in single-molecule experiments.

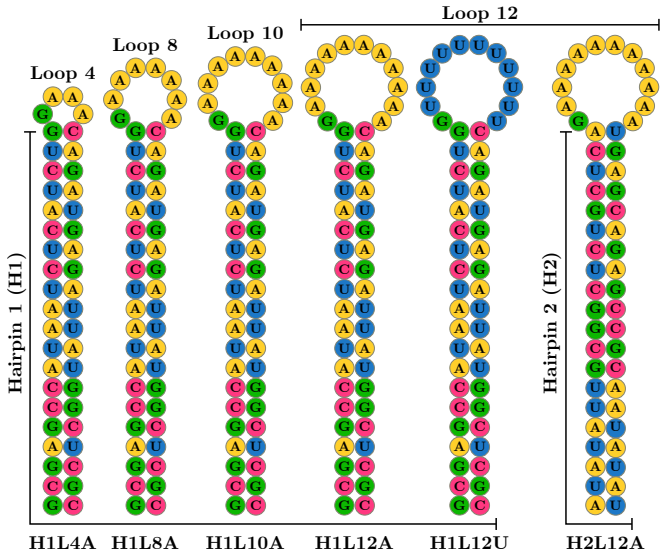


Figure D.2: Molecular structure of the six RNA hairpins. Each hairpin is named after the sequence (H1 or H2), the loop size (L4, L8, L10, or L12), and the loop composition (poly-A or poly-U). The sequences are reported in Table D.3.

D.2 SHORT RNA HAIRPINS

We synthesized six short RNAs with long hybrid RNA/DNA heterohandles ($\sim$ 500bp). The hairpins are made of a 20bp stem chosen between two possible sequences (H1 and H2) and loops of various lengths ($L = 4, 8, 10, 12$ bases) and composition (poly-A or poly-U). The different RNA sequences are shown in Fig.D.2 and reported in Table D.3. The dif-

Table D.3: Sequences of the six RNA hairpins ($5' \rightarrow 3'$ direction).

Molecule	Stem Sequence	Loop Sequence
H1L4A	GCGAGCCAUAAUCUCAUCUG	GAAA
H1L8A	GCGAGCCAUAAUCUCAUCUG	GAAAAAAA
H1L10A	GCGAGCCAUAAUCUCAUCUG	GAAAAAAAAA
H1L12A	GCGAGCCAUAAUCUCAUCUG	GAAAAAAAAAAA
H1L12U	GCGAGCCAUAAUCUCAUCUG	GUUUUUUUUUUU
H2L12A	AUAUAUUGCGGCUCUGCUCA	GAAAAAAAAAAA

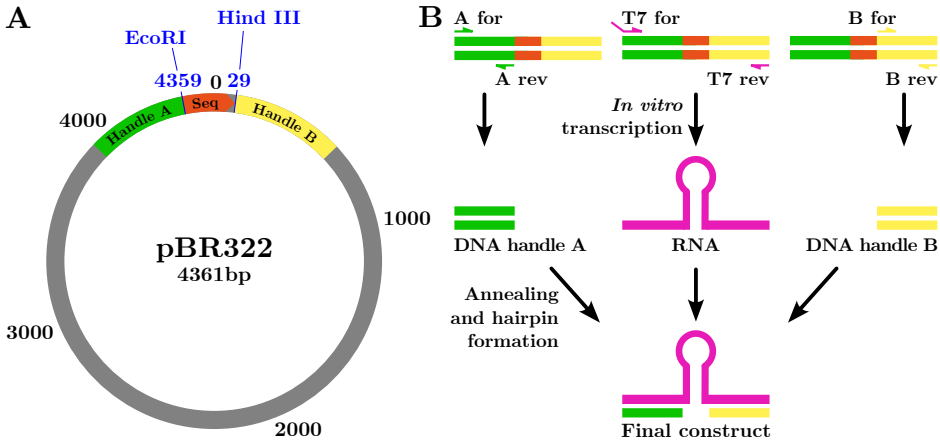


Figure D.3: Synthesis of short RNA hairpins (see text). (**A**) The RNA hairpin (orange) sequence is cloned into the pBR322 plasmid vector. The flanking regions are used to create the handles for the LOT setup (green and yellow). (**B**) Steps of the synthesis protocol. The templates for the RNA transcription and the handles are obtained via PCR amplification of different regions of the plasmid (top and middle). Handles A and B (green and yellow) are tagged with digoxigenin and biotin and annealed to the RNA transcript (magenta) to create the hairpin (bottom).

ferent sequences have been cloned between the EcoRI and HindIII sites of plasmid pBR322 (Fig.D.3A), and the flanking regions of the plasmid were used to engineer DNA/RNA heteroduplex handles for single-molecule pulling experiments (Fig.D.3B). In what follows, we describe each step of the synthesis protocol (adapted from [234]).

1.1 *Preparation of the DNA template*

The molecular template for the RNA hairpins (Eg. the H1L12A RNA hairpin in Table D.3) is made by two complementary DNA oligonucleotides with the sequence of the molecule of interest (see Table D.4). Each oligo has EcoRI and HindIII compatible overhang at one end. The oligonu-

Table D.4: DNA template for the H1L12A hairpin synthesis.

Oligo 1	5′ – *AATTC* **AA** GCGAGCCATAATCTCATCTG
	GAAAAAAAAAAA CAGATGAGATTATGGCTCGC **AA** A – 3′
Oligo 2	5′ – *AGCTT* **TT** GCGAGCCATAATCTCATCTG
	TTTTTTTTTTTC CAGATGAGATTATGGCTCGC **TT** G – 3′

The DNA oligo 1 and the complementary sequence oligo 2 cloned between the EcoRI and Hind III sites of pBR322 plasmid. The sequences include the enzyme restriction sites (in italics) and ssDNA spacers for the LOT setup (in bold).

cleotides are phosphorylated and annealed into their dsDNA using T4 DNA kinase:

1.1 *Phosphorylation and annealing:*

MilliQ water	40μl
Oligo A (100μM in Tris 10mM)	20μl
Oligo B (100μM in Tris 10mM)	20μl
T4 polynucleotide kinase (10U/μl)	10μl
10X T4 ligase buffer	10μl
Total volume	100μl

Protocol: Incubate 1h at 37°C, then anneal the oligos using a temperature gradient (95°C to 4°C in 2h).

1.2 *Digestion of the pBR*322 *plasmid*

To obtain the recombinant plasmid for RNA transcription of the molecular construct, the dsDNA template obtained at step 1.1 is inserted between the HindIII and EcoRI sites of the pBR322 plasmid (Fig.D.3A). To do this, the plasmid is incubated with the two restriction enzymes and with Calf Intestinal Alkaline Phosphatase (CIP):

1.2 *Vector digestion:*

MilliQ water	131μl
pBR322 (0.25μg/ml)	50μl
CIP enzyme	4.5μl
HindIII	3μl
EcoRI	3μl
10X NEB EcoRI buffer	20μl
Total volume	211.5μl

Protocol: Incubate 1h at 37°C.

The digested vector is then gel purified (0.8% agarose gel, 1XTBE) using Sybr Safe staining (5μl) and blue illumination to minimize DNA damage and nicking. The vector is extracted using the Qiaquick gel extraction kit (Qiagen).

1.3 *DNA template and vector ligation*

The linearized and dephosphorylated vector is ligated with the dsDNA template from step 1.1. In this way, the plasmid is circularized again, but it now contains the DNA insert that is the template of the desired RNA molecule.

1.3 *Ligation of recombinant plasmid:*

MilliQ water	3μl
pBR322 (DNA template)	2μl
ATP 10μM	1μl
Digested vector (100 ng/μl)	2μl
T4 DNA ligase (20U/μl)	1μl
10X T4 ligase buffer	1μl
Total volume	10μl

Protocol: Incubate at 16°C overnight (or at least for 4 hours).

After the incubation, the product is gel purified (1% agarose gel, 1XTBE) to ensure the correct ligation of the recombinant plasmid. Finally, the vector is extracted using the Qiaquick gel extraction kit (Qiagen). NOTE: The recombinant plasmid can be frozen for successive usage at this step.

2.1 *Clonation of the recombinant plasmid*

To increase the concentration of recombinant plasmid obtained at the end of step 1.3, we cloned the plasmid using ultra-competent cells (UCC) XL10-GOLD (Quickchange II XL site-directed mutagenesis kit). The transformation has been done following the kit instructions (but replacing the NZY+ Broth with LB):

1. Aliquot 45μl UCC on microtubes previously pre-cooled on ice.
2. Add 2μl β-mercaptoethanol. Mix and incubate for 10 min on ice, mixing every 2 min.
3. Add 1μl plasmidic DNA (20 ng/μl). Mix well and incubate for 30 min on ice. Meanwhile, preheat LB at 42°C.
4. Heat-pulse the transformation reaction at 42°C for exactly 30 seconds. Incubate on ice for 2 min.
5. Add 1ml LB preheated at 42°C and incubate on a shaker at 37°C at $225 - 250$rpm for 1 hour.
6. Prepare Petri dishes with an Agar medium substrate for cell culture. Apply $\approx 15 - 20$ml of Agar medium per dish. Store solidified plates at 4°C facing down in a sealed bag. We prepared 15 plates.
7. Prepare increasing dilutions of the preparation. We prepared 3 different dilutions:
    - *Concentrated.* Centrifuge 400μl preparation for 2 min, remove 300μl supernatant and inoculate the remaining 100μl.

- *Normal.* Inoculate 100μl of the preparation.
- *Eluted.* Inoculate 10μl of the preparation +90μl of LB.

8. Place the dilutions on the dry agar plates (5 per dilution) and spread with a glass spreader sterilized by immersion in 95% ethanol and burnt off with a Bunsen flame before and after inoculation (cooldown by touching an agar plate before culturing). Incubate overnight at 37°C with agar facing down (without agitation). NOTE: The bacteria grown on the LB agar plate can be stored at 37°C for a few weeks.

2.1 *LB liquid culture:*

Triptone	5g
Yeast extract	2.5g
NaCl	5g
MilliQ water	500ml
Total volume	500ml

Protocol: Autoclave and use in sterile conditions.

2.1 *Agar medium:*

Triptone	2g
Yeast extract	1g
NaCl	2g
Agar	3g
MilliQ water	200ml
Total volume	200ml

Protocol: Autoclave, let cool down to $\approx$ 50°C and add 400μl of Ampicillin (50μg/ml) in sterile conditions.

Table D.5: Primers sequence

A–for	5′–ggaattccGACTGGTGAGTACTCAACCAAGTC–3′
A–rev	5′–ATTCTTGAAGACGAAAGGGC–3′
B–for	5′–GCTTTAATGCGGTAGTTTATCACAG–3′
B–rev	5′–Digoxigenin-GCATTAGGAAGCAGCCCAGTAGTAGG–3′
T7–for	5′–taatacgactcactatagggaCTGGTGAGTACTCAACCAAGTC–3′
T7–rev	5′–TAGGAAGCAGCCCAGTAGTAGG–3′

Primers for amplifying the handles and the transcription of the molecular construct. Lowercase sequences correspond to non-hybridizing fragments.

2.2 *Selection of recombinants colonies*

Prepare 15ml Falcon tubes (as many as the Petry dishes) with 3ml sterile LB medium and add Ampicillin (50μg/ml) in sterile conditions. Select the colonies of bacteria, picking them up from agar plates with a sterile micropipette tip or tooth stick. Inoculate a single colony per tube, taking the biggest from each plate. Incubate the liquid cultures overnight at 37°C with agitation. NOTE: leave the tubes' cap slightly loose.

After growing the cultures overnight, select the Falcon tubes showing the highest cell concentration. Centrifuge the selected tube to precipitate the cultures and remove the supernatant fluid with a pipette. Then, the recombinant plasmid is extracted using a miniprep extraction kit (Qiaprep from Qiagen)

3 *PCR amplification of DNA handles and template for in vitro transcription*

The recombinant vector is used both to create the DNA heterohandles and the template for *in vitro* transcription of the RNA structure (see Fig.D.3B). To do this, we performed three different PCR amplifications corresponding to the template for the in vitro transcription (Fig.D.3B, green, orange, and yellow), handle A (Fig.D.3B, green), and handle B (Fig.D.3B, yellow). The sequences of the primers used for the different

PCRs are reported in Table D.5. The primers are designed to create a handle that hybridizes with one end of the RNA transcript (handle A, 535bp), and a handle that hybridizes with the other end of the transcript (handle B, 599bp). The reverse primer for handle B contains a digoxigenin at its 5′ end for the molecular construct tethering with the LOT setup. To improve the reaction performance, both forward primers (following $5' \rightarrow 3'$ direction), and reverse primers (following $3' \rightarrow 5'$ direction) have been used.

Before performing the PCR, the circular plasmid is linearized by digestion with the PtsI restriction enzyme at 37°C for 1 hour. The PtsI enzyme was chosen because it cuts the plasmid close to the T7 primer attaching point. Then, the PCR amplifications of the three segments can be done:

3 1X *PCR reaction (handle A, handle B, DNA template):*

MilliQ water	52.2μl
KOD buffer	10μl
KOD enzyme	0.8μl
dNTPs 2mM	10μl
$MgCl_2$	4μl
Primer Forward 10μM	10μl
Primer Reverse 10μM	10μl
Recombinant Plasmid 1ng/μl	3μl
Total volume	100ml

Protocol (DNA template): 95°C for 1 min; 30×(95°C for 45 sec, 60°C for 1 min, 72°C for 90 sec); 72°C for 5 min.

Protocol (handle A/B): 95°C for 1 min; 30×(95°C for 45 sec, 54°C for 1 min, 72°C for 90 sec); 72°C for 5 min.

The PCR products were gel purified (1% agarose gel, 1XTBE) and extracted using the QIAquiak PCR purification kit (QIAgen).

4 *Handle A tailing*

The handles of the molecular construct have different tags (biotin and digoxigenin, respectively), allowing for specific binding with the SA and AD beads of the LOT experimental setup. This requires tailing the two handles: handle A with biotin and handle B with digoxigenin. In this case, handle B has been purchased already tailed with digoxigenin. The purified strand from the PCR of handle A (step 3) is biotinated according to the following protocol:

4 *Biotination of handle A:*

MilliQ water	44μl
Handle A	30μl
T4 polymerase Buffer (NEB2)	10μl
T4 DNA polymerase	5μl
BSA 100x	1μl
Biotin 16dUTP 1M	10μl
Total volume	100ml

Protocol: Incubate at room temperature for 20 min.

5 *In vitro transcription of RNA*

The transcription of the DNA template to RNA is performed using the T7 Megascript Kit (Ambion). First, the PCR amplified template is concentrated by centrifugal evaporation or ethanol precipitation and eluted in 20μl RNase-free water.

Transcription is then performed according to the kit instructions:

The reaction is stopped by adding to the sample 3μl of TURBO DNase and incubating for 15 min. The RNA template was then purified according to the following protocol:

5 *In vitro transcription:*

	Sample	**Control**
MilliQ water (nuclease-free)	5μl	6μl
Ribonucleotides mix	24μl	8μl
10X reaction buffer	6μl	2μl
Recombinant plasmid	19μl	0μl
pTri-Xef control DNA	0μl	2μl
T7 enzyme mix	6μl	2μl
Total volume	60μl	20μl

Protocol: Keep reagents on ice except for the buffer. Prepare the reaction at room temperature and incubate for 3 hours at 37°C. Seal cap with parafilm.

1. Add 90μl nuclease-free water and 75μl LiCl precipitation solution to the reaction. Mix and chill at -20°C for at least 1 hour.
2. Centrifugate 4°C for 15 min at maximum speed.
3. Remove supernatant, wash the pellet with 1ml 70% ethanol, and centrifuge again.
4. Carefully remove the ethanol, let it dry out, and resuspend in 15μl nuclease-free water. Store at -80°C.

The RNA transcription product can be checked on a 4% a 1.2% RNA denaturing agarose gel (TAE/formamide).

6 *Final construct annealing*

The RNA transcript is finally annealed with the DNA handles to form a hairpin with two dsDNA/RNA heterohandles with biotin and digoxigenin tags at each end. The annealing reaction is done according to the reaction:

6 *Annealing reaction:*	
Annealing buffer	80μl
Handle A	9μl
Handle B	5μl
RNA transcript	1μl
Total volume	95μl

Protocol: 85°C for 10 min, 62°C for 90 min, 52°C for 90 min, ramp to 10°C in 10 min.

6 *Annealing buffer:*	
Fomamide	800μl
EDTA (0.5M, pH 8.0)	2μl
Pipes (1M, pH 6.3)	40μl
NaCl (5M)	80μl
Total volume	922μl

The resulting hairpin can be stored at -80°C to preserve it from degradation.

The above synthesis protocol has been repeated for all six RNA hairpins in Table D.3.

E

BEL AND HYSTERESIS CORRELATION IN RNA: CONTROL TESTS

We performed tests on different control sequences to validate the correlation between loop-BEL and the measured hysteresis of RNA (Sec.9.2). First, starting from the original RNA sequence, for a given stem-loops size, L (in the range $L = [8, 28]$), we generated an equivalent 2027bp hairpin by randomly shuffling sequence segments of size L. Thus, a different shuffled sequence is obtained for each of L. For each of these sequences, we computed the loop-BEL and the maximum average rolling correlation $\langle r_w \rangle_{\mathrm{UR}}^{\mathrm{Max}}$ (shift $s = 0$) with respect to the hysteresis between unfolding and refolding (Eq.(9.11) with $\alpha\beta = \mathrm{UR}$) for a window of size $w = 100$. Fig.E.1A shows the results of this analysis (orange circles) along with the results obtained for the original sequence (blue squares). As discussed in the main text (see Fig.9.13), the correlation of the original sequence increases with the stem-loops size L and has a maximum at $\langle r_w \rangle_{\mathrm{UR}}^{\mathrm{Max}} \simeq 0.25$ for $L \approx 18 - 22$ bases. On the contrary, the correlation for the shuffled sequences varies in the range $[-0.06 : 0.07]$ with no apparent trend. As a second control, we generated three random RNA hairpins with the same GC content as the original sequence. Then, we computed the loop-BEL and the average rolling correlation with the hysteresis for the case $L = 20$ (which roughly corresponds to the maximum correlation observed). The results are shown in Fig.E.1B. Analogously to the previous

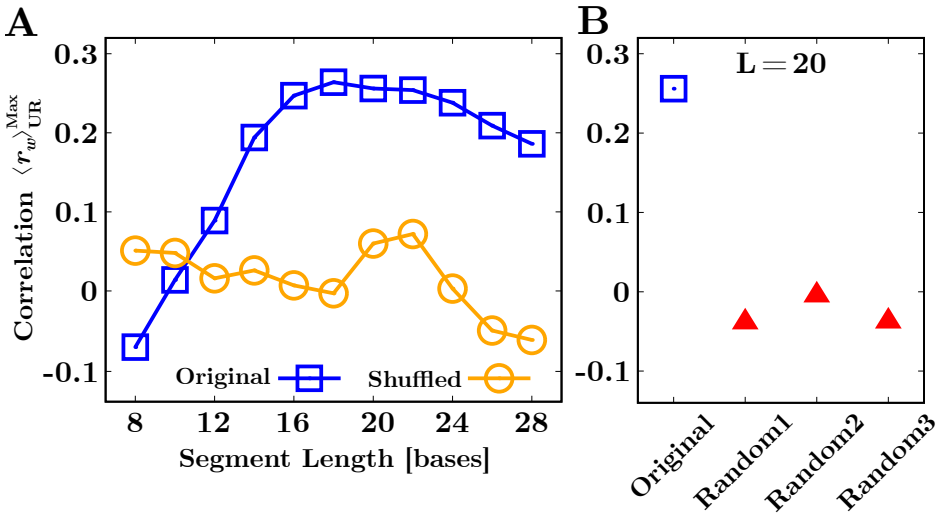


Figure E.1: Maximum average rolling correlation $\langle r_w \rangle_{\mathrm{UR}}^{\mathrm{Max}}$ (shift $s = 0$) between hysteresis and loop-BEL for different control sequences. **(A)** Average rolling correlation computed for the shuffled sequences (orange circles) and the original sequence (blue squares) for each value of $L = [8, 28]$. **(B)** Average rolling correlation of the hysteresis with the Loop-BEL for the random sequences (see text) at the maximally correlated case $L = 20$. The value for the original sequence (blue square) is also reported for a direct comparison.

case, the random sequences (red triangles) do not appear to be correlated with the hysteresis. The correlation falls in the range $[-0.04 : -0.005]$ so that the average of $\langle r_w \rangle_{\mathrm{UR}}^{\mathrm{Max}}$ over the random sequences is roughly equal to -0.02 which magnitude (in absolute number) is ten times lower than the value of the original sequence ($\simeq 0.25$).

The normalized correlation (as defined by the Pearson coefficient) between the hysteresis and the loop-BEL is sensitive to several factors. First, the hysteresis landscape computed in Eq.(9.11) is much more accurate in the regions where the irreversibility is large, i.e. where the difference between the unzipping and rezipping FDCs is large. When the hysteresis is small, i.e. along the reversible regions, Brownian fluctuations and instrumental effects reduce the signal-to-noise ratio of the

measured correlation of the hysteresis landscape with the loop-BEL. In the reversible regions, thermal (Brownian) fluctuations in the unzipping and the rezipping FDCs mask the correlation between the (low) hysteresis and the loop-BEL. The loop-BEL is noiseless by construction, whereas the computed hysteresis is not, so the noise dominates the correlation along the reversible regions. Moreover, instrumental effects are also detrimental in estimating such a correlation in the reversible regions. Slight misalignments between the experimental trajectories produce a spurious contribution to the measured hysteresis that is comparable with that due to the actual irreversibility. Finally, correlation measurements require matching the experimental measure of the hysteresis profile and the loop-BEL, further reducing correlation estimates. These sources of error render the Pearson coefficient in the reversible regions inaccurate.

F

ELASTIC PROPERTIES MEASURE IN TWO STATES SYSTEMS

The accurate measure of the elastic properties of ssNAs is key to characterizing their mechanical properties and understanding their folding mechanisms. In unzipping experiments of short NAs (typically of a few tens of bp) with two states (folded and unfolded), the measured FDC exhibits one force rip corresponding to the complete denaturation (hybridization) of the stem. In this case, the ssNA elastic properties can be measured using two independent methods: the force-jump method and the two branches method [108, 208]. In what follows, we describe these two approaches.

F.1 THE FORCE-JUMP METHOD

In unzipping a two states hairpin (folded and unfolded), the FDC exhibits a sudden force jump, $\Delta f = f_F - f_U$, when the hairpin changes conformation (from folded to unfolded and vice-versa). At the same time, the

corresponding total trap-pipette distance, λ, remains constant ($\Delta\lambda = 0$). Therefore, from Eq.(5.15), we can write

$$\begin{aligned} \lambda(f_F) &= \lambda(f_U) \\ x_b(f_F) + x_h(f_F) + x_d(f_F) &= x_b(f_U) + x_h(f_U) + x_{ss}(f_U)\,. \end{aligned} \tag{F.1}$$

Let us now divide the force jump, Δf, by the effective stiffness of the folded branch in Eq.(5.16). Thus, we get

$$\begin{aligned} \frac{\Delta f}{k^F_{\text{eff}}} &= \Delta f \left(\frac{1}{k_h} + \frac{1}{k_b} + \frac{1}{k_d}\right) = \\ &= \Delta x_h + \Delta x_b + \Delta x_d = \\ &= [x_h(f_F) - x_h(f_U)] + [x_b(f_F) - x_b(f_U)] + [x_d(f_F) - x_d(f_U)]\,. \end{aligned} \tag{F.2}$$

Finally, by substituting Eq.(F.2) into Eq.(F.1), one gets

$$x_{ss}(f_U) = \frac{\Delta f}{k^F_{\text{eff}}} + x_d(f_F)\,, \tag{F.3}$$

which allows for obtaining the ssNA elastic response from unfolding and folding force-jump measurements.

F.2 THE TWO BRANCHES METHOD

When hysteresis effects between unfolding and folding processes are large enough, both the folded (dsNA) and the unfolded (ssNA) branches can be measured from pulling experiments for an extensive range of forces. In this case, the effective stiffnesses $k^F_{\text{eff}}(f)$ and $k^U_{\text{eff}}(f)$ can be measured for the respective FDC branches. Notice that $k^F_{\text{eff}}(f) > k^U_{\text{eff}}(f)$ and that the range of forces where the stiffnesses can be measured increases

with the hysteresis (i. e., with decreasing temperature, increasing salt concentration and loop size). Thus, from Eq.(5.16), we can write

$$\frac{1}{k_{\text{eff}}^{U}(f)} - \frac{1}{k_{\text{eff}}^{F}(f)} = \frac{1}{k_{\text{ss}}(f)} - \frac{1}{k_{\text{d}}(f)} \,. \tag{F.4}$$

The stiffness of the ssNA is modeled according to the WLC and is given by Eq.(5.20). As discussed in Sec.5.3, the extension upon orienting the double helix is modeled as a dipole of length equal to the helix diameter, $d = 2\text{nm}$. By recalling that $k^{-1} = \partial x / \partial f$, from Eq.(5.18) we get

$$\frac{1}{k_{\text{d}}(f)} = \frac{d^2}{k_{\text{B}}T} \left[-\frac{1}{\sinh^2\left(\frac{fd}{k_{\text{B}}T}\right)} + \left(\frac{fd}{k_{\text{B}}T}\right)^2 \right] . \tag{F.5}$$

G

THE BAYESIAN CLASSIFIER

To model the RNA unzipping phenomenology at low temperatures, we used the Bayesian network approach (mixture hierarchical Bayesian model) developed in [235]. This has two advantages. On the one hand, using latent state variables in the model allows for retrieving posterior distributions for the state of each data point, permitting a probabilistic soft clustering of each unfolding trace, i.e. we can assign to each point a probability of being misfolded or native. On the other hand, the use of appropriate likelihood functions in the model gives a range of useful physical parameters, such as the mode and scale parameter of the rupture force distribution of each state (related to the average and the variance of the force distribution), or the weight of each state with respect to the total population. In what follows, we describe the clustering model resumed in Fig.G.1.

Let us recall that Bayesian network models posit that the prior distributions of the parameters to be estimated are known. Given these prior parameters, the likelihood function to observe each data point is also known. The estimation of the model parameters is then obtained by maximizing the posterior distribution of the model, given by the Bayes theorem,

$$posterior \propto Likelihood \times prior\,. \tag{G.1}$$

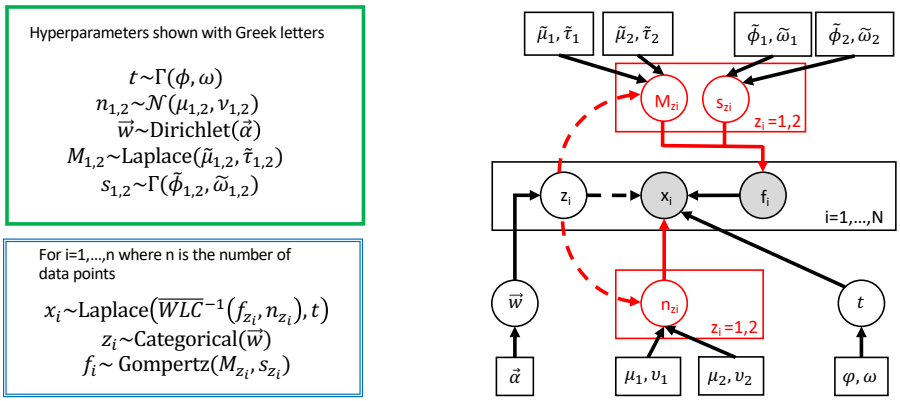


Figure G.1: The Bayesian clustering algorithm. **(Left)** Specification of the prior (in green) and likelihood (in blue) functions used. The hyper-parameters are indicated by Greek letters. Γ stands for the gamma distribution, $\mathcal{N}$ for the normal distribution. **(Right)** Probabilistic graph view of the Bayesian network. Misfolded and native states are represented by the superscript $1, 2$ and are encoded in the latent variable $z_i = 1, 2$. We highlight in red the part of the model which depends on z_i: the rupture force distribution through its mode M and scale s parameters and the number of monomers released within an unfolding event n.

The model's input data are the pairs (f, x), which characterize the rupture force and the released extension of each unfolding event in the forward process. The released extension (x) in an unfolding event depends on the molecule's initial folded state (through the number of released bp, n) and the rupture force as $x(f, n) \approx n d_b \hat{x}(f)$, where $\hat{x}(f)$ is the extension of a single ssRNA monomer and has been modeled by inverting the WLC in Eq.(5.19) (see Sec.5.3 and Appendix A). We assumed this relation holds in the presence of experimental noise, which

can be considered Laplace distributed around 0 with precision t. Thus, for each datapoint (f_i, x_i), one can write

$$x_i - n_{z_i} \cdot d_b \cdot \hat{x}(f_i) \sim \text{Laplace}(0, t)\,, \tag{G.2}$$

where the dependence on the number of released bp (n) for the i-th unfolding event is introduced through the (categorical) latent variable $z_i = 1, 2$ that account for the initial system state ($z = 1$ for the native and $z = 2$ for misfolding). Moreover, we posit that the rupture force distribution also depends on z. Thus, we assumed that rupture forces are Gompertz distributed with mode M_{z_i} and scale $1/s_{z_i}$ depending on the initial state, z_i (native or misfolded).

Therefore, the overall likelihood of observing the experimental point (f_i, x_i) is then:

$$\textit{Likelihood} = p\left(x(f_i, n_{z_i}) - x_i \middle| 0, t\right) \times p\left(f_i \middle| M_{z_i}, s_{z_i}\right) \times p\left(z_i \middle| \vec{w}\right)\,. \tag{G.3}$$

The first term in the r.h.s of Eq.(G.3) is given by Eq.(G.2), being $x(f_i, n_{z_i}) \equiv n_{z_i} \cdot d_b \cdot \hat{x}(f_i)$ The second term is given by the above-mentioned Gompertz likelihood. Finally, the third term represents the likelihood of the latent variable z_i given a weight vector $\vec{w} = (w_1, w_2)$ whose components give the average occupancy of each state. For z_i, we used the standard conjugate pair of a Categorical distribution for the likelihood p combined with a Dirichlet prior for $\vec{w}$.

To complete the model, we defined appropriate priors for the parameters to infer, namely $n_{(1,2)}$, $M_{(1,2)}$, $s_{(1,2)}$, t, and $\vec{w}$. As already mentioned, we use for $\vec{w}$ a Dirichlet prior, while we parameterize t and $s_{(1,2)}$ with

gamma priors, $n_{(1,2)}$ with a normal prior, and finally $M_{(1,2)}$ with a Laplace prior. Thus, the overall prior is ultimately given by:

$$\begin{aligned} \textit{Prior} = & \, p(n_{z_i} | \, \mu_{z_i}, \nu_{z_i}) \times p\left(M_{z_i} | \, \tilde{\mu}_{z_i}, \tilde{\tau}_{z_i}\right) \times p\left(s_{z_i} | \, \tilde{\phi}_{z_i}, \tilde{\omega}_{z_i}\right) \times \\ & \times p(\vec{w} | \, \vec{\alpha}) \times p(t | \, \phi, \omega) \, , \end{aligned} \tag{G.4}$$

where the model hyper-parameters are made explicit and are given by the greek variables $\mu_{(1,2)}$, $\nu_{(1,2)}$, $\tilde{\mu}_{(1,2)}$, $\tilde{\mu}_{(1,2)}$, $\tilde{\tau}_{(1,2)}$, $\tilde{\omega}_{(1,2)}$, α, ϕ, and ω.

Given the likelihood function (Eq.(G.3)) and the priors distributions (Eq.(G.4)), the posterior distribution of the parameters is computed using the Bayes theorem in Eq.(G.1), which gives

$$p(\{z_i\}_{i=1}^N; n_{(1,2)}; w_{(1,2)}; M_{(1,2)}; s_{(1,2)}; \sigma) \propto \textit{Likelihood} \times \textit{prior} \, , \tag{G.5}$$

where we defined for convenience $\sigma = 1/t$. Posterior distributions were obtained running at least 3 Monte Carlo Markov Chains (MCMC) with a burn-in of 1000 iterations and followed by 5000 iterations. We ran the usual convergence and diagnostics test for MCMCs (Gelmann, chain intercorrelation coefficient) and visually inspected the MCMC noise term. We always took the median of the posterior distribution of interest for point estimates (e.g., n_1, n_2).

BIBLIOGRAPHY

[1] *Biophysical Society*, 2022. http://www.biophysics.org.

[2] J. H. McMasters, "The flight of the bumblebee and related myths of entomological engineering: bees help bridge the gap between science and engineering," *Am. Sci.*, vol. 77, no. 2, pp. 164–169, 1989.

[3] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, "Scale-free correlations in starling flocks," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 26, pp. 11865–11870, 2010.

[4] R. Dixit, J. L. Ross, Y. E. Goldman, and E. L. Holzbaur, "Differential regulation of dynein and kinesin motor proteins by tau," *Science*, vol. 319, no. 5866, pp. 1086–1089, 2008.

[5] R. Sorkin, G. Bergamaschi, D. Kamsma, G. Brand, E. Dekel, Y. Ofir-Birin, A. Rudik, M. Gironella, F. Ritort, N. Regev-Rudzki, *et al.*, "Probing cellular mechanics with acoustic force spectroscopy," *Mol. Biol. Cell*, vol. 29, no. 16, pp. 2005–2011, 2018.

[6] M. Piccolino, "Luigi galvani and animal electricity: two centuries after the foundation of electrophysiology," *Trends Neurosci.*, vol. 20, no. 10, pp. 443–448, 1997.

[7] A. Fick, *Gesammelte schriften*, vol. 1. O. Stahel, 1903.

[8] W. C. Röntgen, "On a new kind of rays," *Science*, vol. 3, no. 59, pp. 227–231, 1896.

[9] E. Schrodinger, *What is life? The physical aspect of the living cell*. At the University Press, 1951.

[10] J. D. Watson and F. H. Crick, "The structure of dna," in *Cold Spring Harbor symposia on quantitative biology*, vol. 18, pp. 123–131, Cold Spring Harbor Laboratory Press, 1953.

[11] O. Kratky and G. Porod, "Röntgenuntersuchung gelöster fadenmoleküle," *Recl. Trav. Chim. Pays-Bas*, vol. 68, no. 12, pp. 1106–1122, 1949.

[12] S. B. Smith, L. Finzi, and C. Bustamante, "Direct mechanical measurements of the elasticity of single dna molecules by using magnetic beads," *Science*, vol. 258, no. 5085, pp. 1122–1126, 1992.

[13] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, "Entropic elasticity of λ-phage dna," *Science*, vol. 265, no. 5178, pp. 1599–1600, 1994.

[14] J. M. H. Casades, *Statistical and Thermodynamical Properties of DNA Unzipping Experiments with Optical Tweezers*. PhD thesis, 2010.

[15] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, "Funnels, pathways, and the energy landscape of protein folding: a synthesis," *Proteins*, vol. 21, pp. 167–195, Mar 1995.

[16] B. Aabert, A. L. J. Johnson, M. Raff, K. Roberts, and P. Walter, "Molecular biology of the cell. fourth," *New York: Garland Science*, 2002.

[17] S. W. Englander, L. Mayne, and M. M. Krishna, "Protein folding and misfolding: mechanism and principles," *Q. Rev. Biophys.*, vol. 40, no. 4, pp. 1–41, 2007.

[18] S. W. Englander and L. Mayne, "The nature of protein folding pathways," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 45, pp. 15873–15880, 2014.

[19] G. Felsenfeld and H. T. Miles, "The physical and chemical properties of nucleic acids," *Annu. Rev. Biochem.*, vol. 36, no. 1, pp. 407–448, 1967.

[20] M. H. Shamsi and H.-B. Kraatz, "Interactions of metal ions with dna and some applications," *Inorg Organomet Polym Mater*, vol. 23, no. 1, pp. 4–23, 2013.

[21] C. R. Cantor and P. R. Schimmel, *Biophysical chemistry: Part II: Techniques for the study of biological structure and function*. Macmillan, 1980.

[22] J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 95, no. 4, pp. 1460–1465, 1998.

[23] T. Xia, J. SantaLucia Jr, M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and D. H. Turner, "Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with watson-crick base pairs," *Biochemistry*, vol. 37, no. 42, pp. 14719–14735, 1998.

[24] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3406–3415, 2003.

[25] K. Neupane, D. A. Foster, D. R. Dee, H. Yu, F. Wang, and M. T. Woodside, "Direct observation of transition paths during the folding of proteins and nucleic acids," *Science*, vol. 352, no. 6282, pp. 239–242, 2016.

[26] D. K. Treiber and J. R. Williamson, "Exposing the kinetic traps in rna folding," *Curr. Opin. Struct. Biol.*, vol. 9, no. 3, pp. 339–345, 1999.

[27] S. M. Johnson, R. L. Wiseman, Y. Sekijima, N. S. Green, S. L. Adamski-Werner, and J. W. Kelly, "Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: a focus on the transthyretin amyloidoses," *Acc. Chem. Res.*, vol. 38, no. 12, pp. 911–921, 2005.

[28] M. Manosas, I. Junier, and F. Ritort, "Force-induced misfolding in rna," *Phys. Rev. E*, vol. 78, no. 6, p. 061925, 2008.

[29] R. Russell, "Rna misfolding and the action of chaperones," *Front. Biosci.*, vol. 13, p. 1, 2008.

[30] M. L. Rodgers and S. A. Woodson, "A roadmap for rrna folding and assembly during transcription," *Trends Biochem. Sci.*, vol. 46, no. 11, pp. 889–901, 2021.

[31] A. Jain and R. D. Vale, "Rna phase transitions in repeat expansion disorders," *Nature*, vol. 546, no. 7657, pp. 243–247, 2017.

[32] L. Blaszczyk, W. Rypniewski, and A. Kiliszek, "Structures of rna repeats associated with neurological diseases," *Wiley Interdiscip. Rev. RNA*, vol. 8, no. 4, p. e1412, 2017.

[33] X. Zhao and K. Usdin, "(dys) function follows form: nucleic acid structure, repeat expansion, and disease pathology in fmr1 disorders," *Int. J. Mol. Sci.*, vol. 22, no. 17, p. 9167, 2021.

[34] C. Bustamante, J. Liphardt, and F. Ritort, "The nonequilibrium thermodynamics of small systems," *Phys. Today*, vol. 58, no. 7, p. 43, 2005.

[35] F. Ritort, "Nonequilibrium fluctuations in small systems: From physics to biology," *Adv. Chem. Phys.*, vol. 137, p. 31, 2008.

[36] A. Alemany and F. Ritort, "Fluctuation theorems in small systems: extending thermodynamics to the nanoscale," *Europhys. News*, vol. 41, no. 2, pp. 27–30, 2010.

[37] F. Ritort, "Single-molecule experiments in biological physics: methods and applications," *J. Condens. Matter Phys.*, vol. 18, no. 32, p. R531, 2006.

[38] K. C. Neuman and A. Nagy, "Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy," *Nat. Methods*, vol. 5, no. 6, p. 491, 2008.

[39] G. Binnig, C. F. Quate, and C. Gerber, "Atomic force microscope," *Phys. Rev. Lett.*, vol. 56, no. 9, p. 930, 1986.

[40] F. Moreno-Herrero, J. Colchero, J. Gomez-Herrero, and A. Baro, "Atomic force microscopy contact, tapping, and jumping modes for imaging biological samples in liquids," *Phys. Rev. E*, vol. 69, no. 3, p. 031915, 2004.

[41] T. Ando, "High-speed atomic force microscopy and its future prospects," *Biophys. Rev.*, vol. 10, no. 2, pp. 285–292, 2018.

[42] T. Ando, T. Uchihashi, and S. Scheuring, "Filming biomolecular processes by high-speed atomic force microscopy," *Chem. Rev.*, vol. 114, no. 6, pp. 3120–3188, 2014.

[43] T. Ando, "Directly watching biomolecules in action by high-speed atomic force microscopy," *Biophys. Rev.*, vol. 9, no. 4, pp. 421–429, 2017.

[44] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, "Structural transitions and elasticity from torque measurements on dna," *Nature*, vol. 424, no. 6946, pp. 338–341, 2003.

[45] S. B. Smith, Y. Cui, and C. Bustamante, "An optical-trap force transducer that operates by direct measurement of light momentum," in *Methods Enzymol.*, vol. 361, pp. 134–162, Elsevier, 2003.

[46] A. Ashkin, "Acceleration and trapping of particles by radiation pressure," *Phys. Rev. Lett.*, vol. 24, no. 4, p. 156, 1970.

[47] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu, "Observation of a single-beam gradient force optical trap for dielectric particles," *Opt. Lett.*, vol. 11, no. 5, pp. 288–290, 1986.

[48] A. Ashkin, G. Mourou, and D. Strickland, "The 2018 nobel prize in physics: A gripping and extremely exciting tale of light," *Curr. Sci.*, vol. 115, no. 10, p. 1844, 2018.

[49] A. Ashkin, K. Schütze, J. Dziedzic, U. Euteneuer, and M. Schliwa, "Force generation of organelle transport measured in vivo by an infrared laser trap," *Nature*, vol. 348, no. 6299, pp. 346–348, 1990.

[50] D. G. Grier, "A revolution in optical manipulation," *nature*, vol. 424, no. 6950, pp. 810–816, 2003.

[51] M. Padgett and R. Di Leonardo, "Holographic optical tweezers and their relevance to lab on chip devices," *Lab Chip*, vol. 11, no. 7, pp. 1196–1205, 2011.

[52] M. L. Juan, M. Righini, and R. Quidant, "Plasmon nano-optical tweezers," *Nat. Photon*, vol. 5, no. 6, pp. 349–356, 2011.

[53] O. M. Maragò, P. H. Jones, P. G. Gucciardi, G. Volpe, and A. C. Ferrari, "Optical trapping and manipulation of nanostructures," *Nat. Nanotechnol.*, vol. 8, no. 11, pp. 807–819, 2013.

[54] R. M. Robertson-Anderson, "Optical tweezers microrheology: from the basics to advanced techniques and applications," 2018.

[55] F. M. Fazal and S. M. Block, "Optical tweezers study life under tension," *Nat. Photon*, vol. 5, no. 6, pp. 318–321, 2011.

[56] I. Junier, A. Mossa, M. Manosas, and F. Ritort, "Recovery of free energy branches in single molecule experiments," *Phys. Rev. Lett.*, vol. 102, no. 7, p. 070602, 2009.

[57] I. Tinoco, P. T. Li, and C. Bustamante, "Determination of thermodynamics and kinetics of rna reactions by force," *Q. Rev. Biophys.*, vol. 39, no. 4, pp. 325–360, 2006.

[58] A. Bosco, J. Camunas-Soler, and F. Ritort, "Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions," *Nucleic Acids Res.*, vol. 42, no. 3, pp. 2064–2074, 2014.

[59] J. Camunas-Soler, M. Ribezzi-Crivellari, and F. Ritort, "Elastic properties of nucleic acids by single-molecule force spectroscopy," *Annu. Rev. Biophys.*, vol. 45, pp. 65–84, 2016.

[60] X. Viader-Godoy, C. Pulido, B. Ibarra, M. Manosas, and F. Ritort, "Cooperativity-dependent folding of single-stranded dna," *Phys. Rev. X*, vol. 11, no. 3, p. 031037, 2021.

[61] R. B. Best, E. Paci, G. Hummer, and O. K. Dudko, "Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules," *J. Phys. Chem. B*, vol. 112, no. 19, pp. 5968–5976, 2008.

[62] O. K. Dudko, G. Hummer, and A. Szabo, "Theory, analysis, and interpretation of single-molecule force spectroscopy experiments," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 105, no. 41, pp. 15755–15760, 2008.

[63] M. Rico-Pasto, A. Zaltron, S. J. Davis, S. Frutos, and F. Ritort, "Molten globule–like transition state of protein barnase measured with calorimetric force spectroscopy," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 119, no. 11, p. e2112382119, 2022.

[64] C. J. Bustamante and S. B. Smith, "Light-force sensor and method for measuring axial optical-trap forces from changes in light momentum along an optic axis," Nov. 7 2006. US Patent 7,133,132.

[65] C. J. Bustamante and S. B. Smith, "Optical beam translation device and method utilizing a pivoting optical fiber," Sept. 25 2007. US Patent 7,274,451.

[66] C. J. Bustamante and S. B. Smith, 2013. http://http://tweezerslab.unipr.it.

[67] M. Ribezzi-Crivellari, J. M. Huguet, and F. Ritort, "Counter-propagating dual-trap optical tweezers based on linear momentum conservation," *Rev. Sci. Instrum.*, vol. 84, no. 4, p. 043104, 2013.

[68] J. Gieseler, J. R. Gomez-Solano, A. Magazzù, I. P. Castillo, L. P. García, M. Gironella-Torrent, X. Viader-Godoy, F. Ritort, G. Pesce, A. V. Arzola, *et al.*, "Optical tweezers—from calibration to applications: a tutorial," *Adv. Opt. Photonics*, vol. 13, no. 1, pp. 74–241, 2021.

[69] S. De Lorenzo, M. Ribezzi-Crivellari, J. R. Arias-Gonzalez, S. B. Smith, and F. Ritort, "A temperature-jump optical trap for single-molecule manipulation," *Biophys. J.*, vol. 108, no. 12, pp. 2854–2864, 2015.

[70] S. De Lorenzo Ros, *Single molecule studies in a temperature-jump optical trap*. PhD thesis, 2015.

[71] T. L. Hill, *Thermodynamics of small systems*. Courier Corporation, 1994.

[72] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines," *Rep. Prog. Phys.*, vol. 75, no. 12, p. 126001, 2012.

[73] D. J. Evans and D. J. Searles, "The fluctuation theorem," *Adv. Phys.*, vol. 51, no. 7, pp. 1529–1585, 2002.

[74] J. M. Huguet, M. Ribezzi-Crivellari, C. V. Bizarro, and F. Ritort, "Derivation of nearest-neighbor DNA parameters in magnesium from single molecule experiments," *Nucleic Acids Res.*, vol. 45, no. 22, pp. 12921–12931, 2017.

[75] P. Rissone, C. V. Bizarro, and F. Ritort, "Stem–loop formation drives rna folding in mechanical unzipping experiments," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 119, no. 3, 2022.

[76] E. A. Shank, C. Cecconi, J. W. Dill, S. Marqusee, and C. Bustamante, "The folding cooperativity of a protein is controlled by its chain topology," *Nature*, vol. 465, no. 7298, pp. 637–640, 2010.

[77] K. Hayashi, H. Ueno, R. Iino, and H. Noji, "Fluctuation theorem applied to f 1-atpase," *Phys. Rev. Lett.*, vol. 104, no. 21, p. 218103, 2010.

[78] G. Hummer and A. Szabo, "Free energy reconstruction from nonequilibrium single-molecule pulling experiments," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, no. 7, pp. 3658–3661, 2001.

[79] A. N. Gupta, A. Vincent, K. Neupane, H. Yu, F. Wang, and M. T. Woodside, "Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements," *Nat. Phys.*, vol. 7, no. 8, pp. 631–634, 2011.

[80] A. Alemany, A. Mossa, I. Junier, and F. Ritort, "Experimental free-energy measurements of kinetic molecular states using fluctuation theorems," *Nat. Phys.*, vol. 8, no. 9, p. 688, 2012.

[81] G. E. Crooks, "Path-ensemble averages in systems driven far from equilibrium," *Phys. Rev. E*, vol. 61, no. 3, p. 2361, 2000.

[82] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, vol. 78, no. 14, p. 2690, 1997.

[83] C. H. Bennett, "Efficient estimation of free energy differences from Monte Carlo data," *J. Comput. Phys.*, vol. 22, no. 2, pp. 245–268, 1976.

[84] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, "Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods," *Phys. Rev. Lett.*, vol. 91, no. 14, p. 140601, 2003.

[85] M. Palassini and F. Ritort, "Improving free-energy estimates from unidirectional work measurements: theory and experiment," *Phys. Rev. Lett.*, vol. 107, no. 6, p. 060601, 2011.

[86] F. M. Ytreberg, R. H. Swendsen, and D. M. Zuckerman, "Comparison of free energy methods for molecular systems," *J. Chem. Phys.*, vol. 125, no. 18, p. 184114, 2006.

[87] F. Crick, "Central dogma of molecular biology," *Nature*, vol. 227, no. 5258, pp. 561–563, 1970.

[88] C. R. Calladine and H. Drew, *Understanding DNA: the molecule and how it works*. Academic press, 1997.

[89] W. Saenger, *Principles of nucleic acid structure*. Springer, 1984.

[90] L. Pray, "Discovery of dna structure and function: Watson and crick," *Nature Education*, vol. 1, no. 1, 2008.

[91] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, "Base-stacking and base-pairing contributions into thermal stability of the dna double helix," *Nucleic Acids Res.*, vol. 34, no. 2, pp. 564–574, 2006.

[92] J. Lipfert, G. M. Skinner, J. M. Keegstra, T. Hensgens, T. Jager, D. Dulin, M. Köber, Z. Yu, S. P. Donkers, F.-C. Chou, *et al.*, "Double-stranded rna under force and torque: similarities to and striking differences from double-stranded dna," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 43, pp. 15408–15413, 2014.

[93] H. DeVoe and I. Tinoco Jr, "The stability of helical polynucleotides: base contributions," *J. Mol. Biol.*, vol. 4, no. 6, pp. 500–517, 1962.

[94] D. M. Crothers and B. H. Zimm, "Theory of the melting transition of synthetic polynucleotides: evaluation of the stacking free energy," *J. Mol. Biol.*, vol. 9, no. 1, pp. 1–9, 1964.

[95] K. J. Breslauer, R. Frank, H. Blöcker, and L. A. Marky, "Predicting DNA duplex stability from the base sequence," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 83, no. 11, pp. 3746–3750, 1986.

[96] R. F. Goldstein and A. S. Benight, "How many numbers are required to specify sequence-dependent properties of polynucleotides?," *Biopolymers*, vol. 32, no. 12, pp. 1679–1693, 1992.

[97] P. Licinio and J. C. O. Guerra, "Irreducible representation for nucleotide sequence physical properties and self-consistency of nearest-neighbor dimer sets," *Biophys. J.*, vol. 92, no. 6, pp. 2000–2006, 2007.

[98] D. M. Gray and I. Tinoco Jr, "A new approach to the study of sequence-dependent properties of polynucleotides," *Biopolymers*, vol. 9, no. 2, pp. 223–244, 1970.

[99] S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner, "Improved free-energy parameters for predictions of RNA duplex stability," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 83, no. 24, pp. 9373–9377, 1986.

[100] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Müller, D. H. Mathews, and M. Zuker, "Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 91, no. 20, pp. 9218–9222, 1994.

[101] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure," *J. Mol. Biol.*, vol. 288, no. 5, pp. 911–940, 1999.

[102] A. Severino, A. M. Monge, P. Rissone, and F. Ritort, "Efficient methods for determining folding free energies in single-molecule pulling experiments," *J. Stat. Mech.: Theory Exp*, vol. 2019, no. 12, p. 124001, 2019.

[103] N. Forns, S. de Lorenzo, M. Manosas, K. Hayashi, J. M. Huguet, and F. Ritort, "Improving signal/noise resolution in single-molecule experiments using molecular constructs with short handles," *Biophys. J.*, vol. 100, no. 7, pp. 1765–1774, 2011.

[104] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block, "Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid," *Science*, vol. 314, no. 5801, pp. 1001–1004, 2006.

[105] C. Bustamante, J. Marko, E. Siggia, and S. Smith, "Entropic elasticity of λ-phage DNA," *Proc. Nati. Acad. Sci USA*, vol. 88, p. 10009, 1991.

[106] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, "Stretching dna with optical tweezers," *Biophys. J.*, vol. 72, no. 3, pp. 1335–1346, 1997.

[107] R. Petrosyan, "Improved approximations for some polymer extension models," *Rheologica Acta*, vol. 56, no. 1, pp. 21–26, 2017.

[108] A. Alemany and F. Ritort, "Determination of the elastic properties of short ssdna molecules by mechanically folding and unfolding dna hairpins," *Biopolymers*, vol. 101, no. 12, pp. 1193–1199, 2014.

[109] A. Bosco, J. Camunas-Soler, and F. Ritort, "Elastic properties and secondary structure formation of single-stranded dna at monovalent and divalent salt conditions," *Nucleic Acids Res.*, vol. 42, no. 3, pp. 2064–2074, 2013.

[110] C. V. Bizarro, A. Alemany, and F. Ritort, "Non-specific binding of Na+ and Mg2+ to RNA determined by force spectroscopy methods," *Nucleic Acids Res.*, vol. 40, no. 14, pp. 6922–6935, 2012.

[111] J. Abels, F. Moreno-Herrero, T. Van der Heijden, C. Dekker, and N. H. Dekker, "Single-molecule measurements of the persistence length of double-stranded rna," *Biophys. J.*, vol. 88, no. 4, pp. 2737–2744, 2005.

[112] J. H. Van't Hoff, *Etudes de dynamique chimique*, vol. 1. Muller, 1884.

[113] S. Arrhenius, "Über die reaktionsgeschwindigkeit bei der inversion von rohrzucker durch säuren," *Zeitschrift für physikalische Chemie*, vol. 4, no. 1, pp. 226–248, 1889.

[114] G. I. Bell, "Models for the specific adhesion of cells to cells: a theoretical framework for adhesion mediated by reversible bonds between cell surface molecules.," *Science*, vol. 200, no. 4342, pp. 618–627, 1978.

[115] E. Evans and K. Ritchie, "Dynamic strength of molecular adhesion bonds," *Biophys. J.*, vol. 72, no. 4, pp. 1541–1555, 1997.

[116] A. Alemany, B. Rey-Serra, S. Frutos, C. Cecconi, and F. Ritort, "Mechanical folding and unfolding of protein barnase at the single-molecule level," *Biophys. J.*, vol. 110, no. 1, pp. 63–74, 2016.

[117] A. Alemany and F. Ritort, "Force-dependent folding and unfolding kinetics in dna hairpins reveals transition-state displacements along a single pathway," *J. Phys. Chem. Lett.*, vol. 8, no. 5, pp. 895–900, 2017.

[118] A. Alemany i Arias, *Dynamic force spectroscopy and folding kinetics in molecular systems*. PhD thesis, 2014.

[119] M. Rico-Pasto, A. Alemany, and F. Ritort, "Force-dependent folding kinetics of single molecules with multiple intermediates and pathways," *J. Phys. Chem. Lett.*, vol. 13, no. 4, pp. 1025–1032, 2022.

[120] P. R. Connelly and J. A. Thomson, "Heat capacity changes and hydrophobic interactions in the binding of fk506 and rapamycin to the fk506 binding protein.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 89, no. 11, pp. 4781–4785, 1992.

[121] I. Rouzina and V. A. Bloomfield, "Heat capacity effects on the melting of dna. 1. general aspects," *Biophys. J.*, vol. 77, no. 6, pp. 3242–3251, 1999.

[122] I. Rouzina and V. A. Bloomfield, "Heat capacity effects on the melting of dna. 2. analysis of nearest-neighbor base pair effects," *Biophys. J.*, vol. 77, no. 6, pp. 3252–3255, 1999.

[123] P. Wu, S.-i. Nakano, and N. Sugimoto, "Temperature dependence of thermodynamic properties for dna/dna and rna/dna duplex formation," *Eur. J. Biochem.*, vol. 269, no. 12, pp. 2821–2830, 2002.

[124] P. J. Mikulecky and A. L. Feig, "Heat capacity changes associated with nucleic acid folding," *Biopolymers*, vol. 82, no. 1, pp. 38–58, 2006.

[125] A. Dragan, P. Privalov, and C. Crane-Robinson, "Thermodynamics of dna: heat capacity changes on duplex unfolding," *Eur. Biophys. J.*, vol. 48, no. 8, pp. 773–779, 2019.

[126] M. C. Williams, J. R. Wenner, I. Rouzina, and V. A. Bloomfield, "Entropy and heat capacity of dna melting from temperature dependence of single molecule stretching," *Biophys. J.*, vol. 80, no. 4, pp. 1932–1939, 2001.

[127] O. Gotoh and Y. Tagashira, "Stabilities of nearest-neighbor doublets in double-helical dna determined by fitting calculated melting profiles to observed profiles," *Biopolymers*, vol. 20, no. 5, pp. 1033–1042, 1981.

[128] A. V. Vologodskii, B. Amirikyan, Y. L. Lyubchenko, and M. Frank-Kamenetskii, "Allowance for heterogeneous stacking in the dna helix-coil transition theory," *J. Biomol. Struct. Dyn.*, vol. 2, no. 1, pp. 131–148, 1984.

[129] S. G. Delcourt and R. Blake, "Stacking energies in dna," *J. Biol. Chem.*, vol. 266, no. 23, pp. 15160–15169, 1991.

[130] M. J. Doktycz, R. F. Goldstein, T. M. Paner, F. J. Gallo, and A. S. Benight, "Studies of dna dumbbells. i. melting curves of 17 dna dumbbells with different duplex stem sequences linked by t4 end-loops: Evaluation of the nearest-neighbor stacking interactions in dna," *Biopolymers*, vol. 32, no. 7, pp. 849–864, 1992.

[131] J. SantaLucia, H. T. Allawi, and P. A. Seneviratne, "Improved nearest-neighbor parameters for predicting dna duplex stability," *Biochem.*, vol. 35, no. 11, pp. 3555–3562, 1996.

[132] N. Sugimoto, S.-i. Nakano, M. Yoneyama, and K.-i. Honda, "Improved thermodynamic parameters and helix initiation factor to predict stability of dna duplexes," *Nucleic Acids Res.*, vol. 24, no. 22, pp. 4501–4505, 1996.

[133] J. M. Huguet, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort, "Single-molecule derivation of salt dependent base-pair free energies in DNA," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 35, pp. 15431–15436, 2010.

[134] S. B. Smith, Y. Cui, and C. Bustamante, "Overstretching b-dna: the elastic response of individual double-stranded and single-stranded dna molecules," *Science*, vol. 271, no. 5250, pp. 795–799, 1996.

[135] M. W. Zemansky and R. H. Dittman, *Heat and thermodynamics*. American Association of Physics Teachers, 1998.

[136] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*, pp. 177–186, Springer, 2010.

[137] L. Bottou, "Stochastic gradient descent tricks," in *Neural networks: Tricks of the trade*, pp. 421–436, Springer, 2012.

[138] R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, and J. A. Walder, "Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations," *Biochemistry*, vol. 47, no. 19, pp. 5336–5353, 2008.

[139] C. E. Castro, F. Kilchherr, D.-N. Kim, E. L. Shiao, T. Wauer, P. Wortmann, M. Bathe, and H. Dietz, "A primer to scaffolded DNA origami," *Nat. Methods*, vol. 8, no. 3, p. 221, 2011.

[140] T.-G. Cha, J. Pan, H. Chen, J. Salgado, X. Li, C. Mao, and J. H. Choi, "A synthetic DNA motor that transports nanoparticles along carbon nanotubes," *Nat. Nanotechnol.*, vol. 9, no. 1, pp. 39–43, 2014.

[141] M. Hagiya, A. Konagaya, S. Kobayashi, H. Saito, and S. Murata, "Molecular robots with sensors and intelligence," *Acc. Chem. Res.*, vol. 47, no. 6, pp. 1681–1690, 2014.

[142] F. Wang, X. Liu, and I. Willner, "DNA switches: from principles to applications," *Angew. Chem. Int. Ed.*, vol. 54, no. 4, pp. 1098–1129, 2015.

[143] C. Jung, P. Allen, and A. Ellington, "A stochastic DNA walker that traverses a microparticle surface," *Nat. Nanotechnol.*, vol. 11, no. 2, pp. 157–163, 2016.

[144] H. H. Klump, J. Völker, and K. J. Breslauer, "Energy mapping of the genetic code and genomic domains: implications for code evolution and molecular Darwinism," *Q. Rev. Biophys.*, vol. 53, 2020.

[145] S.-J. Chen and K. A. Dill, "RNA folding energy landscapes," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 97, no. 2, pp. 646–651, 2000.

[146] X. Zhuang and M. Rief, "Single-molecule folding," *Curr. Opin. Struct. Biol.*, vol. 13, no. 1, pp. 88–97, 2003.

[147] S. A. Woodson, "Compact intermediates in rna folding: Annual reviews in biophysics," *Annual review of biophysics*, vol. 39, p. 61, 2010.

[148] D. U. Ferreiro, E. A. Komives, and P. G. Wolynes, "Frustration in biomolecules," *Q. Rev. Biophys.*, vol. 47, no. 4, pp. 285–363, 2014.

[149] S. W. Englander and L. Mayne, "The case for defined protein folding pathways," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 31, pp. 8253–8258, 2017.

[150] J. Camunas-Soler, A. Alemany, and F. Ritort, "Experimental measurement of binding energy, selectivity, and allostery using fluctuation theorems," *Science*, vol. 355, no. 6323, pp. 412–415, 2017.

[151] P. Sonar, L. Bellucci, A. Mossa, P. O. Heidarsson, B. B. Kragelund, and C. Cecconi, "Effects of ligand binding on the energy landscape of acyl-coa-binding protein," *Biophys. J.*, vol. 119, no. 9, pp. 1821–1832, 2020.

[152] Z.-J. Tan and S.-J. Chen, "RNA helix stability in mixed Na+/Mg2+ solution," *Biophys. J.*, vol. 92, no. 10, pp. 3615–3632, 2007.

[153] Z.-J. Tan and S.-J. Chen, "Nucleic acid helix stability: effects of salt concentration, cation valence and size, and chain length," *Biophys. J.*, vol. 90, no. 4, pp. 1175–1190, 2006.

[154] S. H. Cross and A. P. Bird, "CpG islands and genes," *Curr. Opin. Genet. Dev.*, vol. 5, no. 3, pp. 309–314, 1995.

[155] M. Esteller, "Epigenetics in cancer," *N. Engl. J. Med.*, vol. 358, no. 11, pp. 1148–1159, 2008.

[156] J. Lipfert, S. Doniach, R. Das, and D. Herschlag, "Understanding nucleic acid–ion interactions," *Annu. Rev. Biochem.*, vol. 83, pp. 813–841, 2014.

[157] M. Manosas, X. G. Xi, D. Bensimon, and V. Croquette, "Active and passive mechanisms of helicases," *Nucleic Acids Res.*, vol. 38, no. 16, pp. 5518–5526, 2010.

[158] N. R. Markham and M. Zuker, "DINAMelt web server for nucleic acid melting prediction," *Nucleic Acids Res.*, vol. 33, no. suppl_2, pp. W577–W581, 2005.

[159] N. Markham, M. Zuker, and J. Keith, "UNAFold: software for nucleic acid folding and hybridization," *Bioinformatics*, vol. 2, no. 453, pp. 3–31, 2008.

[160] M. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, "Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy," *Biophys. J.*, vol. 86, no. 4, pp. 2530–2537, 2004.

[161] J. M. Huguet, N. Forns, and F. Ritort, "Statistical properties of metastable intermediates in DNA unzipping," *Phys. Rev. Lett.*, vol. 103, no. 24, p. 248106, 2009.

[162] A. Montanari and M. Mézard, "Hairpin formation and elongation of biomolecules," *Phys. Rev. Lett.*, vol. 86, no. 10, p. 2178, 2001.

[163] M.-N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette, "Stretching single stranded dna, a model polyelectrolyte," *Phys. Rev. Lett.*, vol. 89, no. 24, p. 248102, 2002.

[164] X. Viader-Godoy, M. Manosas, and F. Ritort, "Sugar-pucker force-induced transition in single-stranded dna," *Int. J. Mol. Sci.*, vol. 22, no. 9, p. 4745, 2021.

[165] F. Schneider, N. Möritz, and H. Dietz, "The sequence of events during folding of a DNA origami," *Sci. Adv.*, vol. 5, no. 5, p. eaaw1412, 2019.

[166] W. J. Greenleaf, K. L. Frieda, D. A. Foster, M. T. Woodside, and S. M. Block, "Direct observation of hierarchical folding in single riboswitch aptamers," *Science*, vol. 319, no. 5863, pp. 630–633, 2008.

[167] M. T. Halma, D. B. Ritchie, T. R. Cappellano, K. Neupane, and M. T. Woodside, "Complex dynamics under tension in a high-efficiency frameshift stimulatory structure," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 116, no. 39, pp. 19500–19505, 2019.

[168] D. B. Ritchie and M. T. Woodside, "Probing the structural dynamics of proteins and nucleic acids with optical tweezers," *Curr. Opin. Struct. Biol.*, vol. 34, pp. 43–51, 2015.

[169] A. Schug and J. N. Onuchic, "From protein folding to protein function and biomolecular binding by energy landscape theory," *Curr. Opin. Pharmacol.*, vol. 10, no. 6, pp. 709–714, 2010.

[170] N. Goldenfeld and C. Woese, "Life is physics: evolution as a collective phenomenon far from equilibrium," *Annu. Rev. Condens. Matter Phys.*, vol. 2, no. 1, pp. 375–399, 2011.

[171] J. E. Darnell, *RNA: life's indispensable molecule*. Cold Spring Harbor Laboratory Press, 2011.

[172] W. Filipowicz, "Rna, the epicenter of genetic information: wirtten by john mattick and paulo amaral, crc press, boca raton, 2022," 2022.

[173] S. E. Butcher and A. M. Pyle, "The molecular interactions that stabilize rna tertiary structure: Rna motifs, patterns, and networks," *Acc. Chem. Res.*, vol. 44, no. 12, pp. 1302–1311, 2011.

[174] D. Herschlag, S. Bonilla, and N. Bisaria, "The story of rna folding, as told in epochs," *Cold Spring Harb. Perspect. Biol.*, vol. 10, no. 10, p. a032433, 2018.

[175] A. Pyle, "Metal ions in the structure and function of rna," *J. Biol. Inorg. Chem.*, vol. 7, no. 7, pp. 679–690, 2002.

[176] S. A. Woodson, "Metal ions and rna folding: a highly charged topic with a dynamic future," *Curr. Opin. Chem. Biol.*, vol. 9, no. 2, pp. 104–109, 2005.

[177] D. E. Draper, D. Grilley, and A. M. Soto, "Ions and rna folding," *Annu. Rev. Biophys.*, vol. 34, no. 1, pp. 221–243, 2005.

[178] J. C. Bowman, T. K. Lenz, N. V. Hud, L. D. Williams, *et al.*, "Cations in charge: magnesium ions in rna folding and catalysis," *Curr. Opin. Struct. Biol.*, vol. 22, no. 3, p. 262, 2012.

[179] D. K. Treiber and J. R. Williamson, "Beyond kinetic traps in rna folding," *Curr. Opin. Struct. Biol.*, vol. 11, no. 3, pp. 309–314, 2001.

[180] R. Russell, I. S. Millett, M. W. Tate, L. W. Kwok, B. Nakatani, S. M. Gruner, S. G. Mochrie, V. Pande, S. Doniach, D. Herschlag, *et al.*, "Rapid compaction during rna folding," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, no. 7, pp. 4266–4271, 2002.

[181] P. Brion and E. Westhof, "Hierarchy and dynamics of rna folding," *Annu. Rev. Biophys.*, vol. 26, no. 1, pp. 113–137, 1997.

[182] J. A. Cruz and E. Westhof, "The dynamic landscapes of rna architecture," *Cell*, vol. 136, no. 4, pp. 604–609, 2009.

[183] J. Mattick and I. Makunin, "Non-coding," *RNA. Hum. Mol. Genet.*, vol. 15, pp. R17–R29, 2006.

[184] A. P. Aalto and A. E. Pasquinelli, "Small non-coding rnas mount a silent revolution in gene expression," *Curr. Opin. Cell Biol.*, vol. 24, no. 3, pp. 333–340, 2012.

[185] M. Esteller, "Non-coding rnas in human disease," *Nat. Rev. Genet*, vol. 12, no. 12, pp. 861–874, 2011.

[186] M. Matsui and D. R. Corey, "Non-coding rnas as drug targets," *Nat. Rev. Drug Discov*, vol. 16, no. 3, pp. 167–179, 2017.

[187] P. J. Mikulecky and A. L. Feig, "Cold denaturation of the hammerhead ribozyme," *J. Am. Chem. Soc.*, vol. 124, no. 6, pp. 890–891, 2002.

[188] F. Iannelli, Y. Mamasakhlisov, and R. R. Netz, "Cold denaturation of RNA secondary structures with loop entropy and quenched disorder," *Phys. Rev. E*, vol. 101, no. 1, p. 012502, 2020.

[189] E. Loh, E. Kugelberg, A. Tracy, Q. Zhang, B. Gollan, H. Ewles, R. Chalmers, V. Pelicic, and C. M. Tang, "Temperature triggers immune evasion by neisseria meningitidis," *Nature*, vol. 502, no. 7470, pp. 237–240, 2013.

[190] G. Stefani and F. J. Slack, "Small non-coding rnas in animal development," *Nat. Rev. Mol*, vol. 9, no. 3, pp. 219–230, 2008.

[191] T. R. Mercer, M. E. Dinger, and J. S. Mattick, "Long non-coding rnas: insights into functions," *Nat. Rev. Genet*, vol. 10, no. 3, pp. 155–159, 2009.

[192] L. Ma, V. B. Bajic, and Z. Zhang, "On the classification of long non-coding rnas," *RNA Biol.*, vol. 10, no. 6, pp. 924–933, 2013.

[193] A. Fatica and I. Bozzoni, "Long non-coding rnas: new players in cell differentiation and development," *Nat. Rev. Genet*, vol. 15, no. 1, pp. 7–21, 2014.

[194] L. Statello, C.-J. Guo, L.-L. Chen, and M. Huarte, "Gene regulation by long non-coding rnas and its biological functions," *Nat. Rev. Mol*, vol. 22, no. 2, pp. 96–118, 2021.

[195] R. Seidel and C. Dekker, "Single-molecule studies of nucleic acid motors," *Curr. Opin. Struct. Biol.*, vol. 17, no. 1, pp. 80–86, 2007.

[196] X. Zhuang, "Single-molecule RNA science," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 34, pp. 399–414, 2005.

[197] E. A. Alemán, R. Lamichhane, and D. Rueda, "Exploring rna folding one molecule at a time," *Curr. Opin. Chem. Biol.*, vol. 12, no. 6, pp. 647–654, 2008.

[198] C. J. Bustamante, Y. R. Chemla, S. Liu, and M. D. Wang, "Optical tweezers in single-molecule biophysics," *Nat. Rev. Methods Primers*, vol. 1, no. 1, pp. 1–29, 2021.

[199] R. Y. Henley, S. Carson, and M. Wanunu, "Studies of rna sequence and structure using nanopores," *Prog. Mol. Biol. Transl. Sci.*, vol. 139, pp. 73–99, 2016.

[200] F. Bošković and U. F. Keyser, "Nanopore microscope identifies rna isoforms with structural colours," *Nat. Chem*, pp. 1–7, 2022.

[201] R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker, "Viennarna package 2.0," *Algorithms Mol. Biol.*, vol. 6, no. 1, pp. 1–14, 2011.

[202] M. Bon, C. Micheletti, and H. Orland, "Mcgenus: a monte carlo algorithm to predict rna secondary structures with pseudoknots," *Nucleic Acids Res.*, vol. 41, no. 3, pp. 1895–1900, 2013.

[203] S. Janssen and R. Giegerich, "The rna shapes studio," *Bioinformatics*, vol. 31, no. 3, pp. 423–425, 2015.

[204] M. R. Davison, R. Wu, V. G. Danna, and I. G. Godinez, "Uncovering novel rna viruses in permafrost," tech. rep., Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2020.

[205] A. Lemieux, G. A. Colby, A. J. Poulain, and S. Aris-Brosou, "Viral spillover risk increases with climate change in high arctic lake sediments," *Proc. Royal Soc. B*, vol. 289, no. 1985, p. 20221073, 2022.

[206] M. Rico-Pasto, I. Pastor, and F. Ritort, "Force feedback effects on single molecule hopping and pulling experiments," *J. Chem. Phys.*, vol. 148, no. 12, p. 123327, 2018.

[207] M. Rico-Pasto and F. Ritort, "Temperature-dependent elastic properties of dna," *Biophys. Rep.*, vol. 2, no. 3, p. 100067, 2022.

[208] X. Zhang, K. Halvorsen, C.-Z. Zhang, W. P. Wong, and T. A. Springer, "Mechanoenzymatic cleavage of the ultralarge vascular protein von willebrand factor," *Science*, vol. 324, no. 5932, pp. 1330–1334, 2009.

[209] P. Rissone and F. Ritort, "Nucleic acid thermodynamics derived from mechanical unzipping experiments," *Life*, vol. 12, no. 7, p. 1089, 2022.

[210] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, "Verification of the crooks fluctuation theorem and recovery of rna folding free energies," *Nature*, vol. 437, no. 7056, pp. 231–234, 2005.

[211] C. Jarzynski, "Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale," *Annu. Rev. Condens. Matter Phys.*, vol. 2, no. 1, pp. 329–351, 2011.

[212] K. J. Breslauer, J. M. Sturtevant, and I. Tinoco Jr, "Calorimetric and spectroscopic investigation of the helix-to-coil transition of a ribo-oligonucleotide: ra7u7," *J. Mol. Biol.*, vol. 99, no. 4, pp. 549–565, 1975.

[213] M. Petersheim and D. H. Turner, "Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with ccgg, ccggp, ccggap, accggp, ccggup, and accggup," *Biochem.*, vol. 22, no. 2, pp. 256–263, 1983.

[214] N. Sugimoto, R. Kierzek, S. M. Freier, and D. H. Turner, "Energetics of internal gu mismatches in ribooligonucleotide helixes," *Biochem.*, vol. 25, no. 19, pp. 5755–5759, 1986.

[215] J. D. Bryngelson and P. G. Wolynes, "Spin glasses and the statistical mechanics of protein folding," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 84, no. 21, pp. 7524–7528, 1987.

[216] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, "Theory of protein folding: the energy landscape perspective," *Annu. Rev. Phys. Chem.*, vol. 48, no. 1, pp. 545–600, 1997.

[217] M. Manosas, D. Collin, and F. Ritort, "Force-dependent fragility in rna hairpins," *Phys. Rev. Lett.*, vol. 96, no. 21, p. 218301, 2006.

[218] M. Schlierf and M. Rief, "Single-molecule unfolding force distributions reveal a funnel-shaped energy landscape," *Biophys. J.*, vol. 90, no. 4, pp. L33–L35, 2006.

[219] A. Mossa, M. Manosas, N. Forns, J. M. Huguet, and F. Ritort, "Dynamic force spectroscopy of dna hairpins: I. force kinetics and

free energy landscapes," *J. Stat. Mech. Theory Exp.*, vol. 2009, no. 02, p. P02060, 2009.

[220] O. D. Broekmans, G. A. King, G. J. Stephens, and G. J. Wuite, "Dna twist stability changes with magnesium (2+) concentration," *Phys. Rev. Lett.*, vol. 116, no. 25, p. 258102, 2016.

[221] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55. Courier Corporation, 1965.

[222] P. Borwein and T. Erdélyi, *Polynomials and polynomial inequalities*, vol. 161. Springer Science & Business Media, 1995.

[223] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, pp. 400–407, 1951.

[224] H. B. Curry, "The method of steepest descent for non-linear minimization problems," *Q. Appl. Math.*, vol. 2, no. 3, pp. 258–261, 1944.

[225] B. T. Polyak, *Introduction to optimization. optimization software*, vol. 1. 1987.

[226] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.

[227] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[228] Y. Nesterov, "A method for unconstrained convex minimization problem with the rate of convergence o (1/kˆ 2)," in *Doklady an ussr*, vol. 269, pp. 543–547, 1983.

[229] P. Achlioptas, "Stochastic gradient descent in theory and practice," 2019.

[230] A. Untergasser, H. Nijveen, X. Rao, T. Bisseling, R. Geurts, and J. A. Leunissen, "Primer3plus, an enhanced web interface to Primer3," *Nucleic Acids Res.*, vol. 35, no. suppl_2, pp. W71–W74, 2007.

[231] F. Bolivar, R. L. Rodriguez, M. C. Betlach, and H. W. Boyer, "Construction and characterization of new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9," *Gene*, vol. 2, no. 2, pp. 75–93, 1977.

[232] F. Bolivar, R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow, "Construction and characterization of new cloning vehicle. II. A multipurpose cloning system," *Gene*, vol. 2, no. 2, pp. 95–113, 1977.

[233] M. M. Baklanov, L. N. Golikova, and E. G. Malygin, "Effect on DNA transcription of nucleotide sequences upstream to T7 promoter," *Nucleic Acids Res.*, vol. 24, no. 18, pp. 3659–3660, 1996.

[234] J. Camuñas Soler, *Force-spectroscopy of small ligands binding to nucleic acids*. PhD thesis, 2015.

[235] P. Rissone, A. Severino, I. Pastor, and F. Ritort, "Rna cold misfolding," *[Manuscript submitted for publication]*.