UNIVERSITAT DE
BARCELONA

Biophysical properties of single-stranded DNA studied with single-molecule force spectroscopy

Xavier Viader Godoy
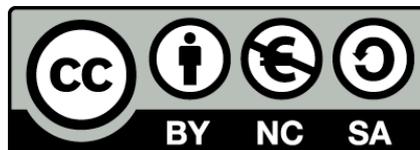
BIOPHYSICAL PROPERTIES OF SINGLE-STRANDED DNA STUDIED WITH SINGLE-MOLECULE FORCE SPECTROSCOPY

Autor:

Xavier Viader Godoy

Directors:

Dr. Fèlix Ritort Farran
Dr. Maria Manosas Castejón

UNIVERSITAT DE BARCELONA

BIOPHYSICAL PROPERTIES OF SINGLE-STRANDED DNA STUDIED WITH SINGLE-MOLECULE FORCE SPECTROSCOPY

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

Programa de doctorat en Nanociències

Autor: Xavier Viader Godoy

Director i Tutor: Fèlix Ritort Farran
Directora: Maria Mañosas Castejón

UNIVERSITAT DE
BARCELONA

El que res no costa res no val.
Refrany popular català

Com que hi ha tanta grip,
han hagut de clausurar la Universitat

Josep Pla, *El Quadern Gris*

AGRAÏMENTS

El camí que he recorregut fins arribar a escriure aquestes línies no ha sigut gens trivial. 6 anys de feina; mesos de motivació, seguits de mesos sense rumb. Setmanes de congressos i setmanes de baralles amb l'instrument. Dies llargs que es feien curts i dies curts que es feien llargs. Hores de profit i hores perdudes. Moltes i moltes hores, al cap i a la fi. Però, com diria el filòsof "al final, el que ens queden són les anècdotes". I d'anècdotes, aquests 6 anys en van plens. Anècdotes que han culminat, directa o indirectament, amb l'escriptura d'aquesta tesi. Per això, vull donar-vos les gràcies a tots i totes les que heu contribuït, personalment o científicament amb aquesta tesi.

En primer lloc, vull agrair a l'altra peça fonamental en la creació d'aquesta meva petita contribució en la ciència: el meu director de tesi, en Fèlix. Gràcies per donar-me l'oportunitat de treballar investigant com a científic i tractar-me com a tal des del primer moment. Amb l'exigència que demana anar sempre més enllà, tot i el cansament que pot comportar seguir excavant sempre una mica més. Gràcies a això he après i crescut un munt a nivell científic (encara que potser no quedi bé que sigui jo qui ho digui). M'emporto un gran record d'aquelles reunions en què anaves fent preguntes en veu alta, esborrant amb una mà bruta de la tinta dels retuladors el que havies escrit en un segon amb l'altra. Quedaran les fórmules, sí -aquesta tesi n'és la prova-, però sobretot les interpretacions i les preguntes que feies. Fer-se les preguntes corresponents i mirar com poder donar-hi resposta, un gran aprenentatge. També vull agrair de tot cor a la meva directora, Maria. Tot i arribar més tard a la meva tesi, la marca que hi deixes és clau. Els primers passos en la redacció dels articles van ser amb tu i segur que algun dia recuperaré alguna fulla plena de les teves correccions. Has sigut una peça clau en aquesta tesi, Maria. Els dos, Fèlix i Maria, heu sigut una direcció de tesi enorme: el fruit d'aquesta tesi és gràcies a la gran contribució que hi heu fet. Mil gràcies.

Tants d'anys rondant pel lab han fet que hi hagi conegut un munt de gent, a qui també haig d'agrair la seva inestimable ajuda. Marco, Sílvia, Joan, Anna, Maria i Sara: vau ajudar-me amb els primers passos (que sempre són els més difícils) i em vau descobrir d'un bon tou de "trucs" experimentals i teòrics que han sigut essencials al llarg dels 6 anys de tesi. Us recordo exultants, en algun sopar de celebració de publicació d'algun paper; però també ajudant-nos als principiants amb totes les tasques que comporta endinsar-se al món de fer experiments amb les miniTweezers. I tot, sense esbufegar amb l'estrès d'haver de treballar, i sense posar mala cara quan us demanàvem ajuda. Moltes gràcies.

Que qui eren els que anàvem demanant ajuda? Doncs érem tres: el Marc, l'Álvaro i jo. Quina sort haver pogut compartir la tesi amb vosaltres! Els cafès (i les cerveses) al bar. Les aventures als congressos. Les bromes de l'Álvaro (encara ric només d'imaginar-me'l rient) i la seva

capacitat científica brutal, que sempre et deixava pensant. L'energia i els "piques" que es portaven amb el futbol que sempre buscava el Marc. Els suggeriments experimentals, sempre encertats, del Marc. Heu sigut els millors companys de tesi que un podria haver demanat.

I el lab és gros, i cabem més persones. Mil gràcies a la Marta, per escoltar-me en els moments més baixos i per apuntar-se a una bona festa sempre que hagi fet falta (ai, plataforma). A la Isabel per sempre ser allà quan es necessita un cop de mà. Al Paolo per sempre tenir una broma a punt, amb una mica d'humor negre per a especiar el dia. I al meu relleu al despatx V337 (*rules!*), el Jaime. Realment crec que no podria haver-hi un millor relleu al grup, no em trobareu gens a faltar.

No oblidaré mai els dies al despatx amb l'Annamaria i la Carmina. Heu format part de l'etapa que recordo amb més carinyo al lab. Entrar al despatx i sempre estàveu de broma les dues, amb el positivisme que emana l'Anna i l'energia que desborda la Carmina. Tot el que m'heu ajudat en la tesi no hi cap en unes poques línies. Moltes gràcies, *grazie mile*, sou un amor.

I els agraïments per a la gent del lab continuen. Fabio, *grazie mile* for that night you helped me programming until 4 a.m., among other countless things I can not even start to thank you for. Aurélien, thank you for your comments, which were always deep and right (and a bit for your swiss diplomacy, too, which I feel sometimes I should have implemented more). Vegard, thank you for the instauration of the beer o' clock tradition, and for even trying to combine it with performing helicase experiments. Some of the best memories I have during the thesis are with you (Alps, hiking, partying) I gràcies a la instauració del beer o'clock també he pogut gaudir de les discussions filosòfiques i l'anàlisi més profunda de la realitat que he escoltat mai (i de l'enfotre'ns del francès), merci, Marius i *multa sanatate! Noroc!.*

A la resta de la gent que va passar pel lab, gràcies! A la Regina, per la "marxa" que portaves a sobre (i per acceptar acompanyar-me a un concert de Zoo al primer dia de conèixer-nos), Zubair, pel teu sentit de l'humor (i pels vins de Porto) i Patrick, per la tenacitat en aprendre català amb un somriure que mai esborres.

El darrer any i mig de la tesi em va portar unes de les millors estones que he passat a la universitat: la pausa del dinar. La xerrera escalfant als micros, arreglar el món mentre fardàvem dels plats que ens havíem cuinat. Infinitament agraït estic de totes aquestes estones. Amb el Carlos, per portar-me a fer una mitja marató, una cosa que sempre havia volgut fer i mai m'hi havia atrevit. Amb la Mariona i el seu bon "rotllo" que tot ho omple. Amb l'Elis i el Sant Albert brutal que ens vam marcar. I per fer-me de taxista per anar a collir bolets a Vall-llobrèga (així ho pronuncio bé, Mariona?).

En aquest grup divers també hi havia un col·lega del màster, el Joan. Amb ell i el Ricard havíem fet dinars de tupper plens de discussions i temes apassionants (bé, i bromes de tot el que passava a la uni). I del màster també h havia el Lluís, sempre de broma menys quan s'havia de posar seriós (i se n'hi posava, i molt!). Amb ells vaig fer els primers passos per la facultat. Mil gràcies per fer sentir-me com a casa! També,

gràcies a ells vaig conèixer els companys de despatx del Joan, l'Eloy i el Paco (per sort mai sabrem qui va guanyar el concurs de qui era el més sonat del despatx). De tu Joan, però, m'agradaria dir que és amb tu que m'ha passat una de les millors coses que m'ha portat el coronavirus: que t'hagi fet quedar per aquí i poder seguir xerrant d'absolutament TOT el que m'interessa. I fer excursions, i reflexionar d'història, llengua, costums, ciència. Ets el *pack* sencer, Joan.

No hauria pogut acabar la tesi d'una peça (bé, que li ho diguin al meu lligament creuat) si no hagués sigut pels sonats del bàsquet: Kim, Xavi, Dani, Pedro, Víctor. El *flying free* cantat pujant per les escales a la pista, fent un tipi amb els pantalons per les aixelles com si fossin els 70s, els sopars d'equip que acabaven pitjor (o millor) que una comèdia surrealista... Aquesta energia i les ganes de passar-s'ho bé han sigut el combustible anímic que he necessitat per a tirar endavant en els moments més foscos. I del bàsquet també vénen el Roger i l'Alberto. Mil gràcies per totes, totes les vivències que ja fa anys que ens hem donat i que seguim donant-nos. Anècdotes i més anècdotes, per a omplir llibres i llibres com aquesta tesi. Els assajos als bucs de la fontana, deixant-nos la veu i la dignitat i destruint els instruments han sigut tant clau com els sopars-festetes fracassades del bàsquet en què acabem els quatre de sempre en aquella foto sopant a ca meva i sense samarreta.

Mil gràcies a la colla de Mollet (o hauria de dir del Vallès-Barcelona?): pels dijous de pintxos al pool, amb la Gemma B. i l'Alba; per les trobades que ja són gairebé històriques amb tota la colla, Gemma T. i Úrsula (us enrecordeu del primer diumenge de març del 2025). Al Carlos per compartir totes aquestes excursions, i rutes pels Pirineus, que, tot i la pluja, el vent, la gana, la son o el fred, sempre han sigut un descans i un canvi de piles al·lucinant. I a l'Anna i el Jofre... ho sento, Anna, però és que he pensat amb ell i ja m'estic rient. Et dedicaria algun insult, Jofre, però no quedaria bé a la tesi. I em respondries, digués el que et digués: "I tu més". Mil gràcies per tot, als dos, pels can Muntaner quan més falta han fet, per picar-nos una mica i reviure junts els infinits records de Menorca...

A l'Aitor i la Meri, per a acollir-nos al seu petit racó de Suïssa, i donar-nos les vacances d'agost més curtes de la vida, però també amb una densitat de felicitat més altes. Ah, i també perquè compartir en paral·lel una mica els patiment de la recta final de la tesi fa que sembli menys dur, en conjunt.

I parlant de colles, mil gràcies a la colla de Barcelona (i Mataró), per les paelles a la platja i les calçotades! Lucía, Núria, Mireia, Isaac, Ari, Damià, per ser-hi sempre allà i animar-me sempre que ha calgut. I celebrar-ho amb totes les forces del món quan ha fet falta. I enllaçant amb la colla, tenim el mal (que no ve d'Almansa), si no que ve d'Alacant. Per l'energia inacabable del Kike (i per sempre sentir-te escoltat al seu costat), l'humor inesgotable de l'Ama (i les seves reflexions sobre el formigó) i el durum del Fubu. Gràcies per uns vinyes inoblidables!

Inoblidables són les estones amb els amics que sempre hi han sigut i sempre hi són. Gerard, moltes gràcies per tot, per escoltar-me en els mals moments, per sempre ser-hi. Per 30 anys més! I mil gràcies a tu, Tiko.

Cada cop que parlo amb tu acabo amb un extra d'energia i motivació per a fer ciència increïble.

Diuen que els amics es trien i la família no. I això és prou cert, però en el cas de l'Isidro, la Núria i la Berta (i el Cosme! que no me n'oblidi d'ell!), els triaria encara que no vinguessin amb el "pack" de la Irene. Moltíssimes gràcies pel vostre suport, en tot moment. Pels sopars i dinars, dies a la platja i per acollir-me com un més. Ah, i gràcies també per deixar-me acabar la tesi, no matant-me durant l'excursió al Capolatell.

Mama, Marc, Núria, Papa. Per ordre alfabètic, perquè em seria impossible ordenar-vos de cap altra manera. Sou la gran base sobre la que he pogut construir la tesi. La samfaina de la mama. Les excursions amb el Marc (una mica més i encara ens tenen en cerca per la Pica). El sortir a córrer amb la Núria. Els dissabtes dels croissants de xocolata del papa. Infinites gràcies per tot. Per intentar entendre el que mal-explico sobre el que faig a la uni. La tesi és vostra, família. I de la família extensa, Gemma, també. Moltes gràcies per escoltar i aguantar el que hagi fet falta quan només era part del "pack" familiar.

I, *last but not least* (ni molt menys), la Irene. Crec que no puc arribar a emfatitzar com d'important has sigut durant tots els anys d'incubació d'aquesta tesi. La paciència que has mostrat en els moments difícils, com has fet que em senti recolzat fins i tot en aquests darrers mesos d'escriptura. Com hem tirat endavant la tesi, a pesar de corones, contractes que ballen, pisos que no es troben. Per tot, i per tot el que encara ha de venir: gràcies infinites (no numerables).

Escrivint aquest agraïments m'he adonat de la dificultat que em suposa l'agrupar a la gent que m'ha ajudat en una única categoria, totes se'm difuminen. En tot cas, gràcies a tothom per haver aguantat qualsevol de les males cares i m'heu apropat una cervesa, o m'heu fet una broma per a alegrar-me el mal moment. Ho heu aconseguit, sempre. I és a vosaltres a qui dec gran part d'haver fet, viscut i sobreviscut aquesta tesi.

Barcelona, 30 de setembre de 2020

CONTENTS

ACRONYMS

Part I

INTRODUCTION

SINGLE MOLECULE BIOPHYSICS

In *The Structure of Scientific Revolutions* [1], Thomas S. Kuhn presents a picture of the development of science quite unlike any that had gone before. Before Kuhn, "there was little by way of a carefully considered, theoretically explained account of scientific change" [2]. The development of science was thought to be a by-product of the prevailing philosophy of science, as well as a popular, heroic view of scientific progress. The evolution of knowledge in science was something linear, based on the addition of new truths to the stock of old truths. Such progress could accelerate in the hands of a particularly great scientist, but progress itself was assumed to be guaranteed by the scientific method. However, in Kuhn's theory, science evolves as a consecuence of the jumps between different fundamental concepts and experimental practices of a scientific discipline, coining the term paradigm shift[1]. A typical example of a paradigm shift would be the rise of quantum mechanics during the 1920s, which would shape the way physics was understood, thought and taught, compared to the previous paradigm, classical mechanics.

With the definition given above, the interdisciplinary work from two existing fields does not necessarily imply a paradigm shift (i.e. the fundamental concepts and experimental practices remain the same for both disciplines). However, the creation of a new field of study, such as is the case for biophysics, can be understood as a paradigm shift in biology (i.e. new concepts and experimental practices were introduced).

The birth of biophysics was a slow process, which started in the mid 19th century, when scientists began to wonder whether the blooming expanding knowledge in physics could be used to answer biological questions. First it started with the so-called medical physics, with the contributions of figures such as Carl Ludwig (1816-1895), who first measured blood pressure, or Emil du Bois Reymond (1818-1897), who first studied the nerve action potential.In 1944, Erwin Schrödinger, one of the main contributors to the development of quantum physics theory, wrote the book "What is Life?" [3], suggesting a physical approach to biological questions.

But the starting point of modern biophysics was settled by the discovery of the DNA double-helix structure, by using X-ray diffraction, an experimental technique developed by physicists. The early stages of biophysics were therefore mainly focused on the structure and interaction of biomolecules and cells. The advent of new theoretical tools, such as graph theory, as well as the development of new experimental techniques, such as single-molecule techniques, during the last decade of

---

1 Even though Kuhn restricted the use of the term to the natural sciences, the concept of a paradigm shift could also be used to define his impact in the philosophy of science

the 20th century, has lead to new biophysical discoveries in all scalles, from nanometric-sized molecules to global ecological networks (Fig. 1.1).

1.1 BIOPHYSICS

Biophysics is a scientific discipline that uses the tools and the methods of physics to study biological systems [4]. Due to the broad and diverse nature of living systems, biophysicists focus on a wide range of processes from the molecular to the ecological scale, approaching to their structures, dynamics or complexity.

Physics and Biology are the two branches of knowledge that blend into this discipline. However, the historical development and tools applied in physics and biology are very different: biology tends to be descriptive and complex, while physics is conceptual and symplistic [5].
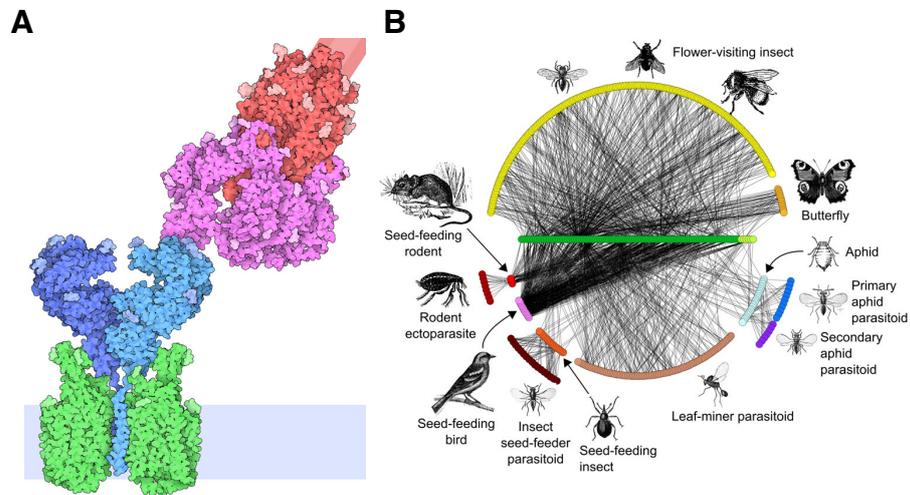
Figure 1.1: **Some topics in biophysics.**. (a) Molecular biophysics: Illustration of a complex of the spike protein of SARS-CoV2 (red and magenta) bound to its human receptor ACE2 (blue). The cell membrane is shown schematically in light blue at the bottom. From the combined images of [6] and [7] of the PDB "Molecule of the Month" of June 2020 [8]. (b) Systems biology: species interaction networks at Norwood Farm, Somerset, UK. From [9].

A physical approach to biophysics is the study a biological system like any other physical system in order to try to find and understand new underlying physical laws. This can be translated as a gain in physics knowledge. An example of this approach is the verification of the force response predicted by the worm-like chain model with the force vs extension curves obtained from the measurement of a single-DNA molecule, first achieved in 1992 using magnetic tweezers [10, 11].

A clear example of a biological approach to biophysics is the discovery of the double-helix of DNA [12], which supposed a great advance in biology: it unraveled DNA as the carrier of genetic information in living organisms. Since X-ray diffraction was developed at the beginning of the 20th century, the discovery did not contribute directly to physics at the methodological level.

Although the border between biophysics and other closely related disciplines such as biochemistry, nanoscience or biomedicine is not clear, biophysics is no longer "physics for biologists" or "physical methods applied to biology", but a discipline defined by its own scientific questions [5].

1.2 SINGLE-MOLECULE EXPERIMENTS AND THE PHYSICS OF SMALL SYSTEMS

Most of the discoveries in molecular chemistry and biology over the 20th century have been obtained from experiments performed in bulk. A bulk experiment involves following a process on a large collection of molecules. These experiments show the average behaviour of the molecules, allowing to extract useful information from the population averages. However, fluctuations in this nano-scale are important (see below) and therefore averages alone do not describe these molecular processes completely. The study of biomolecules therefore benefits from single-molecule techniques, which allow to track the behaviour of individual molecules, complementing the information obtained from bulk assays.

An example where such single-molecule features hidden to macroscopic observations are particularly relevant is regarding the unwinding kinetics of a DNA helicase. As shown in Fig. 1.2 DNA unwinding process in a bulk assay might look like a smooth and continuously varying reaction due to the unsynchronized average over a large number of unwinding enzymes. These experiments provide the average unwinding of the enzyme, but the underlying molecular dynamics of a DNA helicase is far from the smooth, continuous observed phenomenon in bulk. Single-molecule techniques has allowed to follow these enzymes individually giving access to their unwinding kinetics, which consist on a series of quick unwinding of DNA, alternated by pauses and backtracking.

Futhermore, the development of force spectroscopy single-molecule techniques allow to measure physical magnitudes difficult to determine with previous techniques (such as force and torque), which in turn helped to establish a physical approach to the study of biomolecular systems, as discussed above.

A fundamental feature of single-molecule experiments is the observation of fluctuations in the measured quantities. These fluctuations do not play a role in typical bulk experiments containing a large number of molecules or subunits N, since energy fluctuations scale as $1/\sqrt{N}$. In contrast, in single-molecule experiments, $N \sim 1$ and, hence, energy fluctuations are relevant, $1/\sqrt{N} \sim 1$. Moreover, in small systems, such as the ones studied in single-molecule (e.g. nucleic acids, proteins or molecular machines), typical energetic exchanges between the system and the environment are on the order of Brownian fluctuations (i.e. $\sim k_{\mathrm{B}}T$), so they are inherently noisy. Small systems thermodynamics provides the precise framework that allows us to understand the physics of small systems [13]. The study of small systems and the small N regime
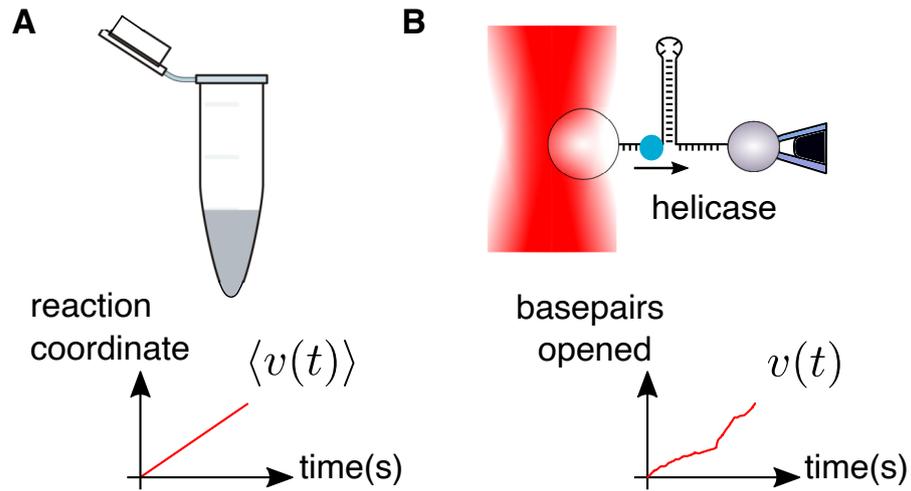
Figure 1.2: **DNA unwinding by a helicase as observed in bulk and single-molecule.**. (a) Bulk measurements allow to obtain average properties, such as the unwinding rate, $\langle v(t) \rangle$. (b) Single molecule experiments allow to obtain precise measurements of the unwinding rate, $v(t)$. Furthermore, different sequence-regions and regimes can be followed in real time. Note that, assuming the same conditions are fullfilled in both experiments, the average value obtained for repeating $N >> 1$ the experiment depicted in (b) , the results of (a) should be recovered.

has become a trending topic due to the recent developments of micromanipulation techniques, among them force spectroscopy single-molecule experiments, which allow to manipulate individual molecules

1.2.1   *Force-spectroscopy techniques*

Among the force spectroscopy single-molecule techniques developed in the past years, three of them stand out due to its widespread use: atomic force microscopy(AFM), laser optical tweezers (LOT) and magnetic tweezers (MT). These techniques allow to manipulate single-molecules and can exert and measure applied forces to the tethered molecules. The combination of their force ranges allow to cover the whole range of forces relevant to biomolecular reactions: from small polymer entropic forces ($\sim 10^{-3}$pN) to the typical force required to break the covalent bonds of the backbone of proteins and nucleic acids ($\sim 10^{3}$pN).

Table 1.1: **Single molecule techniques.** Comparison of the typical characteristics of force spectroscopy techniques. Values from [14].

	AFM	LOT	MT
Force range(pN)	$10^1 - 10^4$	$10^{-1} - 10^3$	$10^{-2} - 10^1$
Spatial resolution(nm)	0.1	1	1
Stiffness(pN/nm)	$10^1 - 10^5$	$10^{-2} - 10^0$	10^{-6}
Temporal resolution(s)	10^{-3}	10^{-4}	$10^{-1} - 10^{-3}$
Parallelization	no	no	yes

Here we briefly introduce these techniques, as well as how their strengths and limitations make them suitable for certain applications. The main characteristics of each experimental technique is summarized in Table 1.1.

1.2.1.1 *Atomic-force microscopy (AFM)*

AFM is a widespread technique used in imaging, which can also be used as a force spectroscopy tool. The AFM is based on the scanning tunneling microscope and consists of a cantilever of $\sim 100\mu m$ with a sharp tip at its end (Fig. 1.3A). The molecules under study are adsorbed on a planar surface that can move relative to the cantilever. The tip is coated with molecules that recognize and bind a site of the molecules adsorbed on the surface. By bringing the tip close to the surface, a single-molecule can be tethered. By moving vertically the surface a force is exerted to the molecule. The cantilever bends due to the exerted force and its bending can be used to measure the applied force by reflecting a laser beam focused on the edge of the cantilever. The force is measured from the signal of the deflected light in a position sensing detector (PSD). AFM can be used to exert forces up to $\sim 10nN$, ideal for studying very stable molecules, such as certain proteins (e.g. titins). The force and spatial resolutions can be estimated from the Equipartition law, giving $\sqrt{\langle \Delta x^2 \rangle} = \sqrt{k_B T/k} \sim 1\text{Å}$, regarding distances, and $\sqrt{\langle \Delta f^2 \rangle} = \sqrt{k_B T k} \sim 10pN$, regarding forces. The high values for the stiffness of the cantilever (up to $k \sim 100pN/nm$) leads to a large spatial resolution of $\sim 1\text{Å}$, but a small force resolution of $\sim 10pN$

One limitation regarding AFM experiments comes from the necessity of the cantilever to be really close to the surface at exerting low forces ($< 10pN$), when many undesired interactions can happen. On the other hand, cantilevers are very expensive and, due to their fragility, have to be constantly replaced, which turns AFM one of the most expensive techniques.

1.2.1.2 *Magnetic tweezers (MT)*

Magnetic tweezers are based on the force exerted upon a magnetic particle, with magnetization $\vec{m}(\vec{B})$, when placed in an inhomogeneous magnetic field. The resulting net force [15] is in the direction of the gradient, $\vec{f} = 1/2\vec{\nabla}\left(\vec{m}(\vec{B}) \cdot \vec{B}\right)$. Therefore, by using permanent magnets or electromagnets it is possible to exert forces on micrometer-sized magnetic particles and perform single-molecule experiments by attaching molecules to the magnetic particle.

As shown in Fig. 1.3B, the setup consists of a pair of permanent rare earth magnets placed above a microfluidics chamber which is mounted in an inverted microscope. Molecules are tethered between the surface of super-paramagnetic beads and the coverslip of the microfluidics chamber. The force applied is controlled by varying the distance of the chamber with respect to the magnets, while the beads are imaged using a CCD camera.
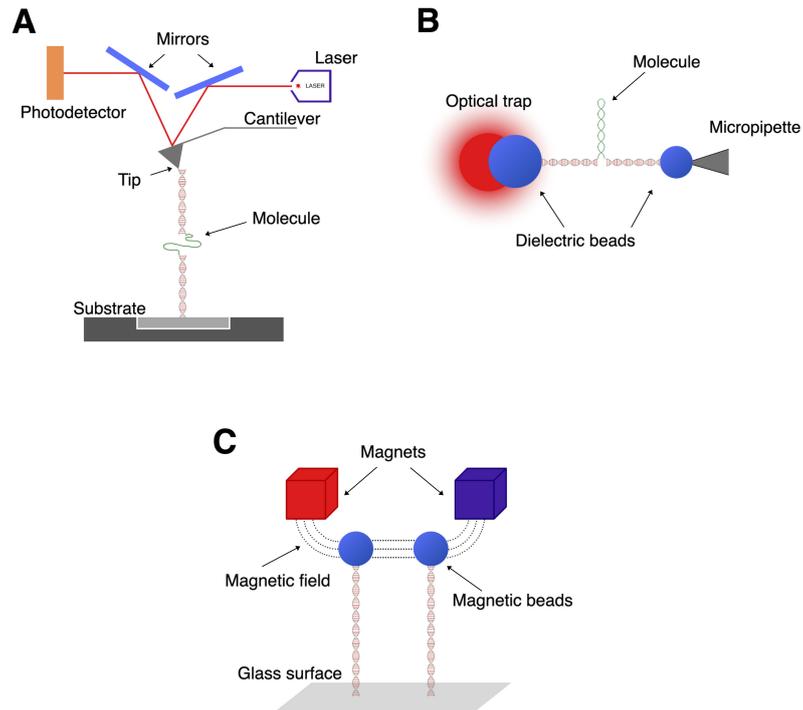
Figure 1.3: **Scheme of single-molecule devices**. (a) Atomic Force Microscope. (b) Laser Optical Tweezers (c) Magnetic Tweezers.

The stiffness of MT is very small, of the order of 10^{-6}pN/nm, which makes magnetic tweezers very well suited to perform constant force experiments as they work as a passive force clamp. On the other hand, the position of the beads (and hence the molecular extension) is tracked by the CCD camera images, which limits their temporal resolution. Typical range of forces achieved in 1μm sized beads is between $0.01 - 30$pN, although higher forces (~ 100pN) can be achieved by using large magnetic beads ($\sim 3\mu$m). Typical resolutions for Magnetic tweezers are ~ 1nm, regarding distance and $\sim 10^{-3}$pN, regarding forces.

A key feature of MT is the possibility to exert torques on molecules, by simply rotating the magnets creating the magnetic field. This fact allows to study the coupling between elastic and torsional properties of biomolecules, which plays an essential role in DNA and its interaction with molecular motors. Another advantage of this technique is that MT allow for high-throughput measurements: up to 100 single-molecule experiments can be performed in parallel.

1.2.1.3 *Laser optical tweezers (LOT)*

The physical principle behind LOT is that light carries momentum which can be used to exert forces on microscopic dielectric particles. An optical trap is usually created by focusing a laser beam with a high Numerical Aperture (NA) objective into a microfluidics chamber. On the one hand, a dielectric particle is captured in the focus of the optical trap. On the other hand, another microparticle is held by air suction at the tip of a micropipette[2]. Both beads are biochemically labelled so that the molecule of study forms a tether between the two beads, as depicted in Fig. 1.3. The usual range of forces in LOT is $10^{-1} - 10^2$pN, depending on the bead size and the laser power (which determine the amount of deflected light).

Since typical stiffness are on the order of $k \sim 10^{-2}$pN/nm, their spatial resolution is of the order of ~ 1nm. One of the advantages of LOT is the capability of some setups to measure force directly, with sub-millisecond and 10^{-1}pN resolution. These facts make LOT one of the most versatile single-molecule techniques: protein-DNA interactions, molecular motors and DNA elasticity have been widely investigated using LOT. The main drawbacks of LOT are the high complexity to setup and align, to parallelize measurements or to exert rotational control on the trapped bead.

Chapter 2 deals with the detailed description of this technique.

1.3 SUMMARY OF THE THESIS

In this thesis, single-molecule experiments using LOT are employed to extract accurate information about the thermodynamics and kinetics of various molecular systems, with special emphasis on the elastic properties of single-stranded DNA (ssDNA). The thesis is divided in three parts.

Part I provides a general description of the research field as well as the main theoretical framework for the basic concepts that will be developed in parts II and III. In Chapter 2 the miniTweezers and the experimental setup used throughout the thesis is described, as well as the physical basis of its working mechanisms, introducing the phenomenon of optical trapping. Chapter 3 contains a brief introduction of the biomolecules of study in this thesis, with an explanation of their historical discoveries, as well as their structure and function. The main focus of this chapter is on ssDNA, which is the main object of study of the thesis. Chapter 4 introduces the polymer models that are widely used in describing the elasticity of nucleic acids and proteins. Specifically, the Freely-Jointed Chain and Worm-Like Chain models are presented.

Part II deals with the elasticity of single-stranded DNA. This is the main part of the thesis, and it includes chapters 5-7. Chapter 5 is about the study of the elasticity of ideal ssDNA chain, i.e. the one that can be modelled as ideal polymers (presented in Chapter 4). The study of the elasticity of different DNA sequences is presented. The

---

2 There is also the possibility of using a dual trap setup, which is not consider in this explanation. Further information of dual traps can be found in [16].

blocking-splint oligo technique is described, a experimental technique developed for studying the elasticity of short (tens of bases) DNA molecules. This study shows the need of using extensible models to succesfully describe ssDNA elasticity over a large range of forces, which explains the previous discrepancies on the elastic parameters obtained in different studies. We also provide an explanation for the required extensibility of the model: a transition experienced at the nucleotide level: a change in DNA sugar pucker conformation. A simple two-states model is introduced and preeliminary results regarding its energetics are presented. The characterization of the ssDNA elasticity is central for the works developed in the following chapters.

Chapter 6 studies the stacking-unstacking transition for ssDNA, previously observed for certain sequences (mainly purine-rich ones). Several molecules, with different degrees of stacking, are studied by obtaining their force extension curves (FECs). A cooperative helix-coil model including heterogeneity is developed and used to fit the obtained FECs, allowing to obtain elastic parameters to describe the stacked chain. The salt dependence of the unstacking transition is also measured by studying two of the sequences by varying the salt concentration over two decades. The free energy of formation of dsDNA duplexes depends on the salt concentration. The obtained salt dependence on the stacking free-energy of ssDNA provides a possible explanation for the salt dependence of duplex formation.

Chapter 7 deals with the non-specific structures that arise at low forces and high salt concentration when pulling ssDNA molecules longer than ~ 100 bases. A helix-coil model with cooperativity is proposed and used to extract some mean-field characteristics of these structures. 8 different sequences are studied, characterizing their elasticity and deviation from the ideal elastic behaviour. The results for a 14kb molecule for 3 decades of varying NaCl and $MgCl_2$ are also shown. All experimental FECs are fitted to the helix-coil model. The model can be used to predict the formation of secondary structures at zero force. A comparison between the predicted structures from the model and those obtained from Mfold is also investigated.

Part III contains two studies which also need of the correct determination of ssDNA elasticity. In Chapter 8, we study the interaction between the RecQ helicase from E. coli and DNA, i.e. how the RecQ unwinds double-stranded DNA molecules, releasing single-stranded DNA. We obtain some of its kinematic properties as well as study the entropy production of the system using the Fluctuation Theorem. In Chapter 9, the effect of DNA mismatches, i.e. non complementary base pairing, on the stability of DNA is studied. To do so, two types of experiments on several DNA sequences are performed: stretching and releasing the molecule by moving the optical trap (pulling experiments) and monitoring the folding/unfolding of the molecule passively (hopping experiments).

OPTICAL TWEEZERS

The first attempts to explain the origin of light had a mythological origin[1]: from Ra, the egyptian Sun-god to the bible, light was created by a superior figure. The Greek philosophers were the first to lay down the study of light, focusing on the problem of explaining vision. Euclid (330-275 B.C.) collected all the existing knowledge about it in his book *Optics*, including the rectilinear propagation of light rays and the law of reflection. However, ancient Greeks never succeeded in explaining the mechanisms behind vision. Until the Arab scientist Alhazen (965-1040), the general knowledge was that vision would require of emanating rays to an object in order to make it visible. Alhazen, not only proved it wrong, but he also performed a large set of experiments and observations regarding refraction and reflection, using lens and mirrors, earning the title of 'father of modern optics".

The improvement of glass-working techniques during the whole medieval era, specially in modern-day Italy, would culminate with the creation of lenses, a fundamental requisit for the invention of telescopes and microscopes. Telescopes and microscopes would discover a new range of worlds, on opposite sides of the length scale. They would also play a key role in the development of optics, e.g. telescopes would be essential in allowing for the earliest measurements of the speed of light, performed by Ole Rømer (1644–1710), measuring the time spent in Jupiter's shadow by its satellite Io.

With the advent of natural philosophy and physics, scientists would seek a definite answer about the nature of light. René Descartes (1596-1690) and Isaac Newton (1642-1727) would give a physical interpretation for light: white light is made of colors, which are in turn given by different types of particles, each type with its own properties. But the so-called corpuscular theory of light would encounter strong opponents. Christiaan Huygens (1629-1695), Thomas Young (1773-1829) and Augustin Fresnel (1788-1827) would settle and boost the opposing interpretation: the wave theory of light. Experiments regarding phenomena that could not be explained by using the corpuscular theory (which would recall any observation requiring the explanation of interferences) would finally settle the debate. The publication of James Clerk Maxwell's (1831-1879) equations seemed to be the final nail in the coffin of the corpuscular theory of light.

However, in 1905[2], only 43 years after the publication of Maxwell's laws, Albert Einstein (1879-1955) proposed a solution to the problem of black body radiation (the so-called ultraviolet catastrophe) that Max Planck (1858-1947) had solved with his theory of quanta.Einstein

---

1 A more detailed description of the origin of Optics and its history is found in [17]
2 With the publication of his paper about the photoelectric effect, A. Einstein would be awarded in 1921 a Nobel Prize, not for his theory of Relativity.

proposed that light was composed of particles, called *quanta*[3], each one carrying an energy, E, proportional to the light frequency, ν,

$$E = h\nu, \tag{2.1}$$

with h being the Planck constant. With the advent of special and general relativity, the energy of each particle of a given mass m would be written as:

$$E^2 = m^2c^4 + p^2c^4, \tag{2.2}$$

where c is the speed of light. For the case of photons, however, they do not have any mass, hence, combining this equation and the so-called Planck-Einstein equation (Eq. 2.1), we obtain the linear momentum of a photon (De Broglie relation):

$$p = \frac{h\,\nu}{c}. \tag{2.3}$$

The advent of quantum mechanics, which was due to the contribution of scientists such as those appearing in Fig. 2.3, would be needed to finally settle the wave-particle interpretation, allowing for the two apparently contradictory interpretations to coexist. This new field of science would lead to the creation of the laser (acronym for Light Amplification by Stimulated Emission of Radiation) in 1960, emerging as a perfect tool with broad applications, among them, optically trapping small objects that would lead to Optical Tweezers.

In this chapter we introduce the principles of optical trapping and the experimental setup used in performing the experiments throughout the thesis: the miniTweezers.

2.1   PRINCIPLES OF OPTICAL TRAPPING

The ability of light to induce forces dates back, at least, since 1619, when Johaness Kepler's (1571-1639) book *De Cometis* described the deflection of comet tails by the Sun's rays. Maxwell's laws also predict that light, as an electromagnetic wave, carries momentum. It is the existence of this linear momentum of light that allows it to be used to trap objects, by using the interaction between the light and the particle to transfer some of the momentum to it.

In 1970, Arthur Ashkin (1922, Fig. 2.2B), who is the oldest person to have received the Nobel Prize in Physics (2018, shared with Gérard Mourou and Donna Strickland), firstly observed experimentally the interaction between a laser beam and a particle. He then made a first experimental attempt to trap particles using lasers, based on compensating the downwards acting force of gravity with the upward force induced by the scattering of a light beam moving upwards. However, the equilibrium reached was very unstable and their team was forced to change their approach. It would not be until 1986 when Ashkin and collaborators[18] would observe optical trapping of particles using a single laser beam,

---

3 The current name of "photon" is generally attributed to Gilbert N. Lewis (1875-1946).

demonstrating that the gradient force due to the focused laser beam was higher than that exerted by the scattering of the light of the same laser beam. This would set the beginning of optical trapping.

.
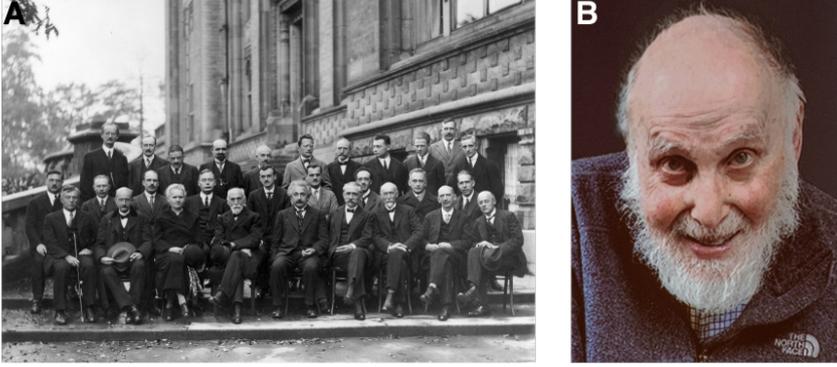
Figure 2.1: **Behind the invention of optical tweezers** (a) Picture of the assistants of the 5th Solvay conference (1927). 17 out of the 29 participants were Nobel awardees (Marie Curie, the only woman in the picture, was awarded twice). Max Planck and Albert Einstein, in the center, laid down the basis of quantum theory. (b) Arthur Ashkin received the Nobel Prize in Physics in 2018 "for the optical tweezers and their application to biological systems."[19]. Ashkin passed away in September 21st, 2020, during the writing of this thesis.

2.1.1   *Ray-optics approach*

If the dimensions of the trapped particle are much larger than the wavelength of the beam, the ray optics approach provides a valid approximation as well as a qualitative interpretation of the creation of the optical trap. In the ray optics description a light beam is decomposed into individual rays which are characterized by their intensity, with their directions described using the laws of geometrical optics. Hence, each individual ray can be traced throughout the particle, allowing to determine the starting and final linear momentum of the light beam. Each ray contributes with N photons to the total momentum:

$$\vec{p} = N \frac{h\,\nu}{c} \hat{u}_p \qquad (2.4)$$

where $\hat{u}_p$ is the unit vector in the direction of propagation of each light beam.

In Fig. 2.2A it is shown a schematic depiction of the trajectories followed by 2 rays of a Gaussian beam, neglecting scattering (reflection). Both rays hit the particle's surface with the same incident angle and at symmetrical positions, hence following symmetrical optical paths. Since the refractive indexes of the particle, n_p , and the surrounding medium, n_m are different, and $n_p < n_m$, the incident rays deviate after entering and exiting the particle, following the Snell's law $n_m \theta_m = n_p \theta_p$. This is translated as a change between the incident ($\vec{p}_i$) and transmitted ($\vec{p}_t$) linear momentum of each ray. Due to linear momentum conservation,

the net variation experienced by both rays has to be transmitted to the particle or bead, $\vec{p}_b$, pushing the particle in that direction. This momentum does not only have a component along the direction of propagation of the beam, but also a component towards the light region highest intensity. The exerted force of the particle will be related to the linear momentum experienced by the particle following Newton's 2nd law:

$$\vec{f}_b = \frac{d\vec{p}}{dt}, \tag{2.5}$$

where it becomes explicit that the direction of the applied force on the trapped particle, $\vec{f}_b$, is the same than that of the transferred momentum. The position of the particle inside this Gaussian beam will be shifted towards the centre of the beam (where the intensity is maximum). While the particle is being displaced towards the center of the Gaussian beam, the temporal variation of momentum, as written in Eq. 2.5, will decrease, reaching a minimum in the beam's center (where both rays' transferred transversal momenta will cancel out). On the other hand, while moving far away from the beam's axis, the variation of momentum will increase, always pointing towards the center. However, this scenario, as well as neglecting the effects of the scattered light, only allows for trapping in one dimension along the direction of light propagation.

In order to create an equilibrium point to trap the particle, an intensity gradient along the direction of propagation of the beam has to be created. To do so, as shown in Fig. 2.2B, a focused laser beam allows for creating a force towards the beam waist. In this scenario, it is possible with a single focused Gaussian beam to compensate for the transferred linear momentum due to the scattered or reflected light, $\vec{p}_r$. The scattered light will always exert a net force towards the direction of propagation of the beam. To effectively trap the bead in the 3-D spatial dimensions[4] the laser beam can be focused using high Numerical Aperture(NA)[5] objectives. In doing so, the equilibrium position of the trapped particle is slightly beyond the focal point of the beam.

Another way to limit the effect of scattered light, is to use two counter-propagating identical beams (then the equally-scattered light cancels out the push from both sides). The use of two counter-propagating lasers offers a lot of advantages, as will be discussed in Sec. 2.2.

It is worth noting that shining a beam light to a particle with higher refractive index than its sorrounding medium, it will always translate into a net force towards the highest intensity regions of the beam[6].

As stated at the beginning of this section, the ray optics approach is an approximation, which provides intuitive explanation of the basic mechanisms behind optical trapping. In order to formulate an exact theory about the trapping happening in most of the optical tweezers setup (including our miniTweezers setup), where the laser's wavelength is

---

4 3D particle trapping can also be achieved by increasing the difference between the refractive indexes of the medium and the particle

5 NA is a dimensionless number that characterizes the range of angles over which the system can accept or emit light

6 For $n_p < n_m$, the force will push the particle towards light-intensity minima.
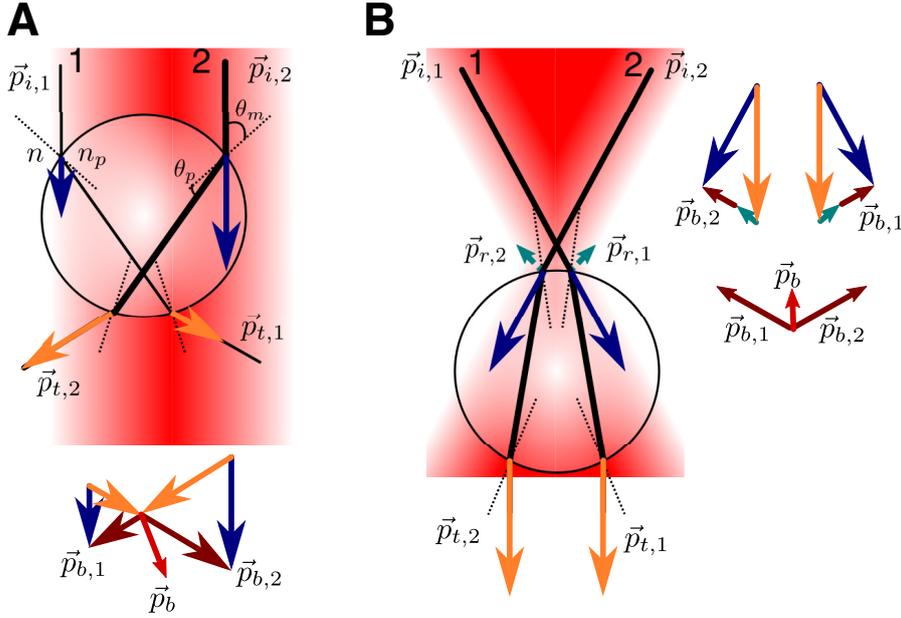
Figure 2.2: **Optical trapping principles.** (a) Ray optics diagram of a particle or bead (with higher refractive index than that of the external medium $n_p > n$) placed in a non-focused and unscattered Gaussian laser beam. Individual rays are represented as black lines, with a thickness proportional to their intensity. The arrows indicate the initial ($\vec{p}_i$, blue) and transmitted ($\vec{p}_t$, orange) light momentum of each ray. The diagram shows the change in momentum that experiences the bead, $\vec{p}_b$, due to total momentum conservation. (b) Same as in (a), but with a scattered focusing beam. Green arrows account for the momentum of the reflected light, which can be compensated by the focusing of the laser beam, as it is shown in the diagram, showing a net momentum transferred to the particle, pushing it towards the focusing point.

of the same order as the bead length ($\sim \mu$m), the generalized Lorenz-Mie Theory is needed, first calculated by Mie[20], using the wave formulation derived from Maxwell's equations. Further explanation of the subject can be found in [21] and [5].

2.2   THE MINITWEEZERS SETUP

The instrument used to performed the experiments throughout this thesis is known as the "miniTweezers", which is based on a miniaturized version of a previous design by Steve B. Smith and Carlos Bustamante in 2003 [22]. It is characterized for using two identical counter-propagating lasers (optics scheme shown in Fig. 2.3A) focused inside a microfluidics chamber that allow trapping polystyrene particles with forces up to ~ 100pN and typical trap displacements of $\sim 10\mu$m. With this configuration the force and trap position are measured with a precision of ~ 0.1pN and ~ 1nm. The data acquisition frequency is 1kHz, although it can be increased up to ~ 1MHz using the data acquisition board. The small size of the miniTweezers(Fig. 2.3B) results in a very short optical path between the emitting point of the laser and the detectors, of around ~ 40cm, which

makes them robust against beam misalignment. A detailed description of the miniTweezers can be found in [5, 23].

2.2.1   *Optics*

The miniTweezers perform a direct measurement of the force using the conservation of the total linear momentum of the system consisting of the trapped bead and the laser beams. The setup consists of two infrared ($\lambda = 845$nm) counter-propagating lasers that focus inside the experimental chamber by using two objectives with high NA. Under-filling conditions in the objectives, allows to collect all deflected light by the trapped bead by the exiting objectives permitting the direct measurement of the force. The optical path followed by each one of
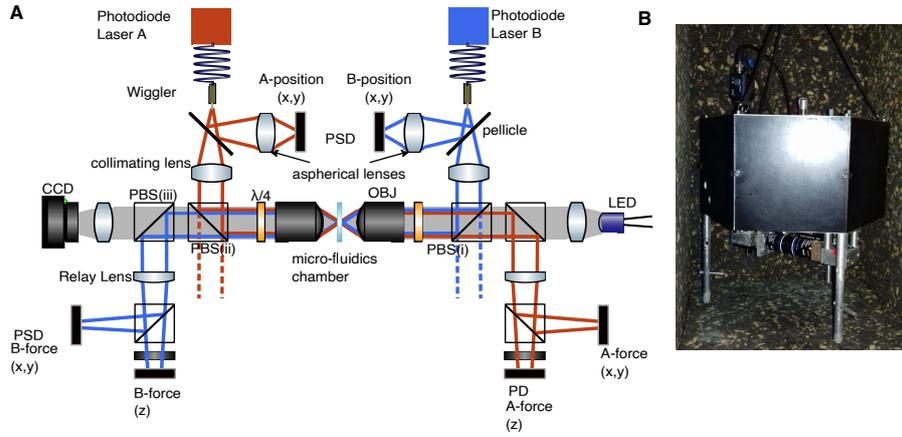
Figure 2.3: **miniTweezers setup**   (a) Schematics of optical paths of the two laser beams and LED. The position of the force-PSD, position-PSD and CCD camera are also shown. The PBS are numbered in roman numerals, following the order that the B-laser beams encounters them. (b) Picture of the miniTweezers setup. All the optics are contained inside the black box, including the lasers, with the chamber held between the objectives. The whole miniTweezers are hung from the ceiling and enclosed inside an insulating box to avoid any vibrations or air currents.

the beams, illustrated schematically in red and blue in Fig. 2.3A, is described as follows. The light emitted from the laser diode is directed through a single mode (Gaussian profile, TM_{00}) optical fiber, which has a bare end, the wiggler, from which the light is then emitted towards a pellicle. This pellicle allows most of the light to get through it ($\sim 95\%$, depending on the initial polarization of the beam), but the reflected light is then focused via a lens to a Position Sensitive Detector (PSD) that is going to register the position of the laser, the so-called position-PSD, or light-lever detector. The position-PSD registers the central position of the illuminated area, as well as the total light captured by the pixels of the detector. This signal allows to determine the laser position in the chamber (see Sec. 2.2.4.2 for details).

   The amount of light deviated towards the light lever and entering the objectives (after the reflection of the first Polarized Beam-Splitter,

PBS) can be regulated by the polarization given by the positioning of the optical fiber prior to the wiggler, which can be modified using small pieces of tape. The light that gets through this first PBS(i), allows to check for the correct colimation of the laser beam, as well as its shape[7]. The reflected light in this PBS (which is reflected with a linear polarization) gets circularly polarized by a $\lambda/4$ waveplate before entering to the focusing objective (Olympus UplanApo 60x, with NA=1.20). The circular polarization of each of the beam is chosen to not interfere with the other inside the chamber, since they will have opposite polarizations[8].

After leaving the chamber, the laser beam is collimated back by the exiting objective, and it gets linearly polarized with another $\lambda/4$ waveplate. The linear polarization of each of the laser beams is also perpendicular to each other, preventing any interference in the shared optical path, as well as to allow the transmission through the first PBS that encounters the other laser beam. The beam then traverses the PBS(ii), and encounters another PBS, which allows the LED light to get to the CCD, in the case of the B-laser[9], while reflecting the laser light towards the force-PSD detectors. A Relay Lens focuses the beam to the two force detectors, which are reached by a split beam (using a non-polarizing Beam Splitter). The x and y force-PSD detector detects the deflection of the light caused by the force exerted inside the chamber along the trap plane, while the z force-PD (Photodetector) detects the force exerted along the propagation axis, which is translated as an increase (decrease) of the size of the spot in the PD. The force detection and their calibration is described in Sec. 2.2.4.1.

As explained in Sec. 2.1, the main advantage of using counter-propagating lasers is that their scattering forces can cancel each other. The correct compensation between the forces is easily achievable by using symmetrical optical paths, such as the one used in the miniTweezers (Fig. 2.3A). The only difference between the lasers is their perpendicular polarization through all the shared optical path, between the first and the third PBS.

2.2.2 *Experimental configuration*

Single-molecule experiments using optical tweezers, are based on controlling the force/position of micron-sized particles (beads) that are coated with the molecule of interest inside a microfludics chamber. The standard experimental configuration for performing experiments is shown in Fig. 2.4. It consists of a molecule tethered between two beads in a dumbbell configuration: one bead is held at the tip of a micropipette by air suction, while the other is held by the optical trap. In order to achieve this configuration, the molecule of study is labelled with a Biotin at one

---

7 The bare end of the optical fiber in the wiggler is extremely sensitive to any type of dust or small breaks. These will alter the shape of the beam forming the optical trap.

8 The correct polarization of the beams is critical, since it will not only avoid any interference in the shared optical path, but it will also prevent any cross-talk between both lasers.

9 In the case of the A laser, the PBS allows the entrance of the LED light as the source of illumination of the chamber in a Köhler configuration.

end, while the other end is labelled with a Digoxigenin. Accordingly, one of the beads is coated with Streptavidin, while the other is coated with anti-Digoxigenin. One of the two beads in the dumbbell can be then captured by the optical trap and brought to the micropipette, where is immobilized by air suction. The other bead in the dumbbell is usually incubated with the molecule of study, allowing it to attach to its surface, using the specific antigen-antibody connections. All the synthesized molecules contain a spacer or linker between the molecule under study and the beads. In Fig. 2.4, we show a DNA hairpin, which is connected to the beads with molecular handles. Notice that the measured trap position does not correspond to the molecular extension, since the bead experiences a displacement relative to the center of the trap, Δx_b, which needs to be accounted for (Sec. 2.2.5). A detailed guide about how to use the miniTweezers setup in order to perform pulling DNA experiments can be found in App. A.
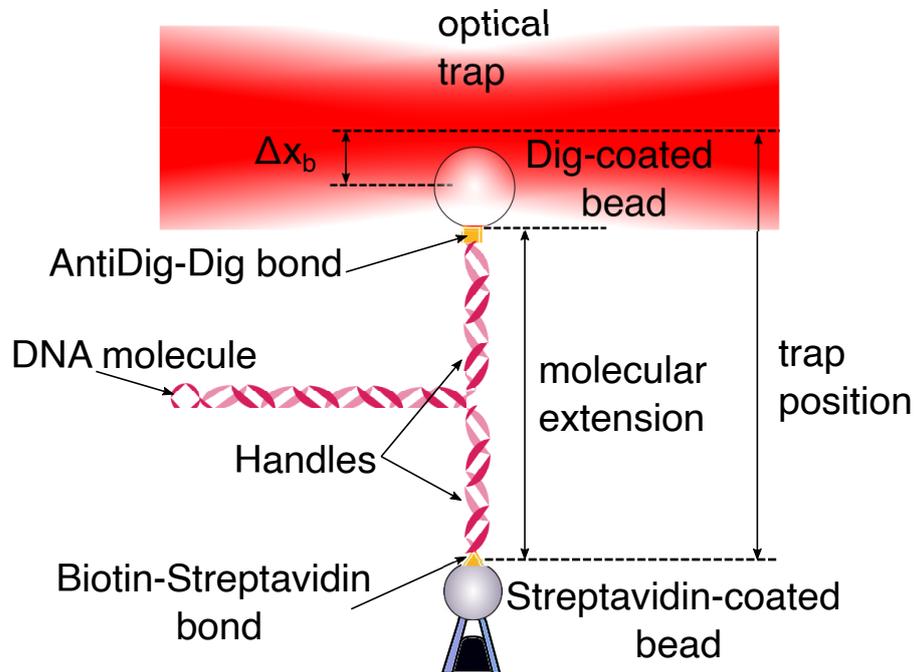
Figure 2.4: **Single-molecule experiments with the miniTweezers:** Experimental setup representing a DNA molecule tethered between the beads.

.

2.2.2.1 *Microfludics chamber*

The experiments are carried out in a microfludics chamber that is placed between the objectives, as shown in Fig. 2.3. A working microfluidics chamber, with all the entrance, exit tubes, as well as dispenser tubes and micropipette, is shown in Fig. 2.5A. The chamber consists on three parallel channels, as shown in Fig.2.5A-B, which are cut on parafilm using a laser cutter. Three entry and three exit holes are cut in a glass coverslip (24x60x0.1mm) to allow for buffer flowing. The cut parafilm is then sandwiched between two glass coverslips (one with the entry/exit

holes cut), with the dispenser tubes and the micropipette between the parafilms. The chamber is then sealed by heating it for several seconds at $\sim 120°$C. The tip of the glass micropipette has a diameter of $\sim 1\mu$m, which can fix a bead at its tip by air suction inside the chamber.
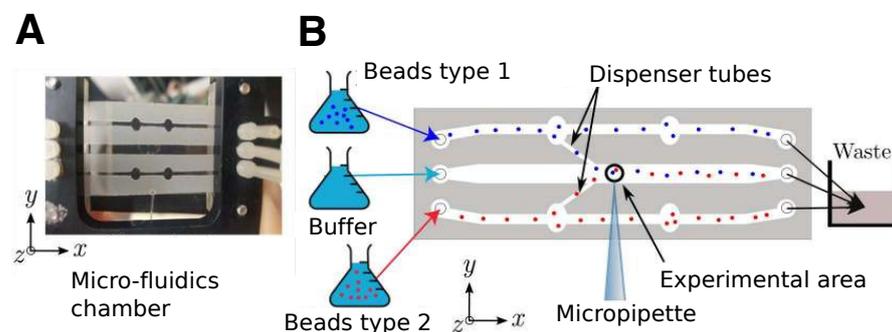
Figure 2.5: **Microfluidics scheme and bead flowing** (a) Picture of the microfluidics chamber within its frame, ready to be used with a single-molecule tethered between the bead held by the micropipette and the optically-trapped one. (b) Schematic depiction of the microfluidics chamber, with the three channels depicted. The two lateral channels are used to flow (from left to right) two different bead types and conected to the central one, where the experiments are performed, close to the tip of the micropipette. Figure adapted from [13].

The chamber is then placed to a metallic mount (Fig.2.5A), where plastic screw and tubes are used to flow in the required buffer for the experiments. Each entry tube is connected to a 1ml syringe, which, after applying some pressure, allows for controlled flow inside the chamber. The three channels have different ends: the central one is where the micropipette is placed, and hence the experiments are performed. The top and bottom channels are used to flow a different type of beads, which then can be flowed to the central channel using a motor stage and captured by moving the fluidics chamber towards each dispenser tube.

The usual approach for performing experiments is as follows. After mounting the chamber in its frame and placing it between the objectives, distilled water is placed between the objectives and the chamber. Initially, the working buffer is floed through all channels and the buffer containing each bead type is flowed on a different channel (top or bottom). For the first bead type, the chamber is moved until the optical trap is close to the dispenser tube, hence allowing for trapping a single bead. Later, the bead is taken to the tip of the micropipette, and immobilized by air suction using a syringe connected to the micropipette. The process is repeated for the second bead, which is brought in the sorroundings of the micropipette's tip (the so-called experimental area). Beads are approached to each other until a single molecular connection is made (a single molecule is tethered between the two beads).

The optical trap is generated close to the center of the microfluidics chamber, avoiding any surface effects (either hydrodynamics or optical ones) from the glass coverslips. All the excess of flowed buffer or beads

are collected in a small, 50ml, waste bin, to which the exit tubes are connected.

2.2.2.2 *Microbeads*

The beads used in the experimental setup are of two different sizes in order to make them visually distinct on the microscope. Also, each type of beads are coated with different molecules, which are going to be bound specifically to a different end of the molecule of interest. In general, this is done by working with 2.1µm polystyrene beads (Spherotech) coated with Streptavidin. Streptavidin binds specifically to Biotin, reaching forces up to ~ 100pN before breaking. On the other end of the biomolecule, a tail of Digoxigenins is placed. These Digoxigenins attach to their antibody, anti-Digoxigenin. We use a coating protocol to cover the surface of the G-coated 3.1µm polystyrene beads (Spherotech) with the antibody.

2.2.3 *Motors*

In order to perform any type of single-molecule experiments, movement of the optical trap around the chamber is required: one type of beads needs to be placed in the tip of the micropipette, while the other needs to be moved in the proximity of the first one, as well as the trap position needs to be precisely controlled in pulling experiments. There are two independent movements that can be performed in the miniTweezers setup:

- Using the motorized stage: Three motors (Thorlabs Z606) are attach in the stage where the chamber is held to allow independent movement in x, y and z. 1 revolution of the screw retracts or extends the screw by 0.05µm. These motors have a wide range of motion (~ 500µm), and are used to performed the large movements, such as trapping beads or moving around the chamber.

- Piezoelectric activators of the wiggler: the optical fiber that exits the laser is held in place by two concentric brass tubes fixed to a solid block (the wiggler). Two piezoelectric actuators push the outer tube, inducing a tilt at the fiber, changing the direction of the light emerging from the bare end of the optical fiber. The movement of the lasers in the chamber created by this tilt is short, $\sim 8 - 12$µm. The range depends on how tight the piezoelectric activators are held in mechanical contact with the wiggler.

2.2.4 *Calibration*

2.2.4.1 *Force calibration*

Each laser beam follows the optical path described in Sec. 2.2.1, exerting a force in a trapped bead inside the chamber. The applied force is translated in the deviation of each laser beam, due to linear momentum conservation, as described in Sec. 2.1. The deviated light then leaves

the chamber and gets to the x-y force-PSDs. The PSD values are a direct measurement of the force in the x or y direction[10] exerted to the trapped bead (with respect to an arbitrary position corresponding to the zero-force baselone):

$$f_{x,y} = C_{x,y} PSD_{x,y} \qquad (2.6)$$

where $C_{x,y}$ are the calibration factors for each component and $PSD_{x,y}$ are the sum of the x, y components of both lasers. The calibration factors are independent of the performed experiments (e.g. independent of each bead diameter, solvent and bead refractive indices). These can be determined after a given known exerted force is applied and used thereafter in regular experiments. Several methods have been used in order to calibrate the force PSD detectors. The most widely used are shown in Fig. 2.6:

STOKES LAW   The independent movement of the chamber with respect to the trap position (Sec. 2.2.3) can be used to create a relative displacement of the bead with respect to the surrounding fluid. This allows to exert forces in each of the three spatial directions by simply moving the chamber. It is well known that a particle that moves relative to a fluid experiences a drag force, proportional to the velocity at which it is moving. The Stokes law predicts a linear relationship between the x and y component of the drag force, $f_{x,y}$, exerted on an spherical particle of radius R when the particle is moving at a velocity $v_{x,y}$ in a viscous fluid (with viscosity η). It is given by

$$f_{x,y} = \gamma\, v = 6\pi\eta\, R\, v_{x,y}, \qquad (2.7)$$

where γ is the drag coefficient of the particle. Typical values in our experiments are $\eta = 8.9 10^{-4}$Pa·s (for distilled water at 298K), $R = 3.0\mu$m. Since the movement of the chamber is controlled by the displacement of the stage (due to well characterized movements of the x and y motors), by varying the velocity, the exerted force will vary too. As it is shown in Fig. 2.6A, we can compare the registered signal in the force-PSD detectors with the corresponding velocity of the chamber, from which the calibration factors can be obtained.

DNA OVERSTRETCHING   When stretching a double-stranded DNA (dsDNA) molecule from its opposed ends with a large force of ~ 70pN, the molecule experiences a highly cooperative transition, increasing the molecular extension by $\sim 70\%$ [24]. The longitudinal stress at which this process occurs depends on the sequence [**BosaeusdsDNA**], temperature [25] and salt concentration [26], but for a sequence with roughly 50%GC content, at $T = 25°$C and [NaCl] $= 1$M, this force is about ~ 67pN. Therefore, by overstretching a DNA molecule, as

---

10 The z component of the force, f_z, is determined by the size of the spot reaching the PD. The exact relation between its signal and the z component of the force is described in [5].
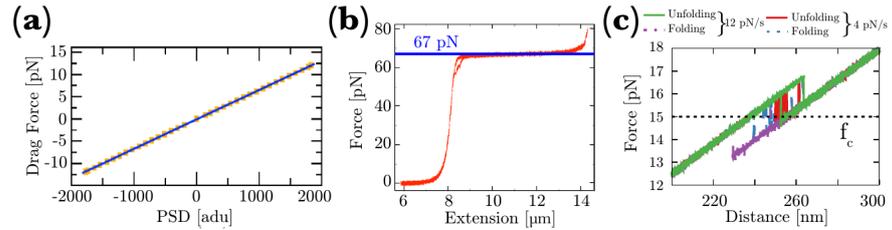
Figure 2.6: **Force calibration of mini tweezers setup.** (a) Stoke's law calibration: The drag force of the relative movement of the particle with respect to the chamber (obtained from Eq. 2.7 is plotted against the registered signal of the PSD detectors. (b) Force-extension curve (FEC) for the half λ dsDNA molecule overstretching (24805bp). The force plateau at which it happens is $\sim$ 67pN, while the extension varies $\sim$ 5.8µm, corresponding to a $\sim$ 70% of the contour length. (c) Force-Distance curve (FDC) for several unfolding(green) and folding (red) cycles for the CD4 hairpin. The measured coexistence force is around 15pN. Figure from [13].

shown in Fig. 2.6B (in our case a 24508bp dsDNAfragment from λ-phage), and comparing the force at the middle of the transition with the expected 67pN, the calibration factor can be measured. This test can also be used for checking the correct distance calibration (described in Sec. 2.2.4.2), since the whole overstretch transition corresponds to about $\sim$ 0.25nm/bp. The measured distance corresponds to the real change in molecular extension, since the force in the transition is kept almost constant and the bead displacement in the trap is negligible (see Sec. 2.2.5).

DNA HAIRPIN UNZIPPING   Another typical experiment is the unzipping of a short DNA hairpin. A hairpin is a segment of dsDNA where the 5'-end and 3'-end of both strands are connected via a short segment of ssDNA, called a loop. When exerting a high enough force from the loop-free ends of a DNA hairpin, the molecule unzips, releasing ssDNA. When the hairpin unzips, the force suddenly drops, and when the hairpin forms back the force increases again. The coexistence force, f_c, is the force at which the unfolded (ssDNA) and folded (hairpin) states are equally probable, and its value depends on the sequence, temperature and salt concentration. As shown in Fig. 2.6C, for the CD4 hairpin (constisting of 20bp and a tetraloop) this coexistence force is $\sim$ 15pN[27] (at $T = 25°C$ and [NaCl] = 1M). This unzipping forcecan be used to adjust the calibration factor (Eq. 2.6).

2.2.4.2   *Distance calibration*

Two different mechanisms permit moving the optical trap with respect to the chamber, either by using the motors connected to the stage at which the chamber is held or by moving the wiggler tip by applying a voltage at the piezo-electric actuator in the wiggler (detailed in Sec. 2.2.3). The wiggler allows for a great precision (on the order of $\sim nm$), and it can be calibrated by taking advantage of the known movement of the

motorized chamber. To determine the calibration factor between the applied voltage in the piezo-electric actuator in the wiggler and the trap displacement, a bead is held by suction on the tip of the micropipette, while the lasers are positioned on top of it. Then, the software is used to keep a zero-constant-force exerted on the bead, which guarantees that the trap position with respect to the bead will not change. Afterwards, by moving the chamber up-down (left-right)[11], the piezo-electric motors have to tilt the tip of the wiggler to induce the equivalent movement of the laser in the chamber. The registered movement in the chamber is then related to the change in the position-PSD registered signal that has been required to compensate for the chamber's movement.

2.2.5 *Trap stiffness determination*

As explained in Sec. 2.2.4.2, the miniTweezers are capable of measuring the position of the optical trap. This position corresponds to the center of the trapped bead at zero force. However, when an external force is applied, the bead will displace from the trap center until this force is compensated by the one exerted by the optical trap. For forces of the order of a few tens of pN (i.e. small displacements), the optical trap can be approximated as an harmonic potential, with a trap stiffness k_b. This allows us to relate the displacement of the bead in the trap, Δx_b, with the applied force: $f = k_b \Delta x_b$. The correct determination of the bead displacement is critical in determining the molecular extension, as it is discussed in Chapter 5.

The bead in the optical trap can be considered as a Brownian particle in a harmonic potential. The power density of force, S_f is expected to follow a Lorenzian distribution:

$$S_f(\nu) = \frac{k_b T}{2\pi^2 \gamma} \frac{k_b^2}{\nu^2 + \nu_c^2}, \tag{2.8}$$

where ν is the frequency and ν_c is the so-called corner frequency ($\nu_c = k_b/(2\pi\gamma)$). The main experimental limitation to measure $S_f(\nu)$ is given by the force-acquisition frequency, that needs to be higher than ν_c. The miniTweezers electronics board allow us to go up to 50kHz acquisition[12]. To obtain the stiffness of the trap at zero force, several seconds of the force signal of the optically trapped bead without any other external force are acquired. A fast Fourier transform algorithm is computed from the obtained data and the log-spaced in time averaged force power spectrum is fitted to the Lorenzian distribution of Eq. 2.8, as shown in Fig. 2.7. The typical range of values for the stiffness is $0.06 \leq k_b \leq 0.09$pN/nm, depending on the bead's size and laser power and optical alignment. This value of the stiffness should be compared to that of the ssDNA.

---

11 Due to its nature, the wiggler does not allow for z-movement, hence z displacement of the trap is not possible.

12 Despite digital signal processed by the software is obtained at a rate of 1kHz, the frequency can be increased by connecting directly an analog voltage signal detector to the electronic board. This voltage signal can be converted to force by callibrating it, applying constant force in the bead, while registering the force signal in the software and mapping it to the voltage signal from the electronic board.
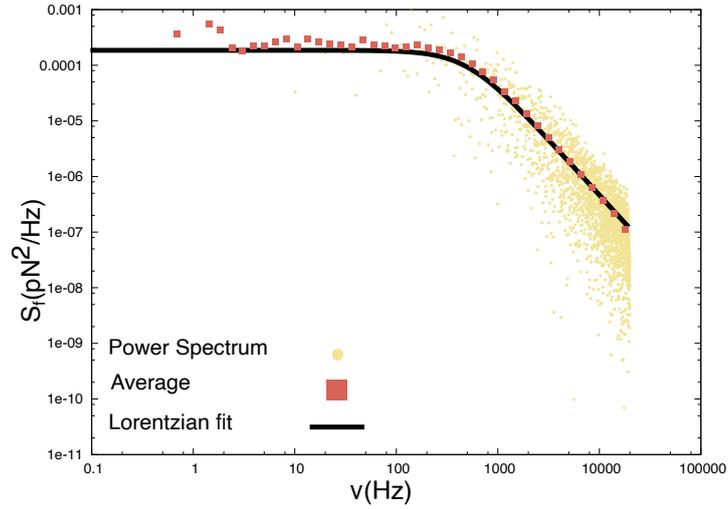
Figure 2.7: **Stiffness determination of the optical trap:** Registering up to 50kHz data of an optically trapped bead allows us to obtain the stiffness of the optical trap at zero force from fitting Eq. 2.8 (black, line) to the Fourier transform of the registered force signal (orange) and a log-spaced average of the power spectrum (red points). The obtained values for the fit are $k_b = 0.081 \pm 0.04$pN/nm and $\gamma = 3.12 \pm 0.08$pN·s/nm. The upper limit in the frequency of this power spectrum is at 20kHz, the force-acquisition frequency at which this data was taken.

This trap stiffness falls within the range given by the molecular stiffness of the ssDNA molecules studied in the thesis: ~ 10pN/nm (high forces for ssDNA of tens of bases) and $\sim 10^{-3}$pN/nm (low forces for ssDNA of tens of thousands of bases).

3

BIOMOLECULES

There is a wide variety of molecules that appear in molecular biology: from carbon hydrates, liquid lipids, proteins and nucleic acids to the main energy source of the cells, ATP. The huge varitation in composition, size and structure makes biomolecules prone to be studied using complementary techniques (X-ray, NMR, fluorescence,...).Within them there are force spectroscopy techniques which allow to manipulate individual molecules and study their mechanical response. Force spectroscopy techniques are specially suited to study nucleic acids and proteins, which are linear polymers of repeating units (monomers) that fold into specific forms in order to achieve their particular functions. In this chapter we introduce their structure and general biological functions, as well as the relationship between them.

3.1 THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

Despite first being formulated by Francis Crick (1916-2004) in 1957, the central dogma of molecular biology is usually stated in the way of James Watson (1928). In the first edition of his book *Molecular biology of the gene*[28], Watson claimed that the flow of information inside the cell is linear. In this formulation, DNA holds all the required information for a living organism, which is replicated in order to allow its reproduction.

Then, in order to accomplish the function that was meant to achieve, the information is trancripted to messenger RNA (mRNA), which acts as an intermediate before its translation to proteins. Proteins, in this frame of reference, are the final elements which develop the function encoded inby the genetic information contained in the DNA.

This linear scheme, depicted in Fig. 3.1A, allows for an easy understanding of the different roles performed of these three type of molecules: DNA acts as an information-holder, RNA acts as the intermediate between this information and the function it was meant to achieve via transcription into a protein. However, as it often happens in biology, the picture is not that simple. Other pathways of information, such as the reverse transcription (from RNA to DNA)[29]; the discovery of other types and functions of RNA (enzymatic RNAs) [30]; or the presence of processes which are able to modify nucleic acids or protein structures, without affecting the DNA sequence (epigenetics)[31].

The complexity of the interaction between DNA, RNA and proteins (Fig. 3.1B) is now better understood. The knowledge gain has been exploited from the medical/diagnose point of view. For instance, reverse transcription of RNA is a critical step in performing Polymerase Chain Reactions (PCRs) that allow us to detect SARS-CoV-2 copies in humans.

Despite all these developments, the first statement of the central dogma still holds true. In 1970, Crick wrote [32]: " *The central dogma*
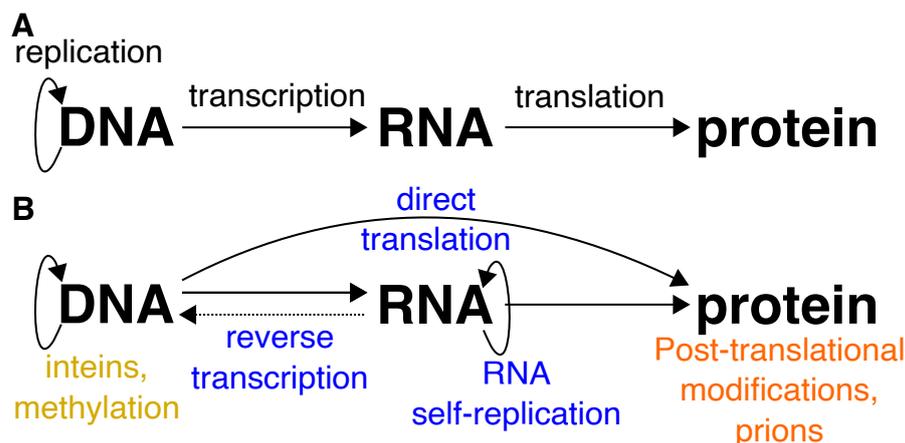
Figure 3.1: **The central dogma of molecular biology.** (a) The linear scheme of the classic central dogma: DNA is replicated in order to allow cells to reproduce. DNA is transcripted to RNA, which in turn is translated into proteins, which are the working machinery in the cell. (b) A more realestic scheme of all possible interactions and transfer of information inside the cell. Activities such as reverse transcription, ribozymes or post-translational modification substantially alter the linearity of 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."

3.2   NUCLEIC ACIDS

All organisms contain the instructions to construct their bodies as well as keeping their functions. The series of genetic instructions, called genome, have to be copied with accuracy during the process of reproduction to be transmitted to the next generations.

The genetic information is codified by nucleic acids. The nucleic acids are nucleotide polymers which are linked by a phosphodiester bond from de 3' sugar carbon of one nucleotide to the 5' sugar carbon of the following nucleotide (Fig. 3.2A). For convention, the directionality of the bond is taken from 5' to 3', which is the way all sequences are written. The primary structure of a nucleic acid is defined as the base sequence composing the nucleotide chain, read from 5' to 3'. In general, the primary structure of any DNA or RNA is defined as the sequence of nucleotides that compose the whole chain.

The only difference between DNA and RNA is found in:

- The RNA sugar ring is a ribose while DNA contains a deoxyribose. Therefore, a H group in the C2' position is found in DNA, while an OH hydroxil group is found in the same position for RNA. The hydroxil group gives RNA more reactivity than DNA.

- DNA has 4 nitrogenous bases (nucleobases or simply bases): Adenine, Guanine, Cytosine and Thymine (Fig. 3.2). RNA has also 4 bases, but Thymine is replaced by Uracil.
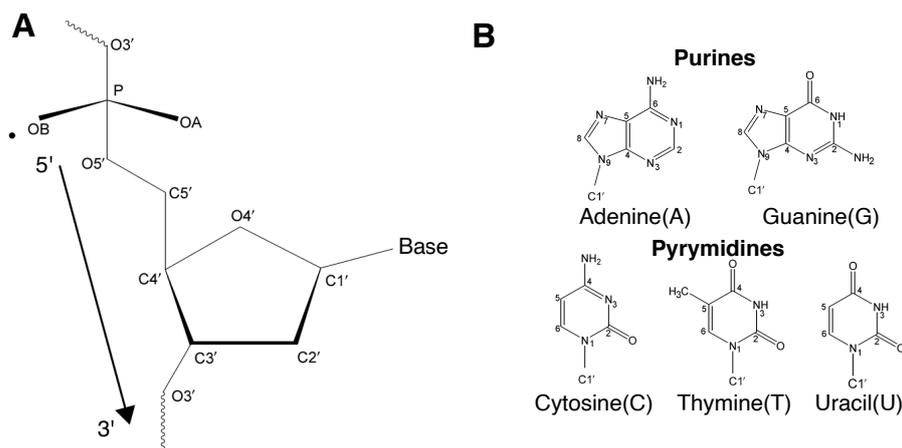
Figure 3.2: **Structure of the nucleotides** (a) Structure of a nucleotide. The directionality of the chain is given, by convention, from the 5' end to the 3' end. (b) Nitrogenous bases that make the DNA and RNA sequence, grouped in double-ring structures (purines) and simple-ring (pyrymidines). Figures from [33].

3.2.1   Nucleic Acids Structure

3.2.1.1   Double helix

DNA was firstly isolated by Friedrich Miescher (1844-1895) in 1869 and later identified as the genetic material carrier by Oswald Avery (1877-1955), Colin MacLeod (1909-1975) and Maclyn McCarty (1911-2005), in 1944. In 1953, Watson and Crick proposed the structure of double-stranded DNA, inspired from X-ray diffraction data from Rosalind Franklin (1920–1958).

In B-DNA the two strands form a right-handed double helix held by hydrogen bonds between complementary bases, forming the so-called Watson-Crick (WC) bonds. As shown schematically in Fig. 3.3, the Adenines of one strand form two hydrogen bonds with the Thymines in the other[1], while the Guanines form three hydrogen bonds with the Cytosines. The higher number of hydrogen bonds is related to the larger stability found in G-C basepairs.

There are other factors contributing to the stabilization of the DNA double helix: the interactions between the phosphate group (negatively charged) and the water molecules and ions in the buffer, and the hydrophobic interactions between adjacent bases that leads to the so-called base stacking, which will be further developed in Sec. 3.2.1.3.

B-DNA is the predominant form of dsDNA, which has been widely studied [34]. As shown schematically in Fig. 3.4A (left) B-DNA is a right-handed double helix that presents a rise along the axis of 0.34nm/bp a helix diameter of 2.0nm and it requires of 10 bases to complete a whole turn. The sugar rings present in this structure are in the C2'-endo conformation (See Sec. 3.2.1.2). However, dsDNA can also be presented

---

1 For dsRNA, which is found in some RNA viruses, Uracils are present instead.
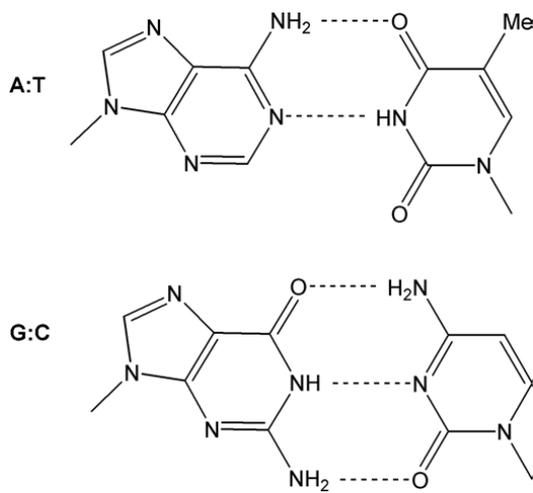
Figure 3.3: **Basepairing structure.** Schematic representation of the two hydrogen bonds that form during an A-T basepair (top) and the three that stabilize a G-C basepair. Figure from [33].

in other forms, among which the most commonly found in nature are A-DNA and Z-DNA.

The A-double helix (Fig. 3.4B, right) is a right-handed double helix with a more compact and wider structure: it has a rise of 0.25nm/bp, a helix diameter of 2.3nm and 11 bases to complete a whole turn. It is usually associated to lowly hydrated solutions (e.g. with ethanol present), and in specific DNA sequences. It has also been found that the A-form is present in dsRNA or hybrid RNA-DNA helices.

The Z-double helix is a left-handed double helix in which the helix winds in a sigzag pattern, with the highest rise (0.38nm/bp) and bp/turn (12) and the shortest diameter (1.8nm) of the three structures described above.

All double-helix structures are compatible with the Chargaff relations, that state that the total number of purines in DNA equals the number of pyrymidines.

3.2.1.2 *Sugar pucker*

The five carbons in the deoxyribose(ribose) form a inherently nonplanar sugar ring in DNA and RNA [33]. This nonplanarity is termed puckering. The energetically most stable conformation for the ring has all substituents as far apart as possible, arising from the effect of nonbonded interactions between them. The puckering can be described precisely in terms of the ring internal torsion angles, although a simple qualitative description of the conformation in terms of atoms deviating from ring coplanarity is widely used. If the major deviation is on the same side as the base and C4'–C5' bond, then the atom involved is termed endo. If it is on the opposite side, it is called exo. In Fig. 3.5 the two most
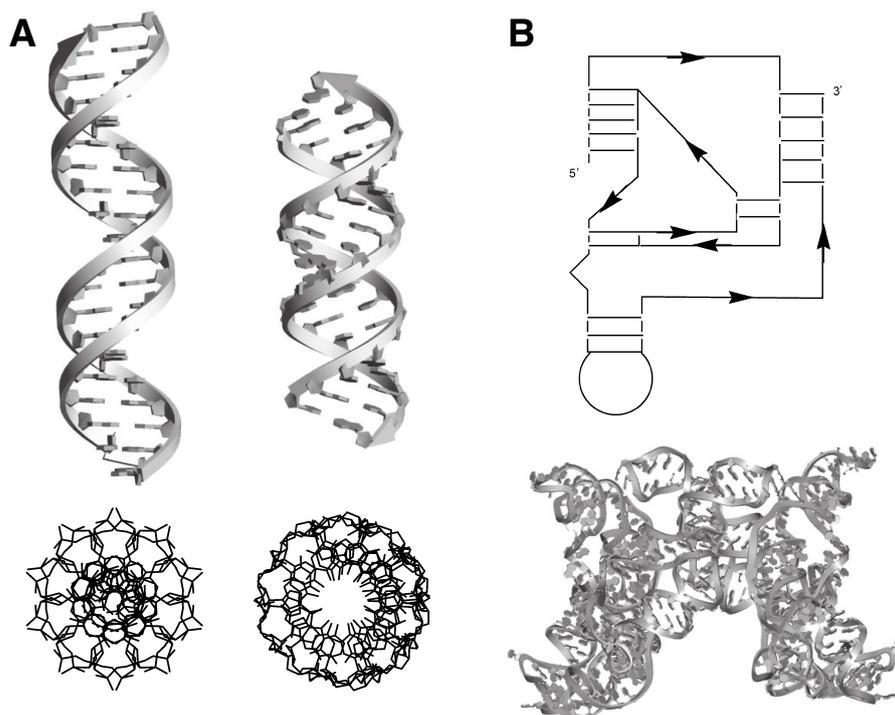
Figure 3.4: **Structure of the double helix** (a) B-DNA (left) and A-DNA(right) double helix forms. In the bottom of each double helix structure, there is a view from top of the double helix in the B-DNA (top) and A-DNA (bottom) conformations. (b) Example of a secondary structure of RNA: hepatitits delta virus ribozyme topology (top) and crystal structure (bottom) [35]. Figures from [33].

commonly observed puckers[2] in crystal structures of isolated nucleosides and nucleotides C2'-endo or C3'-endo types are shown.

As can be observed in Fig. 3.5, the interphosphate distances of this two structures is different. For the case of C2'-endo, the observed distance is 0.7nm, while for the C3'-endo is 0.59nm [34]. This distance corresponds to the observed one from B-DNA and A-DNA, respectively. Also, purines show a preference for the C2'-endo pucker conformational type whereas pyrimidines favor C3'-endo. Deoxyribose nucleosides are primarily (> 60%) in the C2'-endo form and ribonucleosides favor C3'-endo. [33]. Another short set of names is given to both configurations: south (C3'-endo) and north (C2'-endo).

3.2.1.3   *Base stacking*

As stated previously double helix stabilization is not only due to base-pairing. The electrostatic interactions between adjacent bases, called base stacking, are also critical. Stacking forces depend on the aromaticity of the bases and their fluctuating dipole moments (London dispersion forces). For this reason, the strength of the base stacking interaction de-

---

2 In practice, these pure envelope forms are rarely observed, but most observed structures tend to one of these two forms. Further explanation can be found in [33].
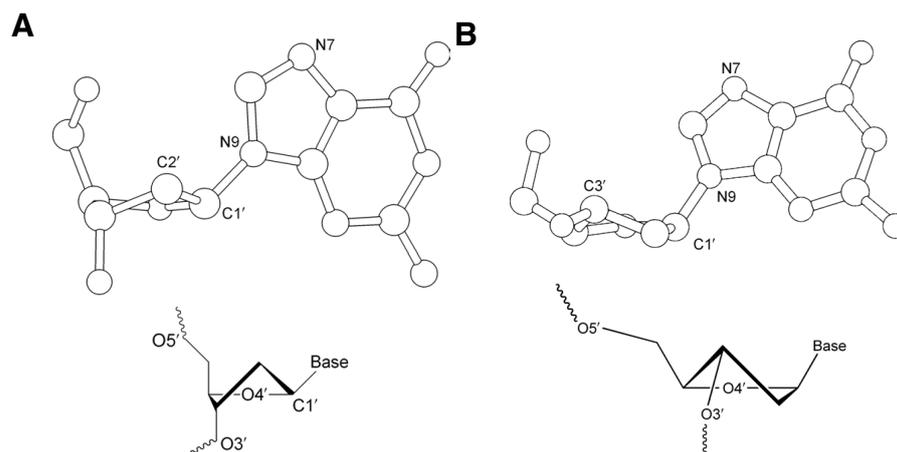
Figure 3.5: **Sugar Pucker.** Schematic depiction of the atoms in the sugar ring in nucleic acids: (a) C2'-endo conformation, also known as south conformation (b) C3'-endo conformation, also known as north conformation. Figures from [33].

pends on the DNA sequence, making consecutive purine bases, with their double aromatic ring in the base, more prone to stack than single-ring pyrymidines. Nearest neighbour base-stacking interactions are important determinants of DNA double strand stability [36]. Base stacking has been also found to play a role in ssDNA and RNA structures.

Because of its electrostatic nature, base-stacking interactions increase with increasing salt concentration, as high salt concentrations mask the destabilising charge repulsion between the two negatively charged phosphodiester backbones.

BASE STACKING STABILITY IN DOUBLE HELIX   A qualitative argument regarding the relevance of base stacking in the double helix stability can be made comparing the ssDNA structure in solution with the dsDNA one. On the one hand, ssDNA in solution will have the bases exposed, allowing for the formation of hydrogen bonds between them and the water molecules. On the other, the stacking of ssDNA is weaker than in dsDNA. Therefore, the free-energy difference between the base-paired double-helix and the random coil ssDNA can not arise from basepairing, since in both states, the bases tend to form hydrogen bonds. Nevertheless, if the formation of these hydrogen bonds is prevented, mainly from having non-complementary sequences, the penalty in free-energy is going to be large (as discussed in Sec. 3.2.1.1). Furthermore, the key ingredient of double helix stability is base stacking, as it is the main difference between the ssDNA structure and dsDNA structure. The essential role of base stacking in the double helix stability has been also proved experimentally in [37].

3.2.1.4 *Other secondary structures and higher order structures in nucleic acids*

Base-pairing in DNA is often considered solely in terms of Watson–Crick hydrogen-bonding, but a large number of other arrangements are possible, and many have been observed experimentally[33]. Besides A-T and G-C Hogsteen and reverse Hogsteen pairs, a whole array possible arrangements, i.e. mismatches, can appear: G-A, G-T, A-A, G-G, T-T, C-C, T-C or A-C. Mismatches generally destabilize duplex DNA as compared to fully-complementary sequences. The difference in energies between mismatches is broad, allowing for keeping the B-form structures in some of them [38]. This can be quantified by using calorimetry [39], as well as force spectroscopy techniques [40, 41]. Obtaining thermodynamic parameters from mismatched structures (non-specific secondary structures) from force spectroscopy measurements is described in Chapter 8.

Despite being mostly found in double-stranded form, DNA is also found in single-stranded form inside some virus or at the end of telomers. Moreover ssDNA is genereted every time the information from the dsDNA needs to be read for its transcription. The flexibility of ssDNA and its basepairing interactions form a rich variety of secondary structure motifs, consisting in helices, loops, bulges and junctions.

On the other hand, RNA is mostly found in single-stranded form with the main exception of dsRNA being present in some viruses. The ability of many RNAs to fold into compact tertiary structures, analogous to the folding of proteins, together with the possession of catalytic activity, found in some RNAs, lead to Alexander Rich to propose the concept of the RNA world in 1962 (concept coined by Walter Gilbert in 1986). The RNA world hypothesizes that life started with self-replicating RNA enzymes [42].

RNAs can have a complex tertiary structure. An example is the hepatitis delta virus ribozyme, shown in Fig. 3.4B . Even short RNA sequences are very prone to secondary structure formation.

DNA also presents higher order structures. For instance, circular DNA in bacteria is found to be in a negative supercoiled state, a type of tertiary structure. The packing of DNA in eukariotic cells is also considered to be a quaternary structure, with the interaction of DNA with DNA-packing proteins, such as histones.

3.3 PROTEINS

Proteins are large biomolecules, consisting of one or more long chains of amino acid residues. Proteins perform a vast array of functions within organisms, including catalysing metabolic reactions (e.g. DNA replication) and providing structure to cells as well as transporting molecules. Proteins differ from one another in their sequence of amino acids, which is dictated by the nucleotide sequence of their genes, and which usually results in protein folding into a specific 3D structure that determines its function (see below).

Once formed, proteins only exist for a given time and are then degraded and recycled by the cell's machinery through the process of protein turnover. They can exist for minutes or years with an average lifespan of 1–2 days in mammalian cells[43].

Gerardus Johannes Mulder (1802-1880) was the first to describe proteins, later named by the Swedish chemist Jöns Jacob Berzelius (1779-1848) in 1838. The term 'protein" derives from the Greek word $\rho\omega\tau\epsilon\iota o\sigma$ (*proteios*), meaning "primary" or "of the first rank", which Berzelius proposed because of the key role of proteins in living organisms.

3.3.1 *Structure*

Like the nucleotides in DNA and RNA, aminoacids are the building blocks of proteins. They consist of a carbon atom, surrounded by an amino and a carboxyl group and a side chain R, as shown in Fig. 3.6A. The individual amino acid residues are bonded together by peptide bonds, which is a covalent bond that takes place between the amino group of one aminoacid and the carboxyl group of another one (Fig. 3.6B). The sequence of amino acid residues in a protein is defined by the sequence of a gene, which is encoded in the genetic code. In general, the genetic code specifies 20 standard amino acids, which in certain organisms can include one more: selenocysteine and –in certain archaea– pyrrolysine. Shortly after or even during synthesis, the residues in a protein are often chemically modified by post-translational modification, which alters the physical and chemical properties, and hence may modify its folding, affecting the function of the proteins. Proteins can also perform a particular function by working together, which leads them to often associate to form stable protein complexes.
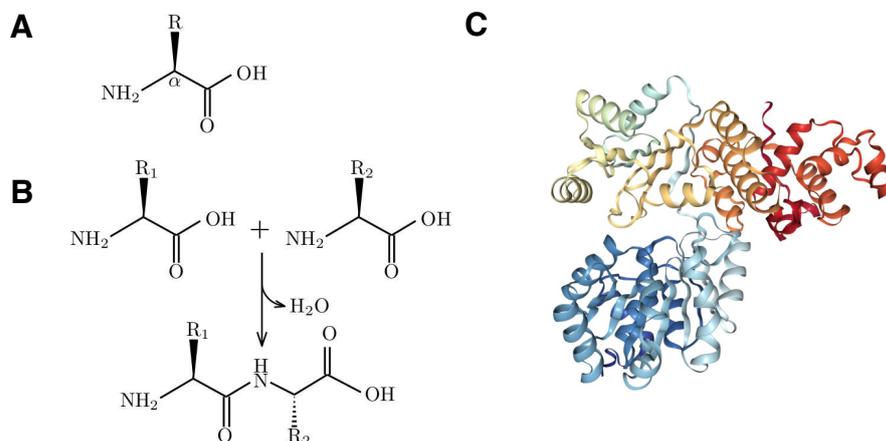
Figure 3.6: **Aminoacid and protein structure.** (a) Chemical structure of an aminoacid (B) Representation of the formation of the peptidic bond between two aminoacids (c) E.coli RecQ helicase catalytic core 3D structure, obtained from crystallography [44]. Figure from [27].

Proteins have four levels of structural organization:

- The primary structure is the sequence of aminoacids of the polipeptidic chain.

- The secondary structure corresponds to the local regular folding of the chain. The two main structures of this level are the α-helices and β-sheets.

- The tertiary structure is the general 3D structure that a single polipeptidic chain adopts.

- The quaternary structure is the final configuration of oligomeric proteins, adopted by the several polipeptidic chains that can form a protein.

3.3.2 Helicases

Helicases are a class of enzymes that are in charge of unpacking organism's genes. They are motor proteins that move directionally along a nucleic acid phosphodiester backbone, separating two annealed nucleic acid strands (DNA and RNA), using energy from ATP hydrolysis. There are many helicase types, representing the great variety of processes in which strand separation must be catalyzed: DNA replication, transcription, translation, recombination, DNA repair, and ribosome biogenesis.

The separation of strands of a DNA double helix or a self-annealed RNA molecule involves the breaking of hydrogen bonds between complementary nucleotides and therefore requires of using the energy from ATP hydrolysis. Helicases move along one nucleic acid strand of the duplex with a directionality and processivity specific to each particular enzyme. In Chapter 8 we study the activity of a DNA helicase (E.Coli RecQ) using optical tweezers.

ELASTIC MODELS FROM POLYMER THEORY

The International Union of Pure and Applied Chemistry (IUPAC), [45], defines a polymer as: " a molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetition of units derived [...] from molecules of low relative molecular mass".

As described in Chapter 3, nucleic acids and proteins are polymers. Their primary structure (sequence of bases or amino acids) determine their function via a very complex folding of the biopolymer in a very specific 3D shape.

In this chapter we introduce the historical context of the emergence of polymer science, focusing on the approach taken from the field of polymer physics, which allows us to model the elastic behaviour of polymers when a force is exerted on their ends. The two descriptions which will be used in the thesis are the ones provided by the ideal chains modelled as the Freely-Jointed Chain and the Worm-like Chain.

4.1 POLYMER THEORY

The first steps in polymer science begun around 1830, with the development of derivatives of the natural polymer cellulose. Henri Braconnot(1780-1855) and Christian Schönbein (1799-1868) were the pioneers in the development of cellulose and cellulose acetate. Remarkably, cellulloid was discovered while investigating a replace material for the extremely expensive ivory billiard balls. The field of polymer science would receive a huge boost by the broad applications these materials present: from the invention of vulcanized rubber to the first synthetic plastic, the Bakelite.

The term polymer was coined by Jöns Jacob Berzelius (1779-1848) in 1833, although a more modern definition was proposed in 1920 by Hermann Staudinger (1881-1965), who described polymers as long chains of atoms linked by covalent bonds [46]. For this work, Staudinger would receive the Nobel Prize in Chemistry in 1953.

The first century of polymer science was basically guided by the potential applications in new materials. For this reason, the main discipline involved in such discoveries was chemistry. However, at mid 20th century, Paul Flory (1910-1985), would include concepts from physics, such as the excluded volume, into the field of polymer science. For "his fundamental achievements, both theoretical and experimental, in the physical chemistry of macromolecules" ([47]) Flory would be awarded the Nobel Prize in Chemistry in 1975.

Polymer physics also became relevant in understanding the underlying principles of techniques that had been used before. An example to this is the description of the reptation of DNA through a porous material, such as an electrophoresis gel, a technique regularly used in biomolecular

laboratories around the world for distinguishing nucleic acids segments of different length. The contribution of Pierre-Gilles de Gennes (1932-2007), Nobel Prize in Physics in 1991) with the later refinement of the theory by Sam Edwards (1928-2015) and Masao Doi (1948) were central for the development of the theoretical framework for polymer physics.

4.2   ELASTIC MODELS

Different models of polymer chains have been developed. Two main descriptions appear: ideal chains assume that there are no interactions between chain monomers, while real chains explicitly model the interactions within the chain (mainly including self-avoidance along the chain as excluded volume effects). More exhaustive details about models and derivation of the parameters can be found in [48, 49].

Several works have proposed different theoretical approaches in order to describe the response of biopolymers to mechanical stress. Within the frame of ideal chains, two main descriptions are used to model the elastic behaviour of biological polymers[24, 50, 51]: the Freeely-Jointed Chain and the Worm-like Chain.

4.2.1   Freely-Jointed chain

The most simple elastic polymer model is the Freely-Jointed Chain (FJC), which consists on N rigid bonds of length c that can orient freely in the space. In some cases, the monomer length c can have a physical interpretation (e.g. in nucleic acids c corresponds to the interphosphate distance). The model assumes that monomers act as rigid bonds and follow a random walk movement, neglecting any kind of interaction between them. Defining $\vec{r}_i$ as the vector position of the monomer i, the end-to-end vector of the monomer, $\vec{R}$, can be written as the sum of all $\vec{t}_i$,

$$\langle \vec{R} \rangle = \sum_{i=1}^{N} \vec{t}_i = 0, \tag{4.1}$$

with $\vec{t}_i = \vec{r}_{i+1} - \vec{r}_i$. In equilibrium conditions the configurational average $\langle ... \rangle$ of $\vec{R}$, $\langle \vec{R} \rangle = 0$, because $\langle \vec{t}_i \rangle = 0$. The fluctuations in each direction verify $\langle \vec{t}_i \vec{t}_j \rangle = \delta_{ij} c^2$ and therefore $\langle \vec{R^2} \rangle = Nc^2$, and the probability density function of $\vec{R}$ reads:

$$P(\vec{R}) = \left(\frac{3}{2\pi Nc^2} \right)^{3/2} e^{-\frac{3R^2}{2Nc^2}}. \tag{4.2}$$

If the chain length is long compared to the scale of short range interactions, any chain can be re-scaled into a freely-jointed chain. To do so, a new segment length named Kuhn length, b, is chosen to be long enough so that neighbouring segments are un-correlated.

When an external force $\vec{f}$ is applied to stretch the polymer, the energy of a single segment can be written as $E_i = -\vec{t}_i \cdot \vec{f}$. Each segment then contributes to the partition function as $Z_i = \int \exp \left(\frac{\vec{t}\vec{f}}{k_\mathrm{B}T} \right) \sin \theta d\theta d\phi$,
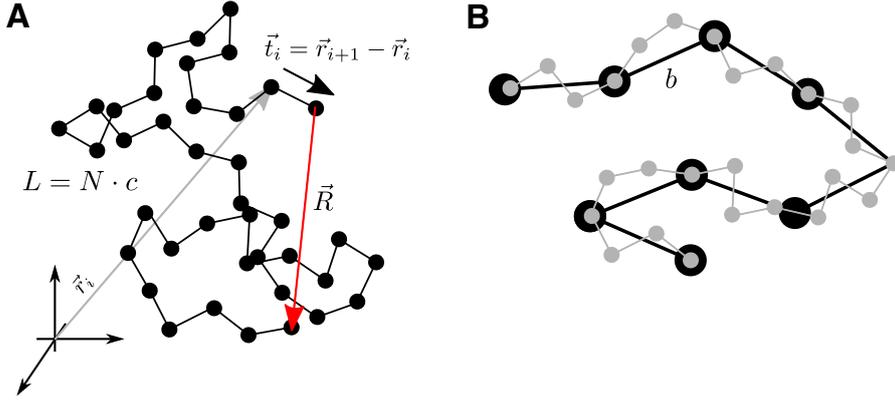
Figure 4.1: **Schematic depiction of the Freely-Jointed Chain model.** (a) Representation of a Freely jointed chain with monomer length c. The end-to-end vector, $\vec{R}$, is also depicted (b) Schematic depiction of the new re-scaling of the polymer, dividing it in longer segments, each of Kuhn length, b.

where ϕ, θ are the polar and azimutal angle of $\vec{t_i}$. The partition function of the whole system can be written:

$$Z(N', f) = \int e^{\frac{\vec{f} \sum_i^{N'} \vec{t_i}}{k_\mathrm{B} T}} \prod_{i=1}^{N} sin\theta_i d\theta_i d\phi_i \qquad (4.3)$$

We choose $\vec{f} = f\hat{z}$ so $\vec{t_i}\vec{f} = bf \cos\theta_i$. Z can be

$$Z(N', f) = \left(\frac{4\pi \sinh\left(\frac{fb}{k_\mathrm{B} T}\right)}{\frac{k_\mathrm{B} T}{fb}} \right)^{N} . \qquad (4.4)$$

The free energy is then computed as $G(N', f) = -k_\mathrm{B} T \log Z(N', f)$, resulting in

$$G(N', f) = -N' k_\mathrm{B} T \log\left[4\pi \sinh\left(\frac{fb}{k_\mathrm{B} T}\right) - \log\left(\frac{fb}{k_\mathrm{B} T}\right) \right], \qquad (4.5)$$

The extension can be directly found by deriving G with respect to f:

$$x_\mathrm{FJC}(f) = -\frac{\partial G}{\partial f} = L\left(\coth\left(\frac{bf}{k_\mathrm{B} T}\right) - \frac{k_\mathrm{B} T}{bf} \right), \qquad (4.6)$$

where the total polymer length, L, Eq. A.5 gives an explicit expression of the polymer extension, x_{FJC} as a function of the force.

It is worth noticing that the expression of the stretching energy, $G(N', f)$, given by Eq. 4.5, has been derived for a constant force ensemble, which corresponds to a constant force applied at the end of the polymer. However, this is not generally the case for the miniTweezers setup, where the control is exerted upon the extension of the molecule, instead of the

force. The stretching free-energy in the extension ensemble, where the molecular extension is externally controlled, is:

$$\Delta G_x(N', f) = \int_0^f f_{FJC}(x')dx' = f \cdot x - G(N', f), \qquad (4.7)$$

where $f_{\text{FJC}}(x')$ is the force required in the FJC model to extend the polymer an extension x' (by inverting Eq. A.5) and f is the force at which the extension x is reached, $f = f_{\text{FJC}}(x)$. A discussion around the effects on the computation of thermodynamic quantities derived from experiments under the constant force ensemble and constant extension ensemble is extensively discussed in [52].

4.2.1.1 Extensible Freely-Jointed Chain

An extension of this model, the so-called extensible Freely-Jointed Chain (ex-FJC), includes bond-stretching, by modifying the extension x_{FJC} as $x_{\text{exFJC}}(f) = x_i(f)\,(1 + f/S)$, where S is the stretch modulus, thus giving:

$$x_{\text{exFJC}}(f) = L\left(\coth\left(\frac{b\,f}{k_B T}\right) - \frac{k_B T}{b\,f}\right)\left(1 + \frac{f}{S}\right). \qquad (4.8)$$

The free energy of stretching, under a controlled force, using the extensible FJC model can also be obtained by integrating:

$$\Delta G_f(N, f) = \int_0^x x_{\text{exFJC}}(f')df', \qquad (4.9)$$

where x_{exFJC} is given by Eq. 4.8 and x is the average extension exhibited by the polymer under an externally applied force f.

The free-energy of stretching under a constant extension can be obtained using Eq. 4.7.

4.2.2 Worm-Like Chain

Another theoretical approach, the Worm-Like Chain model (WLC), is based on describing the polymer as a flexible rod. The first proposal, using a discrete formulation, was developed by Kratky and Porod in 1949 in order to describe the elasticity of single semi-flexible polymers[53]. The model was able to reproduce the force vs distance curves obtained by pulling chains of dsDNA, using magnetic and optical tweezers [10, 24].

In order to describe the elasticity of a long, thin rod, with a total length of L we consider a segment ds of the rod, located at a distance s from one end [54].

The deformations of the segment are given by three quantities:

- The extensional deformation, $u(s)$, is a scalar measuring the fractional change in length of the segment, $u = \Delta(ds)/ds$.

- The bending deformation, $\vec{\beta}(s)$, is a vector which measures how the rod's unit tangent vector, $\hat{t}$, changes along the rod: $\vec{\beta} = d\hat{t}/ds$.

- The torsional deformation, $\omega(s)$, is a scalar that measures how the rod twists an angle $\Delta\phi$ after a displacement ds along its axis, $\omega(s) = \Delta\phi/ds$.

A

B
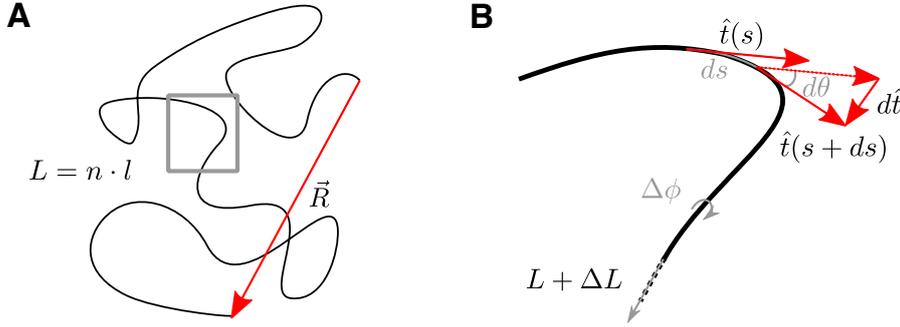
Figure 4.2: **Schematic depiction of the Worm-Like Chain model.** (a) Representation of a WLC rod, with a contour length $L = n \cdot l$. The area inside the gray squared is zoomed in on the right. (b) The three possible deformations that can experience an elastic rod. For the case of the inextensible WLC, only the bending one (regarding energy penalty in changing $\hat{t}$) is considered. The extensional deformation (inducing an extension of the contour length, ΔL) is considered in the extensible WLC (Sec. 4.2.2.1), while the twist ($\Delta\phi$) is included in other models, such as the twistable WLC [55].

We then assume that deformations are small and they can be written in a quadratic form. The energy density, dE, at a segment ds of these deformations is:

$$dE = \frac{1}{2}k_{\mathrm{B}}T\left(A\vec{\beta}^2 + Bu^2 + C\omega^2 + Du\omega\right)ds, \qquad (4.10)$$

where A is the bend persistence length, C is the twist persistence length, $B \cdot k_{\mathrm{B}}T$ is the stretch stiffness and $D \cdot k_{\mathrm{B}}T$ is the twist-stretch coupling. For the WLC model, several simplifications are done: consecutive chemical bonds along the polymer are assumed to rotate freely, so that there is no twist elasticity[1](C=D=0), and the extensional deformation, u, is assumed to be negligible at small enough forces ($f << k_{\mathrm{B}}TB$). The latter assumption can be relaxed by incorporating some degree of extensibility to the rod chain, using the extensible WLC description (see next section).

Therefore, the elastic energy of a continuous rod, assuming that both simplifications hold (B=C=D=0) is given by:

$$dE = \frac{1}{2}k_{\mathrm{B}}T\int_0^L A\vec{\beta}^2 ds, \qquad (4.11)$$

This energy describes a thin, inextensible and continuous rod. Throughout the thesis, this model will be written as the inextensible Worm-Like Chain, to differentiate it from the extensible model described in Sec. 4.2.2.1.

---

1 Twist elasticity has been shown to play a relevant role when performing torsionally constrained experiments, for which magnetic tweezers are ideal[56]. Elastic models that take into account the twist elasticity, such as the one developed in [55], have been used to sucessfully describe such experiments.

Whereas the FJC model is purely entropic, the WLC adds an enthalpic contribution which takes into account a bending penalty. The WLC has two characteristic lengths, l, the contour length per base and, p, the persistence length. The latter is defined as distance along which the tangent vector de-correlates, i.e.

$$\langle \hat{t}(s)\hat{t}(0) \rangle = \cos{(\theta(s))} = e^{-s/p}, \tag{4.12}$$

where θ is the bending angle at the position s in the rod, as shown schematically in Fig. 4.2B. The FEC of the model is not analytically solvable, but an interpolation formula for the low- and high-force[57] regimes is given by [50]:

$$f_{\text{WLC}}(x) = \frac{k_{\text{B}}T}{p} \left(\frac{1}{4}\left(1 - \frac{x}{nl}\right)^{-2} - \frac{1}{4} + \frac{x}{nl} \right), \tag{4.13}$$

where l is the contour length per base and n is the total number of monomers (with total contour length of $L = n \cdot l$). Other interpolation formulas have been proposed[58], as the one given by Bouchiat in 1999 [51], where the term $\frac{x}{nl}$ is expanded up to a 7th degree polynomial, i.e. $\sum_{n=2}^{7} a_n \left(\frac{x}{nl}\right)^n$.

The free-energy of stretching a polymer, ΔG_{WLC}, from zero force ($x = 0$), up to a force f (extension $x(f)$, obtained from inverting Eq. 4.13), for the WLC model is written as:

$$\Delta G_{\text{WLC}}(x, T) = \frac{k_{\text{B}}T}{p} \left[\frac{nl}{4}\left(1 - \frac{x}{nl}\right)^{-1} - \frac{nl}{4} - \frac{x}{4} + \frac{x^2}{2nl} \right], \tag{4.14}$$

which has been calculated by integrating Eq. 4.13: $\Delta G_{\text{WLC}} = \int_0^{x(f)} f_{\text{WLC}}(x')dx'$.

4.2.2.1 Extensible Worm-Like Chain

The only contribution to the energy of the WLC (Eq. 4.11) comes from bending. In the extensible WLC model, the extensional deformation is incorporated. To do so, a Young Modulus is added, following the same approach done in the FJC case (Eq. 4.8) leading to the proposed formula in [59] for the interpolation formula (Eq. 4.13):

$$f = \frac{k_{\text{B}}T}{p} \left(\frac{1}{4}\left(1 - \frac{x}{nl} + \frac{f}{Y}\right)^{-2} - \frac{1}{4} + \frac{x}{nl} + \frac{f}{Y} \right), \tag{4.15}$$

where Y stands for the Young modulus of the chain, $Y = k_{\text{B}}TB$ from Eq. 4.11, while the other parameters are already defined in equation 4.13. Along this thesis both the WLC model and the extensible WLC model have been used to describe ssDNA elasticity (Chapters 5-10).

4.3 COMPARISON OF THE MODELS

The values for the Kuhn length of Eq. A.5, b, and the persistence length of Eq. 4.13, p, are related. In particular, when the value of p is comparable with the segment length, l, the following relation holds [60]:

$$b = 2p, \tag{4.16}$$

When fitting these elastic models to measurements of ssDNA elasticity, different values for p and b are found ($0.5 < p < 2$nm and $1 < b < 4$nm), depending on the DNA sequence, DNA length and salt conditions.

WLC models (inextensible and extensible) have been used to fit the FEC of both ssNA, dsNA and proteins. The FJC model has all been used to describe the elastic behaviour of ssNA [24, 61], where the Kuhn length and the monomer length are comparable, $b \sim c \sim 0.5$nm, giving values of $p \sim 1$nm at [NaCl]=1M and $T = 25°$C.

The WLC, on the other hand, has been used to describe the elasticity of dsDNA and proteins [62]. Typical values for dsDNA are $l = 0.34$nm/bp and $p = 50$nm [24], while shorter values of p are obtained for the case of ssDNA (NaCl=1M): $p = 1.35$nm [63], $p = 0.75$nm [61]. For the case of ssDNA, the low persistence length makes it comparable to the interphosphate distance of DNA, allowing to hold the ideal chain relation: $b = 2p$. The results reported ($0.5 \leq p \leq 2$nm, $1 \leq b \leq 4$nm)[24, 64] are compatible[65].

Apart from characterizing the elasticity of biopolymers, these models are used to determine the stretching free-energies of both models (Eqs. 4.9 and 4.14), which can be used in force-spectroscopy experiments to subtract the elastic contributions of the molecules being studied.

Part II

SINGLE-STRANDED DNA

5

ELASTIC PROPERTIES OF SINGLE-STRANDED DNA

The elastic properties of biomolecules play a central role in several processes occurring at the molecular scale. Examples are DNA packaging inside the nucleus, which is crucial to understand the mechanisms of packaging and reading of the genetic information, and the formation of actin filaments inside the cytoplasm, which is essential for nutrient transport and cell structure [27]. Furthermore, the elasticity of biomolecules is intrinsically related to their molecular structure, which in turn determine their biological function.

DNA is a nucleic acid polymer encoding the genetic information of organisms. In its most common conformation DNA forms a double-stranded helix (dsDNA) stabilized by hydrogen bonds and base-stacking interactions. However DNA can also be present in single-stranded form (ssDNA). ssDNA occurs in many biological processes requiring the reading of the genetic information, such as DNA replication, transcription and repair. The double-stranded and single-stranded forms present very different biochemical and mechanical properties: whereas dsDNA is a pretty stable and stiff polymer, ssDNA is a more reactive and flexible polymer. Force-spectroscopy techniques have been widely used to study biochemical and enzymatic processes involving DNA, such as unzipping of DNA structures, binding of ligands to DNA, unwinding of DNA catalyzed by helicases, etc. In these experiments extracting useful information (e.g. free energies of DNA conformations, position of ligand binding along DNA or determining the translocation rate of DNA motors) requires a detailed knowledge of the mechanical response of ssDNA. Many studies have addressed the elastic properties of dsDNA, however, much less is known about the elasticity of ssDNA. Single-molecule force-spectroscopy techniques have allowed the measurement of ssDNA elastic response. The measured FECs have been fitted to elastic models (WLC, FJC) [24, 61].

The value of the ssDNA persistence length (i.e. the distance over which the polymer bends by thermal forces) falls in the nanometer range) being comparable in magnitude to the interphosphate crystallographic distance (~ 6Å). As as result, the elastic response of ssDNA can be well described with the extensible-FJC (the Kuhn length and the stretching modulus being the parameters of the model) or with the inextensible WLC (with the persistence length and the interphosphate distance being the parameters of the model). In comparison, dsDNA has a much larger persistence length (~ 50nm) which is 150 times larger than the interbase distance (~ 3Å). The disparity of the two lengthscales underlines the importance of bending stiffness in dsDNA (the enthalpy gain upon straightening DNA by pulling). Consequently dsDNA cannot be fit by the FJC description and the WLC must be used instead[10, 50].

In contrast to dsDNA, a large dispersion in the ssDNA persistence and contour lengths has been reported from different experimental techniques and different DNA sequences [24, 61, 64, 66–69]. In particular it has been shown that base stacking affects the elasticity of ssDNA and ssRNA with different purine/pyrimidine content [70–72]. Moreover, single molecule force studies have reported a systematic dependence of the ssDNA persistence length p on the DNA contour length L, with larger p values for short molecules (of few tens of bases) [63, 73–75] and shorter values for longer ones (of few hundreds or thousands of bases) [24, 61, 66, 76–79]. Their apparent length dependence is the opposite of that observed in dsDNA[80, 81]. A dependence of p on L in ssDNA is quite unexpected as p spans at most two bases along the phosphate backbone, which is much shorter than the length of the polymer. How then the value of p can be sensitive to the length of much longer polymers?

Force-spectroscopy techniques have been used to estimate the free energy of DNA native structures, kinetic intermediates and misfolded structures. A paradigmatic case is the determination of the nearest neighbour base-pair energies in DNA to 0.1 kcal/mol accuracy using optical tweezers. This is done by pulling the two strands apart of a several thousands bp DNA hairpin (unzipping) and measuring the FECs at different monovalent [82] and divalent [83] salt conditions. In order to estimate the base-pair energies one needs to subtract the free energy of stretching of the released ssDNA during the unzipping reaction. This shows how important is to accurately determine the elastic parameters of ssDNA.

In this chapter we investigate the effects of molecular length on the ssDNA elasticity.

5.1   MOLECULES STUDIED

The molecules we have studied are based on DNA hairpins, each with N bases in total. Long hairpins ($N > 500$b) are named $H_N L$, with varying sequences in them and a tetraloop that also varies between them. Short hairpins formed of the same 20bp stem, with varying sequences in the 20b-loop: one hairpin has a loop consisting of a single guanine, followed by 19 adenines, while the other has a single guanine, followed by 19 thymines. Respectively, they are named $H_{60}A_{19}$ and $H_{60}A_0$, with the second number indicating the number of consecutive adenines in the loop . We perform two different preparations for the DNA substrates (see App. B for details). The short hairpins $H_{60}A_{19}$ and $H_{60}A_0$ are synthesized by annealing and ligating a different set of oligonucleotides. The longer hairpins, H_{700}, H_{964}, H_{1904}, H_{4508}, H_{7192} and H_{13680} are synthesized using a long DNA fragment obtained either by PCR amplification or by digestion of the linearized λ-phage DNA. All hairpins are annealed to 29bp dsDNA handles that are used as spacers. Labeling of the handles is achieved by a Digoxigenin/Biotin tailing using a terminal transferase.

5.2 ELASTIC PROPERTIES OF LONG ssDNA MOLECULES

ssDNA easily adsorbs non-specifically on surfaces. This makes difficult to manipulate them and to precisely control the number of ssDNA nucleotides that are effectively stretched. In order to avoid this problem, we use DNA hairpins that we mechanically unfold to generate ssDNA. We first mechanically unzip a DNA hairpin by applying a force above 15 pN at the extremities of the hairpin molecule. We next reduce gradually the force and use a loop-complementary blocking oligonucleotide to prevent the rezipping of the hairpin, as described in next section. In Fig. 5.1A we show the unzipping of the H_{7192} hairpin, presenting the characteristic sawtooth shape (orange) [82], which allows to discard that any segments of the DNA molecules are hidden by an incorrect tether of the molecule (described in Sec. 5.2.1). In yellow, the relaxation curve after the oligonucleotide hybridizes to the loop region is shown, corresponding to the elastic curve for the ssDNA molecule. The ssDNA FEC differs from the ideal elastic curve (represented in red using the WLC model) at low forces, where it appears a plateau that has been previously associated with the presence of secondary structure [61, 84].

5.2.1 *Extraction of the molecular extension: the blocking-loop oligonucleotide method*

For long hairpins ($N \geq 100$bp), ssDNA is obtained using the blocking-loop oligonucleotide method as described in [61]. Initially, a 25-30 bases oligonucleotide complementary to the loop region of the hairpin is flowed into the central channel where the experiments are carried out. After mechanically unzipping the hairpin, this oligonucleotide binds to the hairpin-loop region, preventing the re-zipping of the molecule, as shown in Fig. 5.1A. Oligonucleotide-free buffer is flowed into the channel to remove the excess of oligonucleotide, which has been shown to have an impact to the elastic response of the ssDNA molecule [61]. The effects of the dsDNA segment produced by the hybridization of the blocking-loop oligo should be negligible for lengths over 200 bases. Indeed, we verify that for a molecule, H_{204}, the elastic response is not affected by the presence of the oligonucleotide (See App. E).

FECs show two distinct force regimes separated by the unzipping force transition (≈ 15pN). A high-force regime ($\gtrsim 15$pN) that can be well described with elastic polymer models and a low-force regime ($\lesssim 15$pN) presenting a force plateau characteristic of the formation of secondary structure [61]. The fact that ssDNA does not behave as an ideal elastic polymer in the low-force regime is well known, as it tends to form non-specific secondary structures at forces below the unzipping transition. In contrast, for hairpins shorter than 100 basepairs, the secondary structure plateau disappears, allowing us to extract the ideal elastic response at low forces.

To characterize the ssDNA elasticity, several pulling and relaxing cycles are recorded, by moving the optical trap at constant speed ($\approx$

50nm/s) and measuring the force as a function of the molecular extension. The measured distance, λ (corresponding to the relative distance from the center of the trap to the tip of the micropipette), is related to the ssDNA extension $x_{\mathrm{ssDNA}}(f)$ as :

$$x_{\mathrm{ssDNA}}(f) = \lambda(f) - x_h(f) - x_b(f) - x_0, \qquad (5.1)$$

where $x(f)$ is the molecular extension of the stretched ssDNA, $x_h(f)$ is the extension of the handles, $x_b(f)$ is the displacement of the bead from the center of the optical trap, and x_0 is a shift of the trap position relative to the position detector. The extension of the handles, x_h, are modelled as an extensible WLC with the parameters given in [81] and the displacement of the bead, x_b, is computed by assuming a Hookean spring behaviour and using the trap stiffness determined by the force power spectrum of an untethered bead at zero force (as described in Sec. 2.2.5). Typical values in our experiments for k_b range from 0.06pN/nm to 0.09pN/nm.
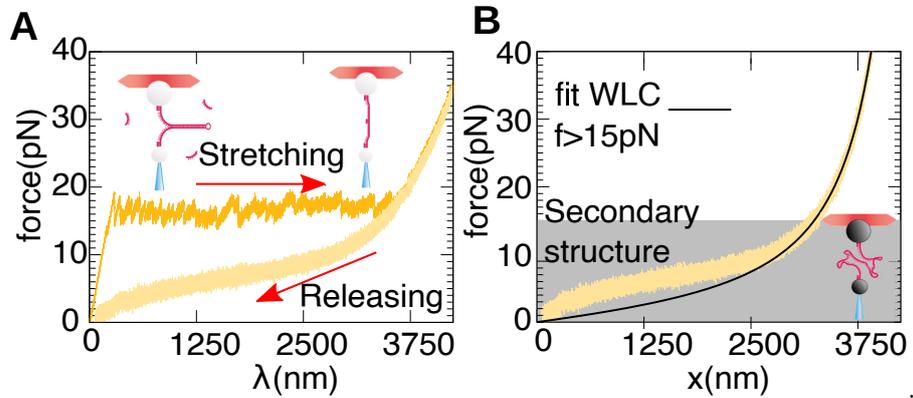
Figure 5.1: **Obtaining ssDNA from long hairpins: blocking-loop oligo method.** (a) Schematic depiction of the blocking-loop oligo technique. The hairpin is mechanically unzipped (dark yellow) by stretching a tethered molecule (in this case, H_{7192}), moving the trap position away from bead in the pipette. After the molecule is completely open, the blocking-loop oligo, diluted with a concentration of ~ 10nM in the buffer, binds to its fully complementary region, which contains the loop region, preventing the formation of the native hairpin upon realeasing the force. (b) Typical FEC of the ssDNA molecule obtained from H_{7192}. Above a force of $f \approx 10$pN, the FEC can be well reproduced and fitted to the WLC model, as shown in black. Below this force, however, the molecule collapses, by forming unespecific secondary structure, deviating from the ideal behavior characterized by the WLC model.

To determine the offset of the distance, x_0, requires the following steps:

- Check that the full molecule is being stretched, i.e. no regions are hidden by any missalignment in the tether. The missalignment happens when the tethered molecule in the bead of the micropipette is on the side of the bead (even in the micropipette), effectively reducing the measured extension of the tehtered molecule. As shown in Fig. 5.2, the experimentally measured unzipping pattern

is compared to a theoretically generated one, which is obtained
using the methodology explained in [5, 82]. The unzipping shown
in Fig. 5.2 is complete, i.e. all the inital rips of the theoretical
curve are reproduced in the experimental unzipping.

- If the experimental pattern matches the one given by the theoret-
ical prediction, the offset, x_0 can be obtained: it is given by the
measrued distance at zero force, $x_0 = \lambda(f = 0\text{pN})$.
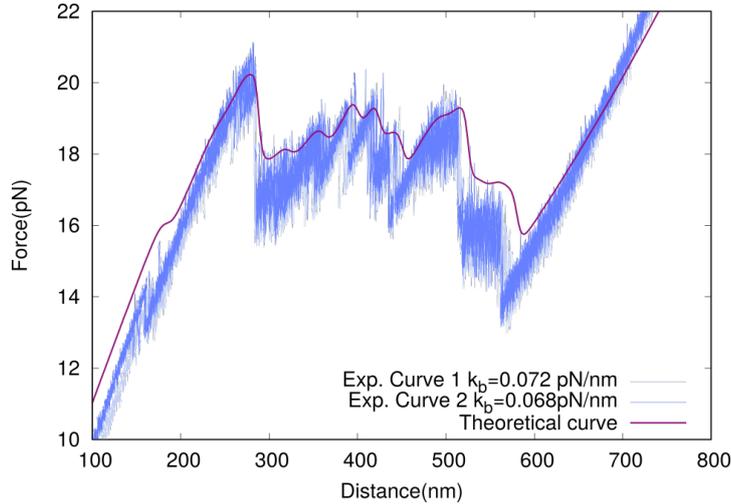
Figure 5.2: Two experimental force-distance unzipping curves (each one with
a different measured trap stiffness k_b) compared to a theoretical one. In both
cases, the agreement between the theoretical rips and the experimental ones
guarantee the right molecule is unzipped. A special attention is paid on the
beginning of the FEC, since if there is an incorrect molecular tethering, initial
rips may be missing. For obtaining the theoretical curve, the elastic parameters
of ssDNA are used from [61], with a bead stiffness of $k_b \approx 0.07\text{pN/nm}$.

A typical force-extension curve (FEC), obtained from plotting the
measured force vs the ssDNA molecular extension, x_{ssDNA} (Eq. 5.1) for
a long ssDNA molecule, $f(x_{\text{ssDNA}})$, is shown in Fig.5.1B.

5.2.2 Force-extension curves

The observed ssDNA elastic curves, such as the one shown in Figs. 5.1B
and 5.3A-B, present two clear regimes. At forces higher than 15pN,
the molecules follow the behaviour predicted by polymer models such
as the WLC model. However, at forces below ~ 10pN we observe a
force plateau that has been previously associated to the formation of
secondary structure [61, 76].

Different hairpins of various lengths were studied using the method
as described above. Representative pulling curves for 6 molecules, com-
prising lengths from 700 and ~ 14000 bases, are shown in Fig. 5.3A. A
collapse of all curves is observed after re-scaling the molecular exten-
sion by the total number of nucleotides (Fig. 5.3B). Both the elastic
response at high forces and the secondary structure formation observed
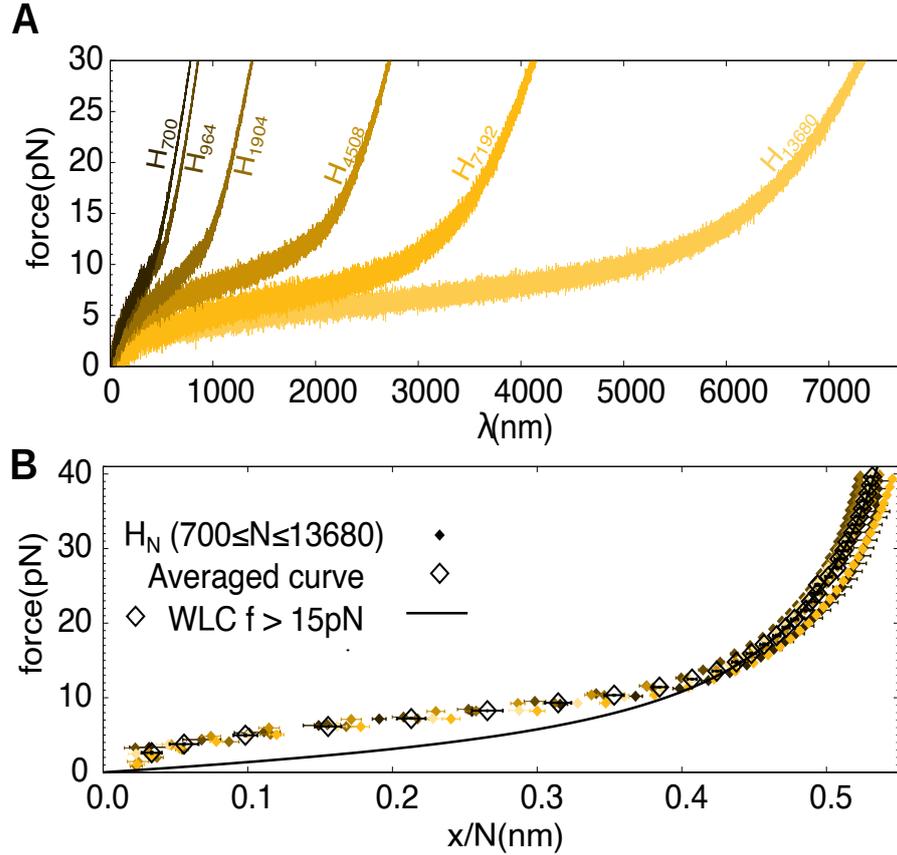
Figure 5.3: **Elastic behavior for long molecules.** (a) Force-distance curves(FDCs) for the 6 molecules studied, ranging lengths from 700 to 13680 bases, from dark to bright yellow. (b) Re-scaled FECs over the number of bases of each sequence for the molecules shown above, keeping the color code. An averaged curve for all the molecules is shown as a black empty diamond. The error bars correspond to the standard error of the mean of different tethers (with a minimum of 4 for H_{13680} and a maximum of 11 for H_{1904}).

at lower forces are independent on the specific sequence and length of the polymer. This result suggests that in this range of polymer lengths interacting forces are local in space. Both the elastic response and secondary structure formation are independent of the specific sequence and length of the polymer.

5.3   ELASTIC PROPERTIES OF SHORT DNA MOLECULES

5.3.1   *Extraction of the molecular extension: 2-branches method*

For short DNA hairpins ($N \leq 100b$), the binding of the oligonucleotide to prevent the re-zipping of the molecule might affect significantly the measured elastic properties. Consequently, we develop an alternative approach. We use a large loop in order to induce a large hysteresis between pulling and relaxing cycles [85], Fig. 5.4A. Then, as it is derived in App. E, the subtraction of the trap position between the folding and

unfolding force-extension branches at a given force f, $\lambda_F - \lambda_U$, gives the extension of the ssDNA, x_{ssDNA}, as follows:

$$x_{\text{ssDNA}}(f) = \lambda_F(f) - \lambda_U(f) + x_d(f) + \lambda_0(t), \qquad (5.2)$$

where $x_d(f)$ is the diameter of the hairpin projected along the stretching direction (as described in E.1) and $\lambda_0(t)$ is the correction of the drift using a spline interpolation after the alignment of the cycles (explained in App. E). The FEC obtained using this method for a 20bp hairpin is shown in Fig. 5.4B. The range of forces in which the FEC can be computed is limited by the unfolding ($\approx$ 15pN) and the refolding ($\approx$ 4pN for a 20b-loop) forces.
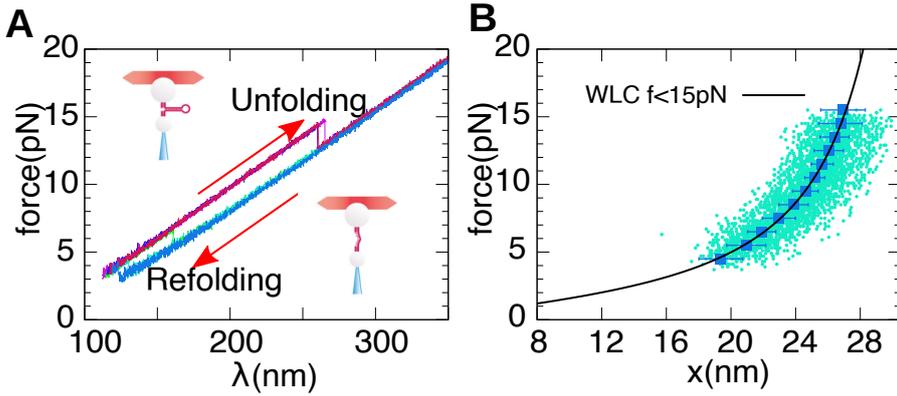
Figure 5.4: **Obtaining ssDNA from short hairpins.** (a) FDC of 5 unfolding and refolding cycles using $H_{60}A_{19}$. (b) FEC obtained from applying Eq. 5.2 to $\sim$ 100cycles of $H_{60}A_{19}$.

5.3.2   *Blocking-splint oligonucleotide technique*

In order to measure the FEC at larger forces, we use a trick to increase the unfolding force. The hairpin is synthesized with ssDNA handles (instead of dsDNA handles) and the experiments are performed with a blocking-splint oligonucleotide (for sake of simplicity, will be called blocking-splint oligo from now on) that is fully complementary to one flanking ssDNA handle (29b) and partially (n bases) to the other ssDNA handle (see App. E), as shown schematically in Fig. 5.5. A spacer of 4 Thymines (about 2nm in length) is inserted between the 29b and nb positions to properly hybridize the oligonucleotides to the flanking ssDNA handles and to stabilize the folded state of the hairpin up to sufficiently high forces $f < f_l$. The particular value of f_l depends on the total length n of the blocking-splint oligo. In our case, a blocking-splint oligowith $n = 15$ gives $f_l \approx 40$pN, high enough to extract $x_{\text{ssDNA}}(f)$ over a wide range of forces. With this new configuration, the extension in the two branches is related with the ssDNA extension by:

$$x_{\text{ssDNA}}(f) + x_{h2}^{n,\text{ssDNA}}(f) = \lambda_F(f) - \lambda_U(f) + x_d(f) -$$
$$- x_{h2}^{n,\text{dsDNA}}(f) + \lambda_0(t), \qquad (5.3)$$

where $x_{h2}^{n,\mathrm{dsDNA}}(f)$ is the extension of n bp dsDNA the hairpin at force f and $x_{h2}^{n,\mathrm{ssDNA}}(f)$ is the extension of the n bases of ssDNA (Fig. 5.5). The derivation of Eq. 5.3 is found in App.E. In all the experiments shown in this chapter, $n = 15$. Notice that the total number of bases in ssDNA form in the l.h.s of Eq. 5.3 is higher than that due to the unzipping of the hairpin. Particularly, when using the blocking-splint oligo method in $H_{60}A_0$, the elasticity is obtained for a total number of bases of 75. For this reason, throughout this chapter, $H_{75}A_0$ will be describing the molecule $H_{60}A_0$ when using the blocking-splint oligo method.
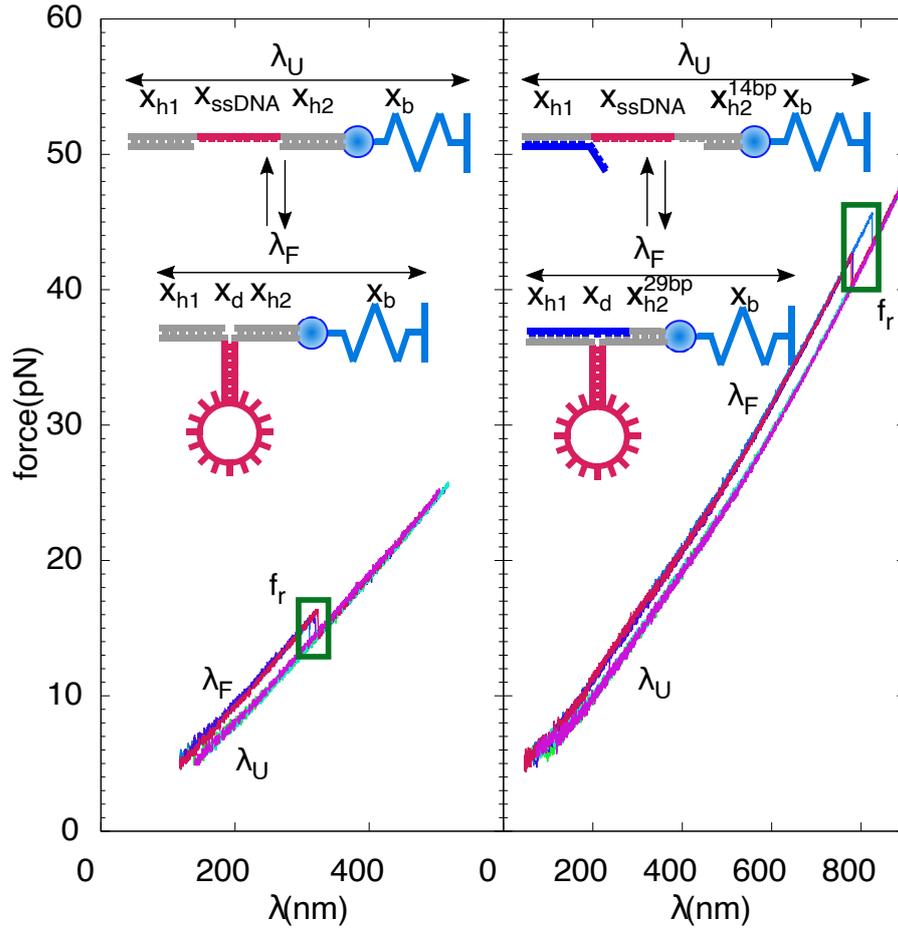
Figure 5.5: **Obtaining ssDNA from short hairpins.** Force-distance curves without (left) and with blocking-splint oligo. A scheme on top of the curves show a schematic representation in the distances involved in Eq. 5.2 and Eq. 5.3. Green squares contain the forces at which the molecule breaks open, f_r.

Even if the length of the hybridized bases, n, in the handle is smaller than the length of the hairpin, the presence of the blocking-splint oligo increases the unfolding force by more than a factor of two. This is a consequence of the force being applied in different geometries: whereas in the hairpin the force is applied perpendicular to its double-helix stem (stretch), in the handles the force is applied parallel to the DNA helix.
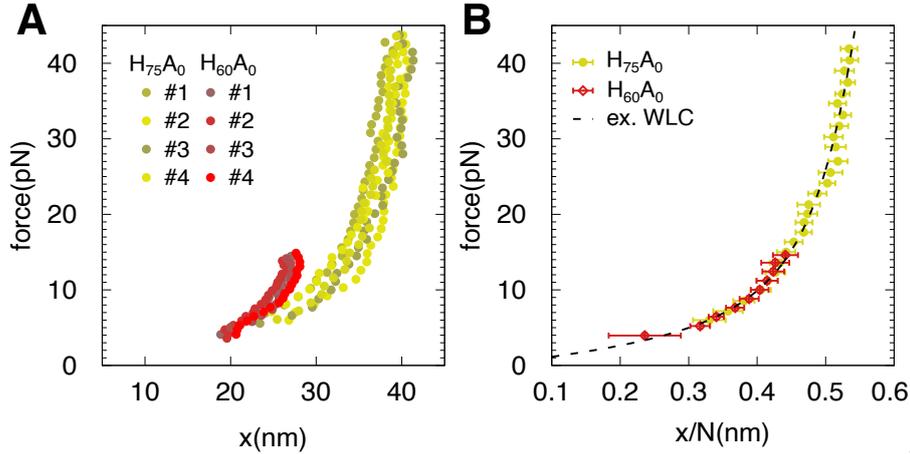
Figure 5.6: **ssDNA extension using the blocking-splint oligo method.**
(a) FEC for several tethers of the $H_{60}A_0$, using the 2-branches method (2BM, red), and of $H_{75}A_0$, using the blocking-splint oligo method (BSOM, yellow). (b) Re-scaled FEC of the $H_{60}A_0$ using the 2BM and $H_{75}A_0$ using the BSOM. The FECs collapse at the force range they share ($4 < f < 15$pN).

When using $H_{75}A_0$ with the blocking-splint oligo technique (schematically depicted in Fig. 5.5), the unfolding force increases to ≈ 40pN, while the refolding one remains at ≈ 4pN, as it is shown in Fig. 5.5. The obtained FEC is in perfect agreement with the one shown for $H_{60}A_0$ (without the blocking-splint oligo) in the range of forces $f \in [4, 15]$ pN, as shown in Fig. 5.6.

5.3.3 *Force-extension curves*

The short hairpins (60 bases) are designed with a large loop of 20 bases to use the two branches method (Sec. 5.3.1 and 5.3.2). The pulling and relaxing cycles present a large hysteresis, allowing us to estimate the ssDNA extension while the folded and unfolded branches coexist (Fig. 5.4A-B). The elasticity measured using this method is restricted to the lower range of forces ($\approx 4 - 15$) pN, where the folded and unfolded branches are simultaneously observed. The re-scaled FEC where the extension per nucleotide is represented are shown in Fig. 5.7A (top) for the two shorter hairpins studied. Interestingly, the secondary structure plateau disappears for these hairpins ($H_{60}A_{19}$, and $H_{60}A_0$,). Moreover, we are able to detect differences between the two FEC, which we interpret to be due to base stacking. $H_{60}A_{19}$,is shorter and stiffer than $H_{60}A_0$, despite having the same sequence in the stem.

Note that when using the blocking-splint oligo method, the FEC at large forces (> 15pN) tends to collapse with the FEC of the average curve for long molecules (averaging the results obtained for the 6 molecules ranging from 700b to ~ 14kb) above 15pN, where there is no secondary structure, as it is also shown in Fig. 5.7A (bottom) .

5.4   FITTING FEC TO ELASTIC MODELS

The Worm-Like Chain model is used to fit the FECs of short and long molecules within different ranges of forces due to experimental limitations. In the case of long molecules, the range fitted is between 15 and 40pN, since below 15pN secondary structure appears and the WLC model fails to reproduce it. For the shortest molecules, the range of forces is mainly limited by the highest unfolding and lowest refolding forces where the unfolding and refolding branches are kinetically stable, which is between 4 and 15pN for our pulling speed. The faster the pulling speed, the larger the hysteresis and the range of forces.

The parameters of the fits of the WLC model (Eq. 4.13) are represented in Fig. 5.7B for long and short molecules. The elastic parameters for the collapsed curve for long hairpins are shown in a dashed line and are consistent with those obtained for each long molecule studied.
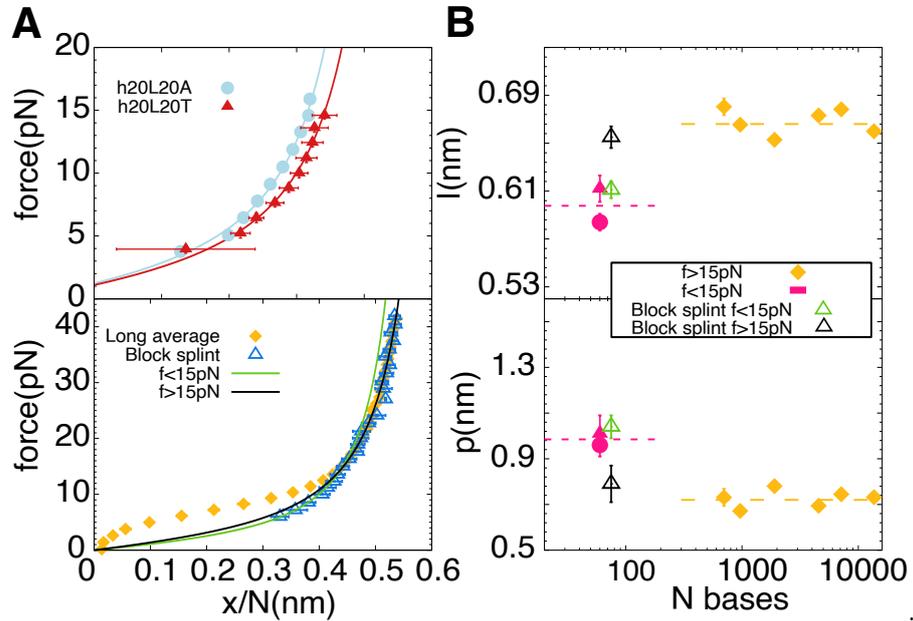
Figure 5.7: **Comparison of the elasticity of short and long molecules.** (a) Top: Re-scaled FEC of the two short hairpins studied, with their respective fits using the WLC model (Eq. 4.13). Bottom: Re-scaled FEC of $H_{75}A_0$, (blue) obtained from the blocking-splint oligo technique and the average FEC obtained for all long molecules (the 6 molecules ranging from 700b to $\sim$ 14kb) (yellow). The black and green lines are the fits of the WLC model on the $H_{75}A_0$, FEC, fitting above and below $f = 15$pN. (b) Contour length (top) and persistence length (bottom) obtained from fitting the different molecules (yellow for long ones, magenta for short ones). The black (green) triangles show the values of the parameters obtained from the fit of the $H_{75}A_0$, FEC shown in panel A bottom, for forces above (below) 15pN.

A systematic difference is observed between short and long molecules, following the trend previously observed [65]. Using the $H_{75}A_0$ with blocking-splint oligo data, we are able to get the FEC of a short molecule in a whole range of forces ($4 < f < 40$pN). First, we notice that the inextensible WLC model(Eq. 4.13) fails to fit all the $H_{75}A_0$ FEC with

just one set of elastic parameters. We fit the $H_{75}A_0$ data in the two force regimes: low ($f < 15$pN, green) and high ($f > 15$pN, black) as shown in Fig. 5.7A, bottom. The elastic parameters obtained from fitting these force regimes, high force (HF) and low force (LF), coincide with the parameters of the fits obtained for short and long molecules (Fig. 5.7B), which were fitted at different force regimes. This proves that long and short molecules do not present different elasticity. The difference in the elastic parameters are rather an artifact coming from the different force regime fitted. However, using the extensible Worm-like chain described by Eq. 4.15, a single curve describes both short and long molecules, with the same elastic parameters: $l = 0.618(14)$nm, $p = 0.95(4)$nm, $Y = 1100(200)$pN, as shown in Fig. 5.8A. Therefore, we conclude that the difference in l (contour length per base) and in p (persistence length) in the WLC fit (Eq. 4.13) is not related to the molecular length but arises from the distinct force regimes fitted ($f < 15$pN and $f > 15$pN).

5.4.1 *Stacking effects*

Polypurine DNA regions have been reported to show stacking effects [72]. In order to study the effects of stacking in the ssDNA elasticity we studied two hairpins with the same stem sequence but different loop content. In Fig. 5.7A (top) we show that the behaviour for polydA- and polydT-loops differs, being $H_{60}A_0$ stiffer and shorter than $H_{60}A_0$. This is expected, since homopolymeric-purine regions, specifically, polydA, have higher stiffness[86], due to the tendency of purines to stack [72]. This fact is also shown in the values of the fitting parameters (Fig. 5.7B), where stiffer values (lower contour length per base, l, and higher persistence length, p) are obtained for the polydA-loop hairpin.

A cooperative unstacking transition is not observed for the $H_{75}A_0$ (measured with the blocking-splint oligo) throughout the whole range of forces (up to 40pN). This is already expected because this sequence is lacking of large purine regions which are the ones that tend to stack [70, 72]. In chapter 6 we explore in detail the effect of stacking on ssDNA elasticity.

5.5   ENERGETICS OF ssDNA ELASTICITY

Different elastic models (Chapter 4) have been used to describe the elasticity of DNA. The elasticity of dsDNA has been usually reproduced by the WLC model [11, 24, 51], whereas ssDNA has been characterized by the FJC [24, 82] as well as the WLC [63], with similar results in the fitness of the fits reproducing the FECs for long ssDNA molecules ($N \sim 10^4$ bases) at forces $f > 15$pN [61].

Despite the convergence of these models at high forces, the two show different elastic behaviours at a low range of forces ($f < 15$pN). As above discussed, the elastic parameters obtained from fitting the WLC to the data of h20L020T at low forces (LF-WLC) and at high forces (HF-WLC) are different. As it is shown in Fig. 5.8A, the difference between the FEC

obtained from using the results of the fits above (black, HF-WLC) and below(green, LF-WLC) 15pN is not negligible. This systematic deviation is expected to become more crytical when considering the free energy of stretching.

The free energy of stretching ssDNA is a quantity that needs to be computed in many experiments to extract information (such as free energy of formation). This free energy is calculated as $\Delta G = \int_{x_0}^{x} f(x)\,dx$, where $f(x)$ is the ideal elastic FEC of the ssDNA (i.e. without any secondary structure formation) and x_0 and x are the initial and final ssDNA extensions. Therefore, the estimation of the ssDNA stretching free energy, and so the base-pair energies extracted, depend on the model (Fig. 5.8B) and the elastic parameters (Fig 5.7B) used to describe the ideal elastic response of ssDNA, as shown in App. E.
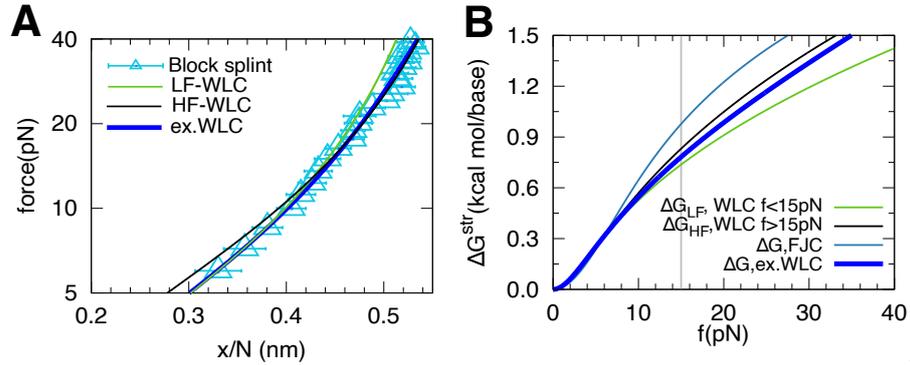
Figure 5.8: **Effect of the range of forces employed in determining the elasticity of ssDNA.** (a) Experimental FEC of the $H_{75}A_0$ molecule (blue triangles), with the fit using the extensible WLC model (dashed lines), which gives $l = 0.618(14)$nm, $p = 0.95(4)$nm, $Y = 1100(200)$pN. The green and black lines show the FECs obtained from the LF-WLC and HF-WLC fits, respectively. All models and force-extension formulas are described in Chapter 4. (b) Free energy of stretching obtained for each one of the models shown in panel A (keeping the color code), adding the FJC (blue, with parameters given in [82]).

In unzipping experiments carried out at room temperature the force value at which DNA unzips is around 15pN (its exact value depending on the salt [82, 83] buffer and the DNA sequence). The estimation of the stretching contribution requires integrating the elastic FEC of ssDNA along the low-force regime from 0 to 15pN. We show in Figure 5.8B that using the elastic parameters describing the high-force regime we get for the stretching free-energy contribution at the unzipping force (15pN), $\Delta G_{HF} = 0.83$kcal/mol·base, while the value obtained from using the extensible WLC , gives $\Delta G_{LF} = 0.78$kcal/mol·base. This difference of $\approx 6\%$ gives a systematic error for the nearest neighbour energies in DNA of about 0.05kcal/mol·base (See App. E), comparable with the current statistical error in the NNBP energymeasurements [82, 83]. This error in the stretching energy leads to an error in the folding free energy of an Nbp DNA structure of $\delta \Delta G = 2N \cdot 0.05$kcal/mol $= N \cdot 0.1$kcal/mol, which, for heterogeneous sequences with $\approx 50\%$GC content, we have $\Delta G_{bp} \approx 1.5$kcal/mol·bp which results in a $\approx 6\%$

ΔG error. The extensible WLC describes the ssDNA elasticity over a large range of forces, leading to a refined precision in the estimation of DNA free energies of formation. The error increases when calculating the stretching free energy of ssDNA using other models, such as the extensible FJC (fitted for $f > 15$pN, as in [82]), as shown in Fig. 5.8B.

5.6 SUGAR PUCKER TRANSITION IN ssDNA

DNA is composed of a phosphate backbone connected to a nitrogenate base via a deoxyribose sugar. The five carbon atoms in the deoxyribose ring are non-planar, as has been described in Chapter 3. The two most commonly observed puckers in crystal structures of isolated nucleosides and nucleotides are the south (C2'-endo) and north (C3'-endo). As shown in Fig. 5.9, depending on the puckering, i.e. how the atoms are non-planarly distributed, the distance between adjacent bases is affected. For the case of the north conformation, the interphosphate distance is 0.59nm, while for the south conformation is 0.7nm [34]. The constraint in the distance is correlated with the structure that dsDNA adopts: while bases in B-DNA form present a south conformation, in A-DNA the north is adopted.

The extensibility required in the elastic model in order to reproduce the FECs of ssDNA has puzzled researchers since it first was proposed as a way to fit the experimental curves [24]. While extensibility has been used in order to describe the elasticity of dsDNA, some physical interpretation has been given to it. The extension along the axis of the double helix in the B-DNA form is shorter than the distance between adjacent monomers: while the first has been measured to be of 0.34nm/bp, the second is 0.70nm, corresponding to the south sugar pucker conformation. This allows for some degree of extensibility in the double helix, by stretching the stacking of bases further away than the 0.34nm/bp above mentioned.

Nevertheless, for the case of ssDNA, this is not the case. The rigidity of the phosphodiester bonds, in the absence of any other secondary structure (either via base pairing or stacked bases), is restricted to 0.59 (0.70)nm, for the north (south) sugar pucker conformation. And still, the FJC or WLC fails to sucessfully reproduce its elastic behaviour, needing to add an extensibility. The values for this extensibility has been found to be around $S \approx 700$pN [24, 61].

As we have just seen, ssDNA elasticity requires of extensible models in order to reproduce the experimental curves throughout all its range of forces. Since the earliest measurements of ssDNA elasticity [24], the extensibility was pointed out to hide a transition happening to the ssDNA (although no configurational explanation was proposed).

Here, we propose a minimal 2-state model, based on the assumption that the measured extensibility is due to a transition in puckering experienced by the ssDNA chain. The model successfully reproduces the obtained FECs, pointing out the transition in puckering as a feasible explanation to the observed chain extensibility.
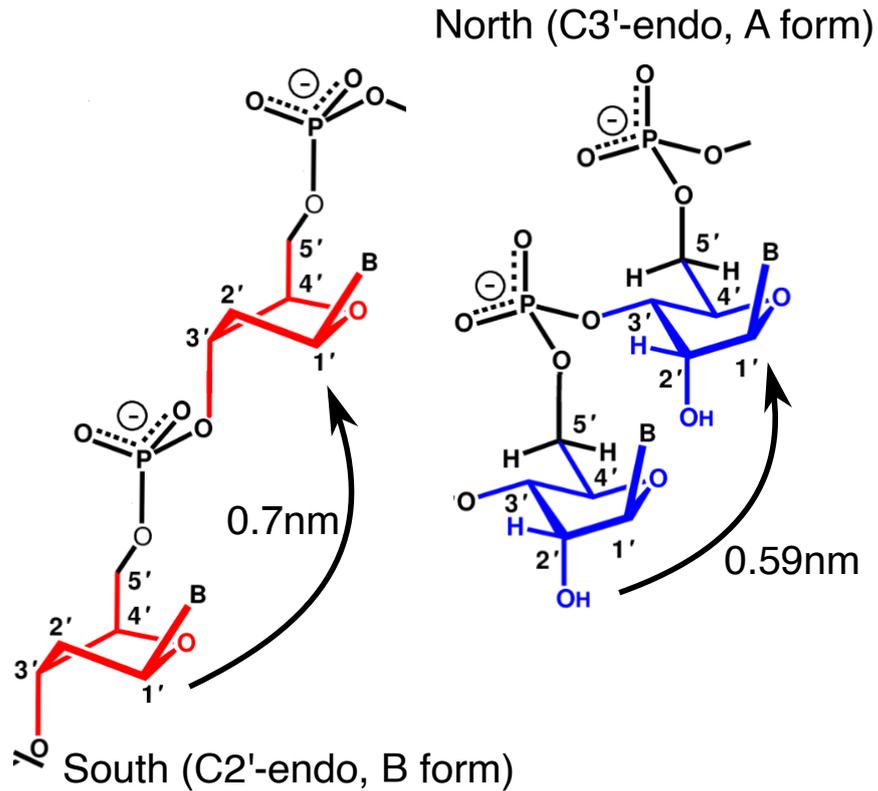
Figure 5.9: **North and south puckering**. The distance between adjacent bases depends on the deoxyribose ring puckering. Adapted from [87].

As discussed before, different DNA sequences, lengths and techniques give different values for the elastic parameters that best fit the ssDNA elasticity. The discrepancy in the parameters obtained between high and low force regimes can be solved by using extensible models, such as the extensible WLC used in Chapter 5 to reproduce the elasticity of ssDNA throughout all the measured range of forces ($5 < f < 45$pN). The underlying question, however, is why would an extensible model be needed for the ssDNA case, if all bonds that compose its backbone are of covalent nature, and hence, rigid? A hypothesis that we explore here is that the extensibility required in reproducing the ssDNA elasticity arises from a transition between the north and south puckering of the deoxyribose ring.

5.6.1   A Minimal 2-state model

The ssDNA molecule is modeled as a chain of N monomers in a thermal bath of temperature T, that can be either in the north ($\sigma_i = -1$) or south ($\sigma_i = 1$) configuration. We assume that each configuration has an elastic behaviour that is described by $x_N(f)$ or $x_S(f)$, respectively. Due to the structural constraints of each configuration, $x_N < x_S$, $\forall f \gg 0$. Under an external constant applied force, f, the Hamiltonian for such a system is,

$$\mathcal{H}(\sigma_i, f) = \left[J - \int_0^f \Delta x(f') df' \right] \sum_i \sigma_i - N \int_0^f \overline{x}(f') df', \qquad (5.4)$$

where J is the energy gap per base between the north and south configurations, $\overline{x} = (x_S(f) + x_N(f))/2$ and $\Delta x(f) = (x_S(f) - x_N(f))$. This system is equivalent to a non-interacting, 1D Ising model. The free-energy of the system is easily obtained:

$$G(f, T) = -N k_\text{B} T \log \left[2 \cosh \left(\frac{\int_0^f \Delta x(f') df' - J}{2 k_\text{B} T} \right) \right] - N \int_0^f \overline{x}(f') df'. \qquad (5.5)$$

From the free energy, the average extension is obtained by using $x(f) = -\partial G / \partial f$:

$$x(f) = N \left[\overline{x} + \frac{1}{2} \tanh \left(\frac{\int_0^f \Delta x(f') df' - J}{2 k_B T} \right) \right]. \qquad (5.6)$$

This model assumes that there is no interaction between adjacent monomers. At zero force, the elastic terms vanish. From the Hamiltonian of Eq. 5.4, the free energy difference per base between having all the monomers in the south conformation ($\sigma_i = 1, \forall i$) with respect to having all bases in the north conformation ($\sigma_i = -1, \forall i$) can be calculated. Hence, the free-energy difference per base between south-north sugar pucker configuration is

$$\Delta G_{pucker} = 2J, \qquad (5.7)$$

while the total change in energy of the chain by changing all its bases from north to south is given by $\Delta G_T = N2J$.

5.6.2 *Fitting the model to experimental FECs*

In order to test the hypothesis and the goodness of the model described in Sec. 5.6.1, we fit the re-scaled FEC of the $H_{75}A_0$ molecule (first introduced in Chapter 5). The elasticity of each one of the states, south and north, is modelled as an inextensible WLC, with a fixed contour length given by its crystallographic one, 0.70nm and 0.59nm, respectively [34]. To the best of our knowledge, no numbers characterizing a chain of bases with the same pucker state have been published[1] Therefore, we have no avaliable information regarding the values of their persistence lengths, which is required to obtain a force-extension relation of each puckering chain type. Nevertheless, we expect the values for both chains to be close to the generally accepted values for the persistence length of ssDNA, $\sim$ 1nm [65].

Therefore, the experimental FEC is fitted using only three parameters:

---

1 The studies published regarding the puckering of DNA only consider an isolated deoxyribose ring and not a whole chain [88, 89].

- J, which is related to the free energy difference of the south and north states (Eq. 5.7).

- p_S and p_N, which are the persistence lengths of each south and north chain type. The persistence length contains information of the bending rigidity of each of the chains
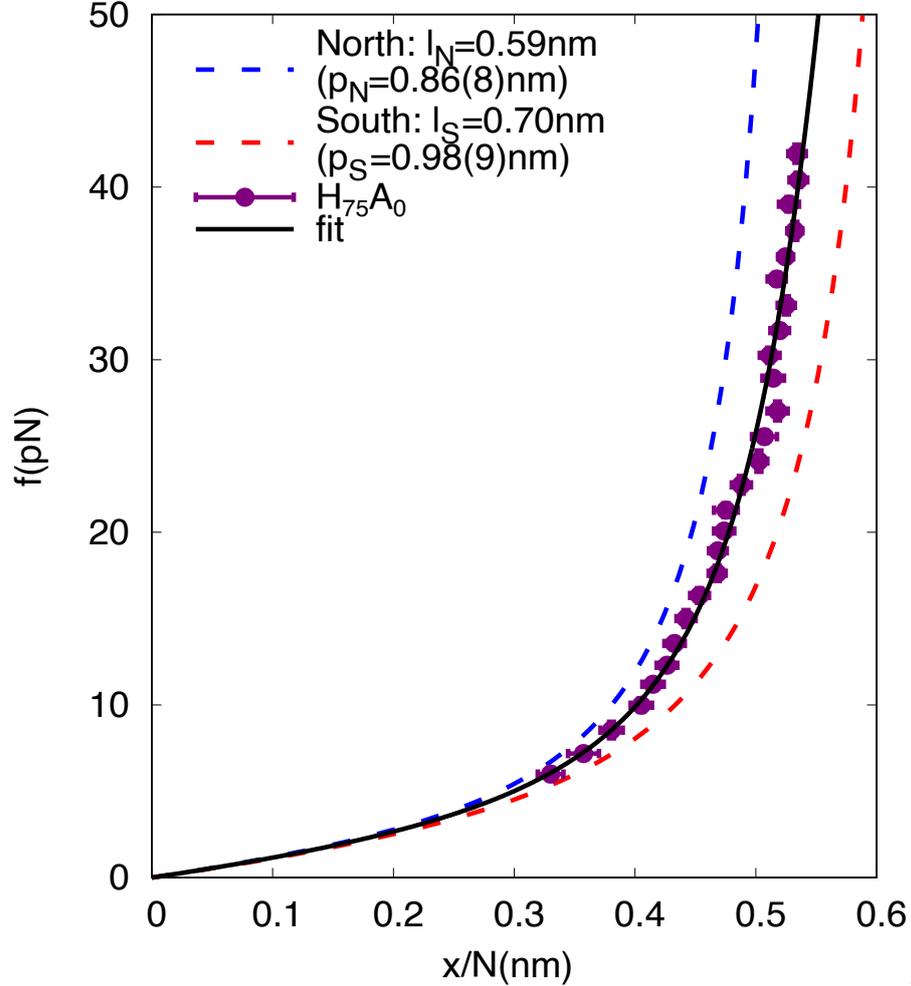
Figure 5.10: **Fit of the re-scaled FEC of $H_{75}A_0$ at 10mM MgCl$_2$**. (a) The experimental points are shown in purple (error bars are the standard error of 7 averaged tethered molecules, each one with at least 30 cycles). The discontinuous lines represent the ideal ssDNA curve of the north (red) and south (blue) configurations, modelled as a WLC. The fit of the curve is the black line, with the parameters: $J = 1.05(12)$pNnm, $p_S = 0.98(9)$nm, $p_N = 0.86(8)$nm.

As can be seen in Fig. 5.10, the fit succefully reproduces the re-scaled experimental curve of the $H_{75}A_0$ molecule.

Regarding the elaticity of the two pucker configurations, the results of the fit return a persistence length of $p_N = 0.86(8)$nm, for the north configuration and $p_S = 0.98(9)$nm, for the south one. The two values obtained are really close to each other (and compatible within errors), which indicates that the orientation correlation in the chain is basically

independent on how close the adjacent bases are. The elastic curves corresponding to these elastic parameters are shown in blue (north) and red (south) in Fig. 5.10.

The experimental FEC is contained in between the two dashed curves, corresponding to each of the conformations. This is one indirect constraint of the model, since from Eq. 5.6, it can be seen that, at a given force f', only extensions between $\overline{x} \pm \Delta x/2 = x_S(f), x_N(f)$ can be obtained from it.

Regarding the free-energy difference between each of the puckerings, the obtained value for J is 1.05(12)pNnm. Hence, the model suggests that the north configuration is favoured, and then gets destabilized by the externally applied force, favouring the longer state one (south). From Eq. 5.5, we can relate J to the free-energy difference per base between the north and south configurations, which gives $\Delta G_{pucker} = 2J = 2.10\text{pNnm} = 0.30(3)\text{kcal/mol}$. This is in contrast with the obtained values using computational techniques, -0.34kcal/mol [88] and -0.32kcal/mol[89], which found that the south configuration is favoured in deoxyadenosines, with respect to the north one. However, the sequence we study has a mixture of all bases, and hence the north puckering may be favoured energetically. Indeed, the free energy differences given for other nucleotides varies between -0.46 to 0.75kcal/mol [89].

Further work is needed in order to test wether the apparent extensibility observed in ssDNA elasticity can be explained with the proposed sugar pucker transition. In particular, experiments with different sequences, each one with different abundance of each nucleotide, might allow to check the free energy difference of puckering of each base, and compare them with the reported values in the literature.

Varying other thermodynamic parameters, such as the temperature or the salt concentration, would help to understand the underlying nature of this transition.

5.7 CONCLUSIONS

In this chapter we have studied the elastic response of ssDNA molecules of different lengths ranging from tens to tens of thousands of bases with optical tweezers. Pulling on ssDNA molecules that fold into DNA hairpins at low forces allows us to have full control over the sequence and length of the molecule under study. Hence, from fitting the measured FECs to elastic models we can directly extract the elastic parameters (persistence length p and contour length per base l) for the different ssDNA sequences. This is in contrast to previous assays where full length control could not be achieved [72]. For molecules with random sequences ($\sim 50\%$GC content) of lengths $\gtrsim 500$ bases we find that the FECs collapse when representing the force as a function of the ratio between the molecular extension and the contour length: at large forces (> 15pN) they follow the elasticity modelled as a WLC, whereas at

lower forces ($<$ 10pN) a force plateau appears, which is regarded as a signature of secondary structure formation.

A more detailed study regarding the secondary structure formation and its length dependence can be found in Chapter 6.

As discussed in the previous section, the elastic parameters describing the different ssDNA molecules studied vary with their contour length, the shorter molecules showing a stiffer behaviour determined by a larger persistence length and lower interphosphate distance, a trend previously reported in [65]. However, rather than the length or the specific sequence of the molecule, this difference arises from the range of forces where the inextensible WLC model is fitted. Indeed, the emergence of secondary structure for long molecules ($\gtrsim$ 500bases) at forces $\lesssim$ 15pN restricts the fitting of the FEC to the high force regime ($\gtrsim$ 15pN), while for short hairpins the fitting region extends over the low range of forces ($\lesssim$ 15pN).

This dependence of the persistence length with force goes in the same direction than [79] where it was argued that at low stretching forces ($f < 2$pN) interactions become relevant, leading to systematic lower elastic strain and higher persistence length. However the higher persistence length we observe occur at higher forces ($5 < f < 15$pN) than those expected in [79]. The investigation of the whole force range with the same molecule (using the blocking-splint oligo method to explore the two force regimes) shows that the dependence of the elastic parameters of the molecule studied is an artifact of the different force regime in which the data is fitted. This discards other possibilities that could explain this discrepancy, such as the stiffening of ssDNA in the vicinity of dsDNA regions (proposed in [90]). The inextensible WLC model is unable to reproduce both force regimes ($f \lesssim$ 15pN, $f \gtrsim$ 15pN) with only two elastic parameters. However, using the extensible WLC, we have also studied whether base stacking could affect the FEC of short ssDNA $\lesssim$100b. We have detected stacking effects in hairpins $H_{75}A_{19}$ and $H_{75}A_0$ which sequences differ in the 19 bases of the loop, $\approx 30\%$ of the total number of bases. The measured FECs show a difference in total extension between the two sequences of ≈ 2nm with respect to a full length of ≈ 40nm (5% difference). This difference is in agreement with [72], where the polydA molecule is $\approx 20\%$ shorter at 10pN than a polypyrymidine sequence. The stacking in ssDNA is explored in detail in Chapter 6.

The need of using an extensible WLC to succesfully reproduce the elasticity of ssDNA can be explained in terms of a transition in the puckering of the ssDNA. Indeed, by using a very simple 2-state model, we can reproduce the experimental FEC. Furthermore, we obtain the free energy difference between the 2 conformations, with $\Delta G_{pucker} = 0.30(3)$kcal/mol, favouring the north conformation.

STACKING IN ssDNA

Stacking plays a critical role in the formation and stability of dsDNA, as we have explained in Chapter 3.

Calculations of the free energy of stacking have been obtained using bulk techniques [91, 92] as well as single-molecule techniques techniques: unzipping of DNA molecules [82], stretching different configurations of dsDNA [71] or using origami DNA to stack the bases and unstack them by an externally applied force [93].

However, stacking has been also observed in ssDNA, using a large variety of techniques: calorimetry measurements [94], X-ray diffraction [95] or single-molecule experiments, fluorescence, AFM and magnetic tweezers [71, 72, 96]. These studies showed it has been generally concluded that purines have a higher tendency to stack, which was expected [33]. Indeed, when stretching ssNA sequences consisting of consecutive adenines, the obtained FECs show a force-plateau that has been associated to a stacking-unstacking (S-U) transition, as was shown for the first time for ssRNA [97]. Nevertheless, initial studies have been mainly focused in sequences containing consecutives purines (adenines) or pyrimidines (thymines). Later experiments studied sequences with alternating purines-pyrimidines and CG content, proving the phenomenon to be more complex than initially thought [94]. Remarkably, the thermal stability of the stacked conformation is dependent on the presence of a possible nucleation based on at least 4 bases (quadruplets) in the ssDNA sequence and not on the GC content of the ssDNA sequence, nor on the number of purine bases in the sequence.

In this chapter we study the elasticity of ssDNA molecules presenting various degrees of stacking, via the blocking splint oligo method (BSOM) explained in Chapter 5. Also, a two state model is developed in order to explain the experimental results and predict the elastic properties of the unstacked state and the stacked one. Finally, we perform experiments by varying two decades the salt concentration which allow us to characterize the salt dependence of the free energy of stacking, as well as the elastic properties of the unstacked state and the stacked one.

6.1 THE 2S-WLC MODEL

Here we have developed a model in order to reproduce the experimental FECs containing a stacking-unstacking (S-U) transition. The ssDNA molecule is modelled as a chain of N monomers in a thermal bath of temperature T is under tension by an external, controlled, force, f. σ_i stands for the state (equivalently to the spin in the Ising formalism) of each one of the monomers. The two states considered are stacked ($\sigma_i = 1$) or unstacked ssDNA ($\sigma_i = -1$) with a free energy difference at zero force between the two states of $2J_i$. As has been previously shown [72, 94],

not all ssDNA sequences experience the S-U transition. We incorporate this heterogeneous behaviour by considering two different types of bases, with different $J_i > 0$: ones that stack (usually purines, $J_i = J$), and ones that don't (usually pyrimidines, $J_i = J'$)[1]. Let $n_s = \sum_i^N \frac{1+\sigma_i}{2}$ and $n_u =' \sum_i^N \frac{1-\sigma_i}{2}$ be the number of monomers in the stacked and unstacked state, respectively. A general Hamiltonian for such a chain is

$$\mathcal{H}_\mathcal{T} = -\sum_i J_i\sigma_i - n_s \int_0^f x_s(f')df' - n_u \int_0^f x_u(f')df' - \gamma \sum_i^N \sigma_i\sigma_{i+1}.$$
(6.1)

J_i can take two different positive values, depending on whether the ith base is allowed ($J_i = J$) or not ($J_i = J'$) to be stacked. The stacking formation is heavily penalized for bases with $J_i = J'$, thereby imposing $J' \gg J > 0$.

The second and third terms take into account the stretching energy contribution of all the monomers in stacked conformation (with an extension described as $x_s(f)$) and unstacked conformation ($x_u(f)$). The last term is a nearest-neighbour interaction term that gives the energy penalty(reward) of having neighbouring monomers in different(equal) states. For the sake of simplicity, we assume that the coupling term ($\gamma > 0$) is the same for both types of bases, regardless of wether they are of the stacking tyoe (J_i) or the non-stacking type (J'_i).

The model includes a heterogeneity parameter, p, to account for molecules with different degrees of stacking, induced by varying sequence and salt dependence. Assuming there is a fraction $1-p$ of the monomers which are penalized towards showing stacking ($J' \to \infty$), and a fraction p that can experience the S-U transition (finite J), which are uncoupled one from another (their distribution along the monomeric chain is basically separated and hence their interactions may be neglected), the Hamiltonian of Eq. 6.1 can be splitted in two terms:

$$\mathcal{H}_\mathcal{T} = \mathcal{H}' + \mathcal{H},$$
(6.2)

$\mathcal{H}$ corresponds to the hamiltonian for the bases experiencing S-U transition whereas $\mathcal{H}'$ describes the monomers which do not show any transition. Using $\Delta x = x_u - x_s$ and $\overline{x}(f) = (x_u(f) + x_s(f))/2$, the Hamiltonians of Eq. 6.2 read as:

$$\mathcal{H}' = -J' \sum_{i=pN}^N \sigma_i - \gamma \sum_{i=pN}^N \sigma_i\sigma_{i+1} - \int_0^f \Delta x(f')df' \sum_{i=pN}^N \sigma_i - \int_0^f \overline{x}(f')df' \sum_{i=pN}^N \sigma_i,$$
(6.3)

$$\mathcal{H} = -J \sum_{i=1}^{pN} \sigma_i - \gamma \sum_{i=1}^{pN} \sigma_i\sigma_{i+1} - \int_0^f \Delta x(f')df' \sum_{i=1}^{pN} \sigma_i - \int_0^f \overline{x}(f')df' \sum_{i=1}^{pN} \sigma_i.$$
(6.4)

---

1 Calorimetry experiments [94] have shown that single-stranded DNA sequences that present stacking are more complex to define than just them having consecutive purine or pyrimidine bases.

Where we have assumed that the system is self-averaging, as well as neglected the constant terms. The summation terms of equations 6.3 and 6.4 can be written as

$$\sum_{i=pN}^{N} = (1-p)\sum_{i=1}^{N}, \tag{6.5}$$

$$\sum_{i=1}^{pN} = p\sum_{i=1}^{N} \tag{6.6}$$

In order to force the fraction $1-p$ of the bases to remain unstacked ($\sigma_i = -1$) and not to experience the S-U transition, we take the limit $J' \to -\infty$. Hence, the Hamiltonian of Eq. 6.3 is reduced to ($\sigma_i = 1, \forall i$):

$$\mathcal{H}' = (1-p)N \int_0^f x_1(f')df' + \kappa \tag{6.7}$$

where κ, is a constant term which will be neglected onwards. Substituting Eqs. 6.7 and 6.4 to Eq. 6.2, the final Hamiltonian reads,

$$\mathcal{H}_{\mathcal{T}} = (1-p)\sum_1^N \int_0^f x_u(f')df' + p\Big(J\sum_{i=1}^N \sigma_i - \gamma \sum_{i=1}^N \sigma_i\sigma_{i+1} -$$
$$- \int_0^f \Delta x(f')df' \sum_{i=1}^N \sigma_i - \int_0^f \overline{x}(f')df' \sum_{i=1}^N \sigma_i \Big). \tag{6.8}$$

The normalized extension, per base, of such a system can be written as (See Appendix C):

$$x(f)/N = (1-p)\cdot x_1(f) + p\left[\overline{x}(f) + \frac{\Delta x(f)}{2} \frac{\sinh\left(\alpha(f)\right)}{\sqrt{\sinh^2\left(\alpha(f)\right) + \exp^{-4\gamma'}}} \right]. \tag{6.9}$$

Where we have defined $\alpha(f) = (\int_0^f \Delta x(f')df' - J)/2k_BT$ and $\gamma' = \gamma/k_BT$. Another way to write Eq. 6.9, is by writing it as a function of the fraction of bases in the unstacked and stacked conformation, $\phi_u = n_u/N$ and $\phi_u = n_u/N$:

$$x(f)/N = \phi_u x_u(f) + \phi_s x_s(f) \tag{6.10}$$

with the explicit formulae for each fraction of bases,

$$\phi_s = \frac{p}{2}\left(1 - \frac{\sinh\left(\alpha(f)\right)}{\sqrt{\sinh^2\left(\alpha(f)\right) + \exp^{-4\gamma'}}} \right), \tag{6.11}$$

$$\phi_u = (1-p) + \frac{p}{2}\left(1 + \frac{\sinh\left(\alpha(f)\right)}{\sqrt{\sinh^2\left(\alpha(f)\right) + \exp^{-4\gamma'}}} \right). \tag{6.12}$$

A schematic depiction of the model is found in Fig. 6.3A. Starting at zero force, a fraction of the bases start in the stacked state (dark blue, shorter extension, x_{-1}). As the force increases, the unstacked state (yellow), with a longer extension, x_1, is energetically favoured, forcing the bases to unstack. Finally at a high enough force, all bases are in the unstacked state. Notice there is a fraction of bases that are always in the unstacked state (yellow), since they do not experience the S-U transition (the stacking is heavily penalized) .

6.2 SEQUENCE DEPENDENCE OF THE STACKING TRANSITION

Stacking of DNA bases depend strongly on the sequence (as discussed in Chapter 3). For DNA, purines have been shown to exhibit a larger stacking effect than pyrimidines [72, 96]. Here we study the elasticity of ssDNA with different sequences, designed to present different degree of stacking. The ssDNA is generated by unzipping different DNA hairpins, and using the BSOM described in Chapter 5. The four different molecules (Fig. 6.1) present a stretch of consecutive n adenines, which are expected to present stacking. The molecules are named as $H_N A_n$, where N indicates the total number of ssDNA bases present when the molecule is unzipped, A stands for Adenine and n is the total number of consecutive adenines in the loop. Three out of the four molecules ($H_{75}A_0$, $H_{75}A_{15}$ and $H_{85}A_{39}$) are hairpins with a varying stem (20bp, 20bp and 15bp) and loop length (20, 20 and 40 bases)[2]. The last of the molecules, $H_{55}A_{40}$, is a stem-less structure, with a 40 bases long loop, which connects the two double-stranded DNA handles which are used to pull the molecules.

The expected difference in the degree of stacking arises from the sequences in the loop, and its relative weight over the total molecular length. Assuming that only consecutive adenines (present in the loop) will exhibit stacking, the fraction of bases that can stack can be calculated, $p_{seq} = n/N$. The proportions given by p_{seq} for $H_{75}A_0$, $H_{75}A_{15}$, $H_{85}A_{39}$ and $H_{55}A_{40}$ are, respectively, 0, 0.25, 0.46 and 0.73.

The experiments of the four types of molecules were performed at $T = 25°C$, in a buffer containing $10mM$ of $MgCl_2$, 10mM Tris pH 7.5 and $0.01\%NaN_3$. By performing pulling and relaxing cycles (with the BSOM) and using the two branches method, we obtained the force as a function of the ssDNA extension for each molecule.

Typical FECs are shown in Fig. 6.2B for three molecules presenting some stacking. Several cycles are represented to show the typical dispersion in the measured extension (cloud of points).The difference in the average extension for each molecule is due to the different number of bases composing them. In order to compare them, the average curve for each molecule (obtained by averaging at least 4 different tethers) is normalized over the total number of bases of the molecule. As shown in Fig. 6.2C, all normalized FECs collapse at high forces, above ~ 25pN.

---

2 The elasticity of $H_{75}A_0$ molecule has been previously discussed in Chapter 5 (named h20L20).
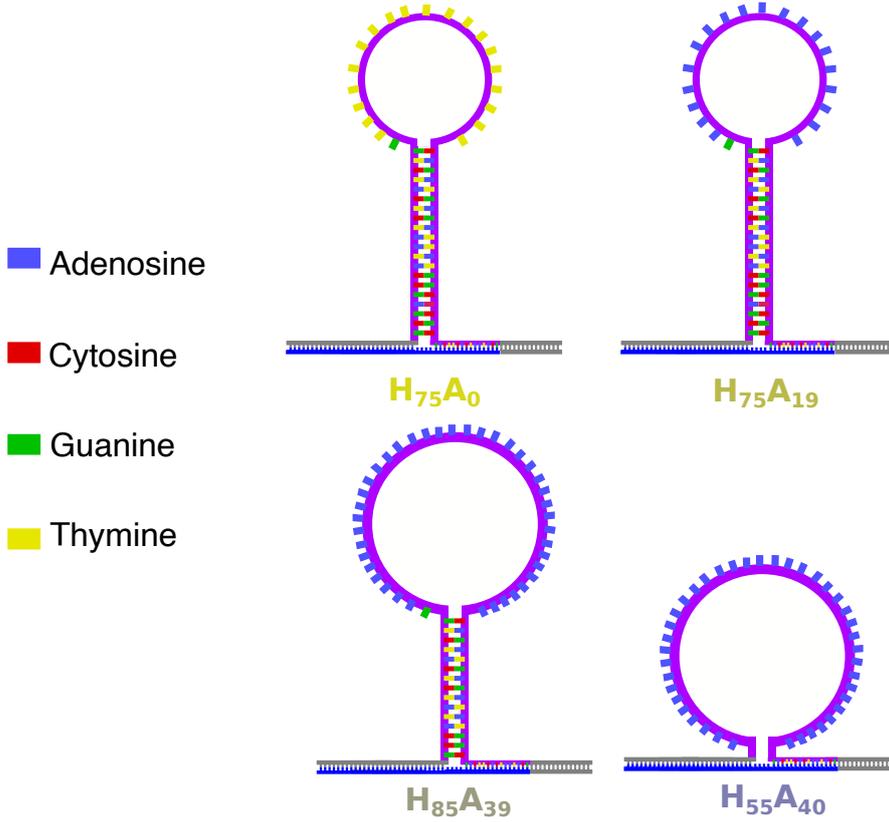
Figure 6.1: **Hairpins with various degrees of stacking.** Scheme of the 4 hairpins studied, with a stem of 20, 20, 15 and 0 basepairs and a loop of 20, 20, 40 and 40 bases. The handles are shown with the molecular construct given by the blocking splint oligo method, described in Chapter 5. The bases represented follow the color code shown in the left.

This is a proof that at high forces, all the molecules show the same elastic behaviour, which corresponds to the ideal (non-stacked configuration). Remarkably, the deviation of the curves with respect to the unstacked elasticity (dashed line in Fig. 6.2C) is larger for larger p_{seq} of the sequence.

In order to fit the model to the FECs, the elasticity of the unstacked state was assumed to be described as an extensible WLC with the parameters derived in Chapter 5 ($l = 0.62$nm, $p = 0.95$nm and $Y = 1100$pN), as shown as a discontinuous line. This curve describes well the elasticity of the $H_{75}A_0$ molecule ($p = 0$), as previously shown in Chapter 5. The colored continuous lines are the obtained fits of the model described in Sec. 6.1.

We use the model described in Sec. 6.1 to fit all experimental results. In order to do it, we have assumed that the parameters of the model related to the free energy of stacking, J, and the cooperativity term, γ, are the same for all the molecules. For the elastic parameters of the stacked state, we use an inextensible WLC (Chapter 3) with a contour length per base, L_0, and the persistence length, P_0, for all sequences. The only parameter that was allowed to vary between the sequences is
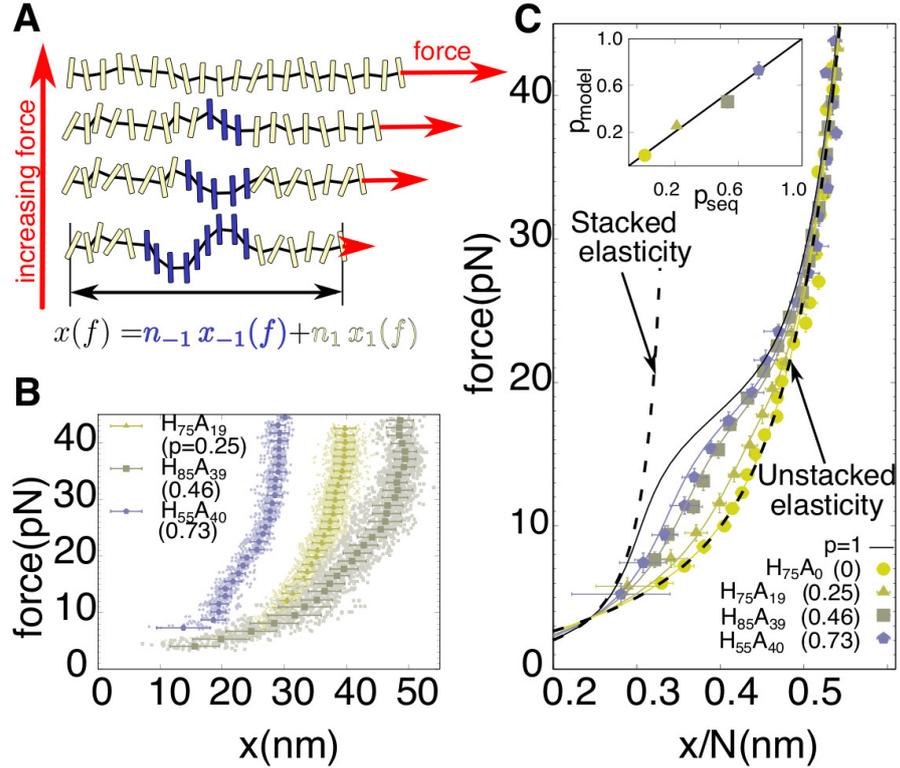
Figure 6.2: **Unstacking transition for ssDNA**. (a) Schematic depiction of the model, the total extension of the molecule is the sum of two contributions: the bases in ideal ssDNA state, and the stacked bases. (b) FECs for the molecules that show some degree of stacking. $\sim$30 trajectories of the same tethered molecule are shown for each one. The points show the extension obtained by the blocking splint method, while the bigger points are the averaged and standard deviation of all of them. (c) Collapsed FEC for the 4 molecules studied and fit of the model (Eq. 6.9). A higher degree of stacking is observed as a higher force plateau. The inset shows the proportion of bases that show stacking with respect the total number of bases of each sequence, p. p_{model} is obtained from the fitted value of the model, while p_{seq} considers that only consecutive dA bases present in the loop contribute to stacking. The uncertainties of all parameters obtained were determined by bootstrapping the combined data of the 4 sequences studied. Each experimental FEC is obtained by averaging from 4 to 10 tethers.

the degree of stacking of the molecule, p, which describes the proportion of bases that experience the S-U transition. The combined fit of the different FECs has 8 parameters: γ, J, P_0, L_0, and p for each of the 4 sequences studied.

From the combined fit we obtain $J = 0.089(3)$kcal/mol and $\gamma = 0.92(4)$kcal/mol. The $\sim$ 10-fold difference between J and γ show how important is the cooperativity, i.e. the interaction between adjacent bases that enhance the formation of the stacked structures. Indeed, without coupling between adjacent bases ($\gamma = 0$), the model fails to reproduce the FECs.

Regarding the elastic properties of the stacked state, we obtain $L_0 = 0.41(3)$nm and $P_0 = 3.4(4)$nm. The obtained contour length is similar to

the one observed per base for B-DNA (~ 0.34nm), which suggests that the stacked ssDNA elasticity takes a similar geometry than the strands on the double helix conformation. On the other hand, the persistence length is also larger than that obtained for ssDNA at similar salt conditions, ~ 0.75nm ([61]), which is expected since the rigidity of the stacked bases is larger than the unstacked ones. Besides, this persistence length is much smaller than the one obtained for dsDNA, ~ 50nm ([24, 80]). Again, this is expected due to the increased rigidity conferred by the double helix.

The free-energy difference of base-stacking can be computed: $\Delta G^0_{stack} = 2J$. The obtained value from substituting J from the fit gives $\Delta G^0_{stack} = 0.180(6)$kcal/mol, which is in agreement with the free-energy differences measured in [72], $\Delta G^0_{stack} = 0.16$kcal/mol.

6.3 SALT DEPENDENCE OF THE S-U TRANSITION

In order to investigate how the stacking depends on the salt conditions we perform experiments varying the NaCl concentration over 2 decades and using two different sequences showing different degree of stacking: $H_{75}A_0$ ($p \equiv 0$) and $H_{85}A_{39}$ ($p \equiv 0.46$).

6.3.1 Elasticity of the unstacked ssDNA as a function of the salt concentration

In order to explore the energetics of base stacking at different salt conditions, a good characterization of the elasticity of the unstacked behaviour as a function of the salt concentration is required. To do so, experiments using the BSOM in the $H_{75}A_0$ molecule ($p = 0$) have been performed(Fig. 6.1B). This technique (described in Chapter 5) allows us to obtained precise FECs for short molecules (< 100bases) over a wide range of forces. The stability of the hybridized dsDNA handles depends on the salt concentration. Therefore, the range of forces which can be used to extract the molecular extension (described in Chapter 5), depends also on the salt concentration used. The values obtained vary from $0 < f < 25$pN, for the lowest salt concentration studied ([NaCl] = 10mM), to $4 < f < 45$pN for the highest one([NaCl] = 1000mM).

Finally, the values obtained for the degree of stacking, p, coincide perfectly with the percentages of contiguous adenines in the sequence (Fig. 6.2C, inset). This result further validates the proposed model.

The obtained FECs for $H_{75}A_0$ ($p = 0$) at different salt concentrations ([NaCl] of 10mM, 50mM, 100mM 500mM and 1000mM) are shown in Fig. 6.3. We assume, like in Sec. 6.2, that this sequence does do not present any stacking effects, e.g. $p = 0$. The FECs of the $H_{75}A_0$ molecule, have been all fitted using the extensible WLC model (See Chapter 4), assuming for the persistence length a salt dependence proposed in [61]:

$$P([Ion]) = p_0 + A \cdot [Ion]^{-1/2}, \qquad (6.13)$$
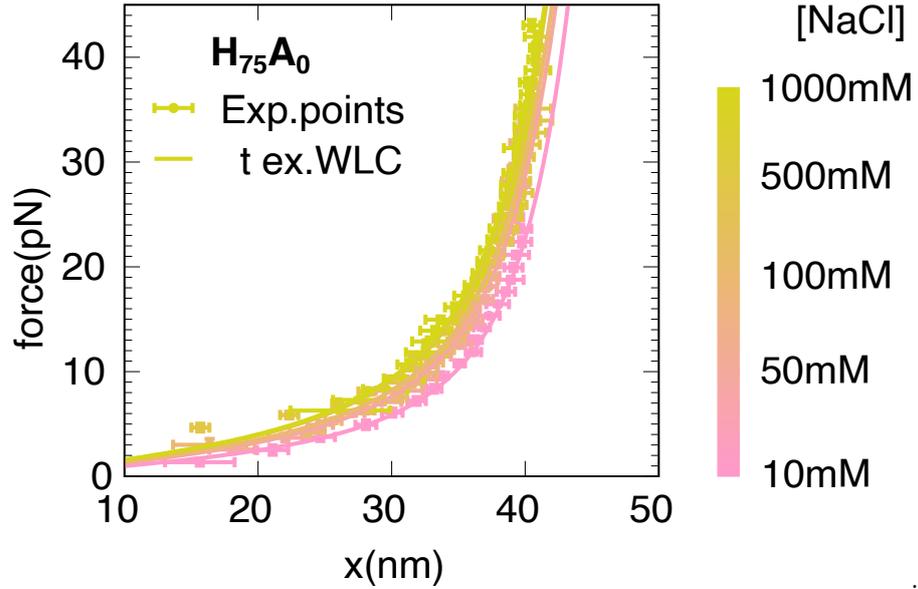
Figure 6.3: **Salt dependence of the FECs for the ssDNA obtained for the $H_{75}A_0$ molecule**. The experimental data is shown in colored squares, at different salt concentrations. Color code indicated in the right legend. The fits are done by fitting the extensible WLC expression, described in Chapter 4.

where $[Ion]$ is the ionic concentration in mM, p_0 corresponds to the persistence length obtained for $[Ion] \rightarrow \infty$ and A is a coefficient that accounts for the dependence of the persistence length with $[Ion]$.

The combined fit leads to $p_0 = 0.87(7)$nm, $A = 1.97(18)$nm, and $l = 0.643(6)$nm, for the contour length per base and $Y = 1900(300)$, regarding the Young modulus. The results are quite compatible with those given in [61]: $p_0 = 0.68 \pm 0.04$nm and $A = 2.19 \pm 0.10$nm. The difference in p_0 might come from two factors: 1) the above mentioned study uses the inextensible WLC model (while we use the extensible one), 2) the fits to the FECs are performed in a different range of forces $15 \leq f \leq 40$pN (while we fit in a larger range of $\sim 5 - 45$pN).

The results at 1M NaCl are in agreement with the ones obtained at $[MgCl_2] = 10$mM in Chapter 5, for the 60b0dA molecule: $p = 0.95(4)$nm. This is expected, since the electrostatic effects can be described using the $100/1$ phenomenological rule (described in Chapter 5) which states that the non-specific binding affinity of a given concentration of divalent cations is roughly equal to that of 100-fold times large concentration of monovalent cations[98, 99], leading to $[MgCl_2] = 10$mM equivalent to $[NaCl] = 1$M. This rule has been experimentally tested using optical tweezers for RNA in [100] and DNA in [83].

6.3.2   *Elasticity of ssDNA presenting stacking as a function of the salt concentration*

The same pulling experiments varying the salt concentration in the buffer are performed for the $H_{85}A_{39}$ ($p = 46$) molecule (Fig. 6.4). Data for

the lowest salt concentration ([NaCl] = 10mM) could not be obtained, since after the unfolding of the hairpin, the molecule does not fold again, probably due to the presence of the 40 bases of the loop. At increasing
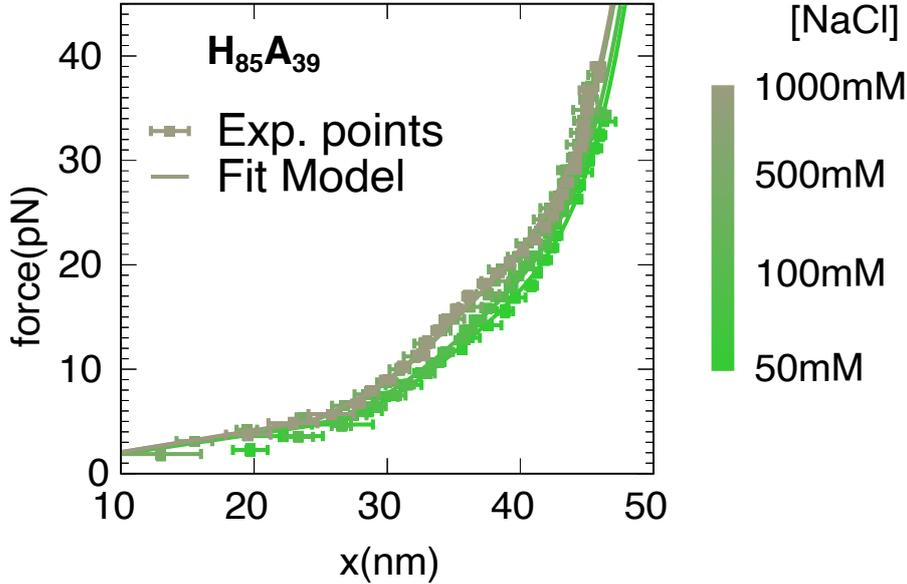
Figure 6.4: **Salt dependence of the FECs for the ssDNA obtained for the $H_{85}A_{39}$ molecule**. The experimental data is shown in colored squares, as indicated in the right legend. The fits are done by fitting the 2S-WLC model expression of Eq.6.9.

salt concentration, the molecular extension at a high force ($\sim$ 30pN) becomes shorter, as observed for the unstacked sequence (Fig. 6.3), showing smaller persistence length for increasing salt dependence, as expressed from (Eq. 6.13).

Remarkably, as shown in Fig. 6.5, the S-U plateau is larger for the highest salt concentration. The pulling experiments are performed slow enough for the FEC to be reversible, hence the area between the partially stacked sequence ($H_{85}A_{39}$) and the unstacked curve ($H_{75}A_0$) corresponds to the free-energy difference of the stacked and unstacked conformations of Np bases, $Np\Delta G^0_{stack}$. This area increases with salt, which indicates that ssDNA stacking is stabilized by salt concentration, contrary to what has been previously reported [72].

In order to fit the curves, we have followed a similar approach to the one performed in Sec. 6.2. In this case, we have imposed that all molecules have the same degree of stacking, which has been taken from the fits performed in Sec. 6.2, $p = 0.52(6)$. The values for the cooperativity, γ, and the contour length per base of the stacked state, L_0, have been assumed to be the same for all salt concentrations. Therefore, the only varying parameters with salt concentration are the persistence length, P_0 and the difference of energy between unstacked and stacked states, $2J$. The latter was assumed to follow a relation such as Eq. 6.13, and hence the salt dependence of the elasticity of the stacked ssDNA is reflected in two' parameters: A and p_0.
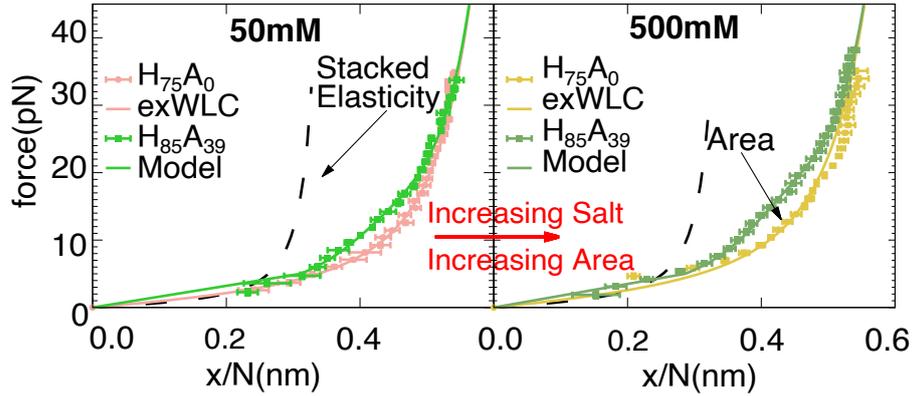
Figure 6.5: **Salt dependence of the FECs for the ssDNA obtained for the $H_{85}A_{39}$ molecule**. The experimental data is shown in colored squares, as indicated in the right legend. The fits are done by fitting the 2S-WLC model expression of Eq.6.9.

The results for the combined fits give $\gamma 0.97(11)$kcal/mol, $L_0 = 0.45(4)$nm and $P_0 = 1.2(3) + 4.9(7)[Ion]^{-1/2}$(nm), with the uncertainties given by bootstraping all the experimental points of all curves. Interestingly, both the cooperativity and the contour length are in agreement with the values obtained in Sec. 6.2, $\gamma 0.92(4)$kcal/mol and $L_0 = 0.41(3)$nm. The main discrepancy found is regarding the obtained persistence length, which, if applying the $100/1$ rule, appears to be smaller than the equivalent at 10mM $MgCl_2$. Indeed, the value obtained for the persistence length in this combined fit, for 1M of NaCl, is $P_{0,} = 1.4(4)$nm, while the one obtained for 10mM$MgCl_2$ is a factor of two higher ($P_0 = 3.4(4)$nm). This might be explained as follows. On one hand, that the $100/1$ rule does not apply in a strongly structured chain such as a stacked chain. On the other hand, the values for the contour and persistence length are highly anti-correlated. This translates to having a different set of parameters giving similar extension at a given force: a large contour length with a short persistence length gives a similar molecular extension than a short contour length with a large persistence length. This leads to errors in the elastic parameters, specially if the range of forces in which the fit is performed is small.

Regarding the comparison of the elasticity of the partially stacked sequence with the unstacked sequence, we observe the same trend as the one observed at Sec. 6.2. The stacked sequences present a larger persistence length than the unstacked ones, which is easily interpreted in terms of rigid, stacked bases forming the chain. Moreover, the contour length is in agreement with a base-to-base distance corresponding to a helix-like conformation (such as a B-DNA form), as observed in previous analysis (Sec. 6.2)

Finally, the free-energy difference of the stacked-unstacked state, $\Delta G^0_{stack} = 2J$, increases with the salt concentration. In the next section we investigate in detail the free energy of stacking using different approaches.

6.3.3 *Free-energy of stacking*

We use two methods for obtaining the free energy of stacking at zero force, ΔG^0_{S-U}. The first one consists in obtaining the free energy difference from fitting the model to the experimental data (Sec. 6.1). The value obtained from the J parameter is directly related to the free energy of stacking ΔG^0_{S-U}. Indeed, if we compute $\Delta G^0_{S-U} = G_U(f = 0) - G_S(f = 0) = \mathcal{H}_U(f = 0) - \mathcal{H}_S(f = 0)$, where we have used that the entropy of the totally stacked ($G_S(f = 0)$) and totally unstacked ($G_U(f = 0)$) configurations is zero.

Using Eq. 6.8, one can easily see that each base that stacks contributes with $2J$ to the free energy difference between the stacked and unstacked states, ΔG^0_{stack}:

$$\Delta G^0_{stack} = 2J. \tag{6.14}$$

On the other hand, the free energy change between the stacked and unstacked conformations of ssDNA sequences with a p degree of stacking reads:

$$\Delta G^0_{S-U} = 2J(1 - p)N. \tag{6.15}$$

Another approach is to directly use the normalized FECs and measure the area between the normalized FEC of a given ssDNA sequence with p degree of stacking and that of the normalized FEC corresponding to the totally unstacked conformation ($p = 0$). This area can be related to ΔG^0_{stack} as follows:

$$\Delta G^0_{stack} = \frac{Area}{1 - p}. \tag{6.16}$$

This results have implicit the approximation that $2J \approx Area/(1 - p)$. This approximation is valid since at zero force, which is the starting value of the force in the integration, the fraction of bases in the stacked state obtained from the values of the model is almost 1, for all salt concentrations studied.

The agreement between the two approaches to calculate the free energy of stacking, Eqs. 6.16 and 6.14, is shown in Fig. 6.6B. In order to compare the free energy of stacking of a single base ΔG^0_{stack} with that of dsDNA we plot in Fig. 6.6B ΔG_{bp} as $2\Delta G^0_{stack}$. The values obtained for ΔG_{bp} seem to be smaller than free energy of basepair stacking [93, 101] and nearest neighbour hybridization [82] which varies between ~ 0.8 kcal/mol to ~ 3 kcal/mol, depending on the sequence. This can be explained by a higher stabilization of the dsDNA stacking given by the complementary strand, perhaps due to the enhanced proximity of the bases in the dsDNA conformation and the formation od hydrogen bonds between the two strands.

A salt dependence relationship of the stacking free energy of dsDNA is proposed in [91], which considers that the free energy depends on the salt concentration through:

$$\Delta G_{bp} = m \cdot \log([\text{NaCl}]) + \Delta G_0 \tag{6.17}$$
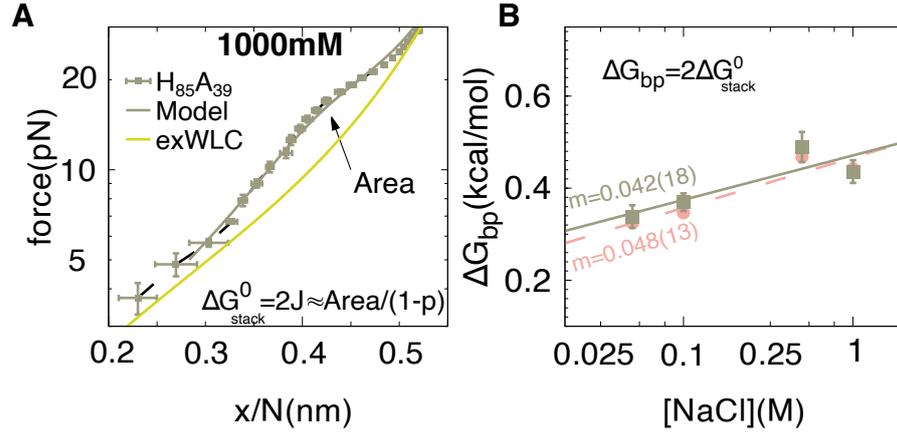
Figure 6.6: **Salt dependence of the FECs for the ssDNA obtained for the $H_{85}A_{39}$ molecule**. (a) FEC of $H_{85}A_{39}$ molecule at 1000mM NaCl concentration (squares, brown). The solid brown curve is the obtained fit using Eq. 6.9, while the discontinuous black curve is a linear interpolation of the experimental points. In yellow the elasticity of the unstacked state at 1M NaCl (Sec. 6.3.1) is shown. (b) Stacking free energy per basepair as a function of the salt concentration. In brown it is shown the obtained values of the fits, while in pink we show the values from the area between the experimental curve and the unstacked state. There is good agreement between both.

with [NaCl] in M, hence ΔG_0 correspond to the free-energy of stacking per basepair at 1MNaCl.

By fitting the data for ΔG_{bp}, we obtain $m \sim 0.04 - 0.05 \text{kcal/mol}$, which is smaller than that of dsDNA (e.g in Mfold [91, 102]), $m = 0.110 \text{kcal/mol}$. To the best of our knowledge, this is the first time that the salt dependence of the energy stabilization due to base stacking in ssDNA is reported. Previous studies using force-spectroscopy techniques [72] failed to prove its salt dependence, despite the strong evidence for it in the case of dsDNA.

This may be due to the fact that the effect for ssDNA is very small and coupled to the totally unstacked ssDNA elasticity, which makes it difficult to detect. The high resolution of our optical tweezers experiments has allowed us to resolve this small, but relevant, effect.

6.4   CONCLUSIONS

In this section we have studied the S-U transition using pulling experiments. We have demonstrated that the blocking-splint oligo method (BSOM) is useful to study this transition in short molecules, for which the control in the sequence is easier than in long sequences, such as the one studied in [72].

We have developed a 2-states model, which allow us to obtain the degree of stacking of each sequence studied as well as the elastic parameters that characterize the totally stacked and unstacked conformations, using the WLC description.

By studying 4 different sequences at the same salt concentration (10mMMgCl$_2$), we have checked that the degree of stacking given in each one of them is a function of the consecutive number of adenines present in them, as was expected from previous works. We have also demonstrated that for a short random sequence without consecutive adenines, the behaviour corresponds to an ideal ssDNA, without any signatures of the S-U transition, well captured by an extensible WLC.

The elastic parameters are obtained by fitting the model to experiments with different sequences and varying salt concentration, from 50mM NaCl to 1M NaCl.

Furthermore, the fitting model is used to determine free-energy of stacking. The obtained value at 10mM MgCl$_2$, $\Delta G^0_{stack} = 0.180(6)$kcal/mol, is in agreement with previous works [72], as well as with our results obtained for the 1M NaCl experiments $\Delta G^0_{stack} = 0.210(11)$kcal/mol, in agreement with the 100/1 rule. Also, we have obtained a salt dependence of ΔG^0_{stack}, which indicates that, at least partially, the salt dependence of the free energy of formation (given by $m = 0.11$kcal/mol) is due to the stacking. This is in agreement with the electrostatic nature of stacking, as described in Chapter 3, but it is in contrast with previous results shown in [72].

Further work could be done, by varying the MgCl$_2$ concentration in the experiments, characterizing the elastic response of the unstacked and stacked state regarding the difference between monovalent and divalent ions. Also, a deeper study of the stacking formation in short molecules could be performed by performing experiments varying the loop sequence. Then, by varying the motifs present in the loop and repeating the analysis presented in this chapter one could investigate more in detail which motifs act as nucleation sites for the stacking of the sequence. Another approach would be to verify the sequences studied using calorimetry measurements [94].

SECONDARY STRUCTURE

Bases in nucleic acid chains (ssDNA and ssRNA) interact forming base-pairs through hydrogen bonding and stacking leading to the formation of secondary structures, which consist of helices, loops, bulges and junctions. Biologically active RNAs are often found as folded stranded polymers, with base-pairing stabilizing relatively rigid structures and flexible non-base-paired regions allowing for functional RNA conformational changes. On the other hand, even if DNA is mainly found in the double-stranded form, conforming the cell or bacterial genome, the single-stranded form is generated continuously, whenever DNA is expressed, duplicated or repaired. The spontaneous formation of non-specific secondary structure in ssDNA impedes the acces of enzymes processing DNA. For this reason, all living organisms contain single-stranded binding (SSB) proteins that bind to DNA protecting it from degradation and avoiding secondary structure formation.

The formation of secondary structure in nucleic acids can be investigated by means of force-spectroscopy techniques, such as optical and magnetic tweezers, which allow for testing the elasticity of the nucleic acid chain. By applying mechanical forces (in the pN range), the force as a function of the ssDNA/RNA extension can be measured [63, 103], as we have seen in previous chapters. The presence of secondary structure induces a condensation in the ssDNA/RNA molecule at forces below ~ 15pN, which is revealed as a force plateau in the measured force-extension curve (FEC) [61, 104] (see for example Chapter 5, Fig. 5.1B). Whereas the elastic response at high forces can be described using ideal elastic polymer models (such as the the Worm-Like Chain –WLC– model) at low forces models that include the secondary structure formation are needed. Here, we aim at investigating the formation of secondary structure in DNA by measuring the ssDNA elasticity using optical tweezers over a wide range of experimental conditions (varying the ssDNA molecular length and the ionic strength). We introduce a helix-coil model to interpret the results.

7.1 EXPERIMENTS

7.1.1 *Obtaining ssDNA elasticity*

We use 8 different DNA hairpins that we mechanically unzip to generate ssDNA. The hairpins used are the ones introduced in Chapter 5: H_{700}, H_{964}, H_{1904}, H_{4508}, H_{7138} and H_{13680}. Two shorter hairpins are also studied: H_{120} and H_{204}. The synthesis of all the hairpins is described in App. B.
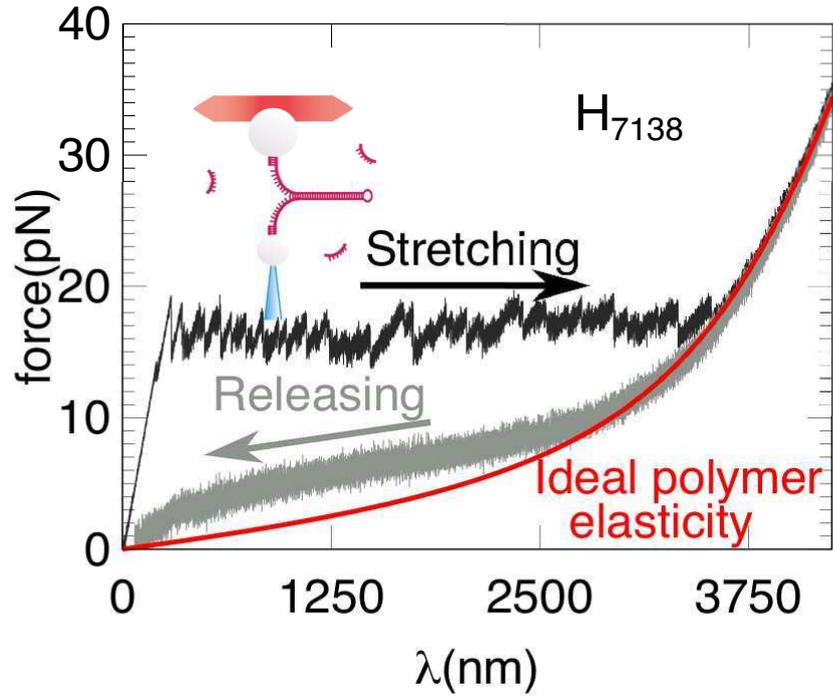
Figure 7.1: **Obtaining ssDNA for long molecules.** The hairpin is mechanically unzipped (black) by stretching a tethered molecule (in this case, H_{7138}), moving the trap position away from the bead in the pipette. After the molecule is completely open, the blocking oligonucleotide binds to its fully complementary region, which is now exposed, preventing the rezipping of the hairpin.

In order to measure the elasticity of ssDNA we first mechanically unzip a DNA hairpin by applying a force above 15 pN at the extremities of the hairpin molecule. We next reduce gradually the force and use a loop-complementary blocking oligonucleotide to prevent the rezipping of the hairpin, as described in Sec. 7.1.1. In Fig. 7.1 we show the unzipping of the H_{7138} hairpin, presenting the characteristic sawtooth shape[82] (black), and the relaxation curve after the oligonucleotide hybridizes to the loop region (grey). The latter corresponds to the elasticity curve for the ssDNA molecule. Several stretching (force increasing) and releasing (force decreasing) cycles are recorded for several tethers and the force extension curve is estimated as the average curve between different tethers. The ssDNA molecular extension $x_{\text{ssDNA}}(f)$ is related to the measured distance, λ, as described in Eq. 5.1: $x_{\text{ssDNA}}(f) = \lambda(f) - x_h(f) - x_b(f) - x_0$, where $x_h(f)$ is the extension of the handles, $x_b(f)$ is the displacement of the bead from the center of the optical trap, and x_0 is a shift of the trap position relative to the position detector. All the quantities and how to obtain them is detailedly explained in Sec. 5.2.1.

For the 120b construct, the 2-branches method, described in Sec. 5.3.1 is used. At forces above the hairpin refolding force and below the hairpin unfolding force the subtraction of the trap position between the folding (pulling cycle) and unfolding (relaxing cycle) force-extension branches at a given force f, $\lambda_F - \lambda_U$, gives the extension of the ssDNA, x_{ssDNA}

(Eq. 5.2): $x_{\text{ssDNA}}(f) = \lambda_F(f) - \lambda_U(f) + x_d(f) + \lambda_0(t)$, where $x_d(f)$ is the diameter of the hairpin projected along the stretching direction (which is equivalent to a dipole orientation under an external field, as described in Chapter 5) and $\lambda_0(t)$ is the correction of the drift.

Both methods are used to obtain the force-extension curves (FECs) of the ssDNA generated by the eight different hairpins studied (Fig. 7.2A) with molecular extensions spanning over two orders of magnitude. The elasticity of molecules of different lengths can be compared by dividing the extension over the number of nucleotides of each sequence, N, as shown in Fig. 7.2B.

The ssDNA FEC differs from the ideal elastic curve (represented in red using the WLC model) at low forces, where it appears a plateau that has been previously associated with the presence of secondary structure [61, 84] (see Fig. 7.2B). The formation of the secondary structure results in a condensation of the ssDNA polymer and hence the measured extension is shortened as compared to the ideal case. Several factors, such as the DNA sequence, length and ionic strength, influence the relative weight of the secondary structure and the ideal ssDNA form. Since ssDNA is a highly charged polymer (2 electrons per phosphate group), the monovalent and divalent ions of the buffer screen the negative charge of the phosphate backbone, favoring the secondary structure form. In order to investigate in detail the transition from the ideal chain regime to the secondary structure condensed-like regime we performed ssDNA elasticity experiments changing the concentration of monovalent and divalent ions as well as the DNA sequence and length (see below).

7.1.2 *Elastic model*

Different polymer models have been used to describe the elasticity of nucleic acids, discussed in Chapter 4. In this chapter, we use the Worm-Like Chain (WLC) model, to interpret the measured ssDNA elasticity.

The elastic parameters used in this work are the ones reported in [61]: a constant contour length per base of $l = 0.69$nm/base and a persistence length which depends on salt concentration, $[C](\text{mM})$, as:

$$p(C) = p_0 + A[C]^{-1/2}, \tag{7.1}$$

where $p_0 = 0.68$nm and $A = 2.19$nm mM$^{-1/2}$ for NaCl (with C =[NaCl]) and $p_0 = 0.70$nm and $A = 1.6$nm mM$^{-1/2}$ for MgCl$_2$, with $C = 100 \cdot [\text{MgCl}_2]$.

7.2 HELIX-COIL MODEL

We have developed a minimal helix-coil cooperative model to describe the ssDNA elasticity taking into account the formation of secondary structure that leads to the deviation from the ideal polymer behaviour. The ssDNA molecule is considered as a chain of N monomers(1 monomer=1 base) in a thermal bath at temperature T, which is stretched by a controlled force, f. Each monomer is described by its state, σ_i (equiva-
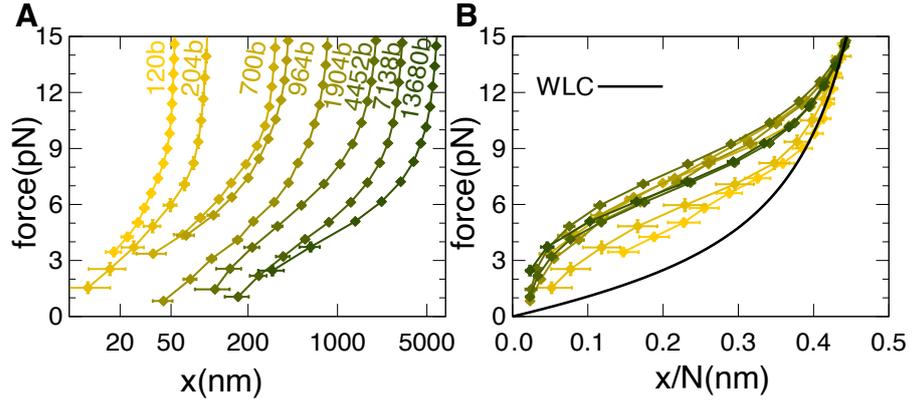
Figure 7.2: **Generating ssDNA elasticity curves from DNA hairpins**.
(A) ssDNA FECs obtained from eight different hairpins with sizes ranging from
120b to ∼14000b, obtained as described in Sec. 7.1.1. The x axis is represented
in log scale to visualize better all the data. The color code is chosen to indicate
the molecular length: darker(brighter) colors show shorter(longer) molecules.
(B) Re-scaled FECs, where the molecular extension measured for each hairpin
has been divided over the number of bases N conforming the hairpin. The
color code as in panel A. The secondary structure emerges at forces below
∼ 10pN, as a plateau deviating from the inextensible WLC(Eq. 4.13) shown
as a continuous black line. The ssDNA elasticity obtained for hairpins with
$N \gtrsim 500$ collapse in a single curve, whereas the 204b and 120b molecules
present a force plateau at lower forces, a signature of finite size effects. The
uncertainties are the standard errors obtained from averaging at least 4 different
molecules for each hairpin. The experiments were performed at 25 °C in a
buffer with 10mM $MgCl_2$ 10mM Tris-HCl pH7.5, 0.01%NaN$_3$.

lent to the spin in the Ising formalism). In the simplest approximation,
each monomer i can only be in one of the two possible states: $\sigma_i = -1$,
forming secondary structure, and $\sigma_i = 1$ forming an ideal elastic polymer
chain. We model the ideal polymer state using the elastic WLC model,
whereas the secondary structure state is considered as a hairpin-like
structure stabilized by an average base-pairing energy ϵ. The model
includes cooperativity through an energy coupliing parameter γ that
penalizes(rewards) adjacent monomers in different(equal) σ_i nearest-
neighbour states. Let $N_1 = \sum_i^N \delta_{\sigma_i,1}$ and $N_{-1} = \sum_i^N \delta_{\sigma_i,-1}$ be the
total number of bases in the ideal polymer chain state and the total
number of bases contained in the secondary structure, respectively. The
Hamiltonian of such a system reads as:

$$\mathcal{H} = -\frac{\epsilon}{2}N_{-1} - N_h \int_0^f x_h(f')df' - N_1 \int_0^f x_1(f')df' - \gamma \sum_i^N \sigma_i\sigma_{i+1}. \quad (7.2)$$

The first term accounts for the energy of formation of secondary structure;
the second term represents the stretching energy of the N_h hairpins in
the polymer, each contributing with an extension $x_h(f)$ given by a single
bond oriented due to an externally applied force (App. E.4); the third
term represents the stretching energy of the ideal ssDNA form with an
extension per base of $x_1(f)$, obtained from Eq. 4.13, as described in
Section 7.1.2; and the last term stands for the coupling energy between
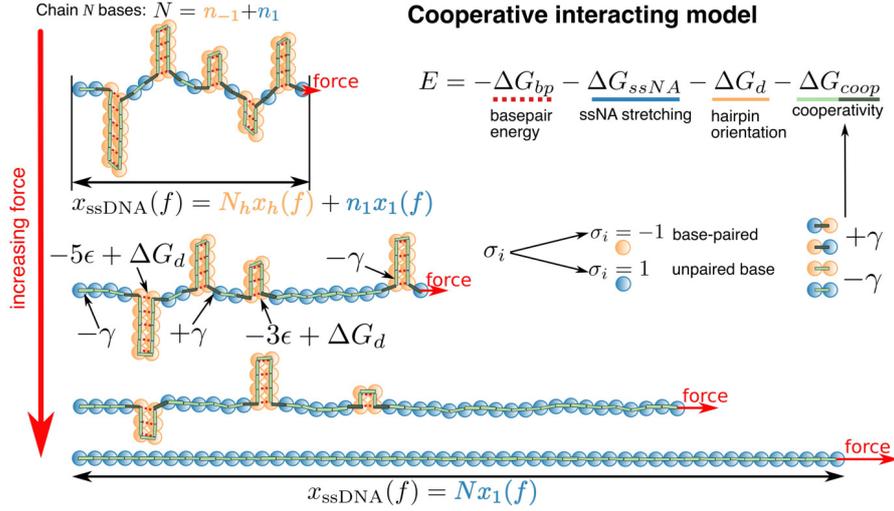
Figure 7.3: **Sketch illustrating the cooperative interacting model for secondary structure formation**. Each nucleotide can be in either single-stranded (blue, state $\sigma_i = -1$) or hairpin (orange, state $\sigma_i = -1$) conformation. Each conformation has its elastic contribution to the energy: ΔG_{ssNA}, for each nucleotide in ss form, and ΔG_d, for each nucleotide in a hairpin-like form. A domain of consecutive nucleotides in $\sigma_i = -1$ constitutes a hairpin-like structure. Each nucleotide in the hairpin conformation contributes with an extra $\epsilon/2$ to the energy, with ϵ representing the average energy per base pair. The model also includes an interaction term that takes into account the state of neighbouring nucleotides (ΔG_{coop}): the neighboring nucleotides are energetically favored when they are in the same state (represented as the connecting line beteen adjacent monomers being in light green), and they are energetically penalized when they have different states (represented in dark green). Starting at low force, many nucleotides are forming hairpin-like structures. As the force increases, some of the hairpins start to unfold, creating larger domains in the single-stranded form, until reaching the fully stretched unfolded state where all the nucleotides are in single-stranded form.

adjacent monomers. A schematic representation of this model is shown in Fig. 7.3. At a given force, N_1 bases are at the ideal polymer state (+1 state), each base contributing $x_1(f)$ to the total extension. The rest of nucleotides, N_{-1}, are divided into N_h hairpin-like structures, each hairpin contributing $x_h(f)$ to the fullextension. Hence, the total molecular extension is given by: $x_{\text{ssDNA}}(f) = N_h x_h(f) + N_1 x_1(f)$.

The Hamiltonian described by Eq. 7.2 is analytically solvable (See Sec. 7.2), and leads to a extension per nucleotide given by:

$$\frac{x_{\text{ssDNA}}(f)}{N} = \phi_1(f) x_1(f) + \frac{N_h(f)}{N} x_h(f), \tag{7.3}$$

where the fraction of bases in ideal polymer state state is written as:

$$\phi_1(f) = \frac{1}{2}\left(1 + \frac{\sinh\left(\beta A(f)\right)}{\sqrt{e^{-4\beta B(f)} + \sinh^2\left(\beta A(f)\right)}}\right) \tag{7.4}$$

Where $\beta = 1/k_B T$, $A(f) = -\epsilon/4 + 1/2 \int_0^f x_1(f')df'$ and $B(f) = \gamma - 1/4 \int_0^f x_h(f')df'$. Note that this expression can be used to express the

fraction of bases in secondary structure form as $\phi_{-1}(f) = 1 - \phi_1(f)$. The number of secondary structure domains equals the number of formed hairpins, N_h,

$$N_h(f) = \frac{N}{2} \frac{\sinh(\beta A(f))}{\sqrt{e^{-4\beta B(f)} + \sinh^2(\beta A(f))} + \cosh(\beta A(f))}. \qquad (7.5)$$

From these two equations, the expression for the average number of bases per hairpin follows:

$$n_{av}(f) = \frac{N\phi_{-1}(f)}{N_h}. \qquad (7.6)$$

With only two parameters, ϵ and γ, we fit all the experimental FECs ($2 \leq f \leq 15$pN) in the different conditions tested (different DNA sequences and different salt conditions, see below) using Eq. 7.3, after imposing the ideal elastic ssDNA response ($x_1(f)$) and the contribution of each hairpin to the full extension ($x_h(f)$). We consider the elastic behaviour of the bases that are not forming the secondary structure ($x_1(f)$) to follow the WLC model (Eq. 4.13), with the parameters described in Sec. 7.1.2. In order to describe the elastic contribution from the hairpin-like structures ($x_h(f)$) we consider hairpin-like structures with a diameter d that align along the stretching direction(E.4). Interestingly, using d from 0 to 2 nm(the diameter of a standard double helix DNA hairpin), the results are not much affected (F.1), meaning that the model reproduces well all experimental results and the values obtained from the fits for ϵ and γ change less than $\sim 10\%$. In other words, in the present conditions, the elasticity term due to the hairpin orientation can be neglected. In what follows, for the sake of simplicity we have taken $d = 0$.

7.3 EFFECTS OF THE MOLECULAR LENGTH ON THE FORMATION OF SECONDARY STRUCTURE

In order to investigate the effects of sequence and length of the ssDNA on the formation of secondary structure, we measure the elasticity of the eight different ssDNA sequences (having N bases, from $N \sim 100$ b to $N \sim 14$ Kb), in a buffer containing 10mM $MgCl_2$, 10mM Tris-HCl pH7.5, 1mM EDTA, 0.01%NaN_3. Interestingly, as shown in Fig. 7.2B, the FECs measured for molecules with $N \gtrsim 500$b collapse into a single curve upon rescaling the moleular extension by the number of bases, deviating from the ideal behaviour at forces $f \lesssim 10$pN. Sequence effects might be observed for sequences presenting different GC/AT content, since GC base pairs are about twice more stable than AT base pairs. However, all the sequences we study are close to 50% GC content, so the sequence effects we observe are mild (F.3). On the other hand, the ssDNA molecules with $N \leq 500$b deviate from the collapsed behavior with a force plateau at lower forces ($f \approx 7 - 8$pN). This is a signature of finite-size effects and suggests that secondary structure formation, despite of being essentially originated from short-range interactions, requires a

minimum number of bases. As shown in Fig. 7.4A, the FECs for all the molecules studied (yellow to dark green points) are well reproduced by the proposed model, Eq. 7.3 (green curves). The fitted values for ϵ and γ as a function of the molecule length are represented in Figs. 7.4B-C. The average energy of base-pairing increases with the ssDNA molecular length, until saturating at ~ 0.35kcal/mol, for molecules with $N \gtrsim 500$b. The length dependence of the base-pairing free energy is found to follow the regular correction,

$$\epsilon(N) = \epsilon_0 - \frac{b}{N},\tag{7.7}$$

with $\epsilon_0 = 0.36(2)$kcal/mol and $b = 37(7)$kcal/mol base (Fig. 7.4B), whereas the cooperativity energy γ remains approximately constant throughout all the ssDNA lengths studied (Fig. 7.4C), with an average value of $\gamma = 0.61(2)$kcal/mol. For long molecules, with $N \gtrsim 500$b, the best fitted values of ϵ and γ do not change significantly, in agreement with the observed collapse of the FECs for molecules in this length regime. For shorter molecules ($N \lesssim 500$b), the fitted values for ϵ deviate from the long-length saturation value ($\epsilon_0 \approx 0.35$kcal/mol), showing finite size effects. The smaller values of ϵ obtained in the short length regime indicate that the secondary structure is constrained by the finite length of the molecule in this regime.

7.4 SALT DEPENDENCE OF THE SECONDARY STRUCTURE OF SSDNA

In order to investigate how the concentration of monovalent and divalent ions in the buffer affects the formation of ssDNA secondary structure the ssDNA elasticity we analyze the data from [61], which performed experiments at different NaCl and MgCl$_2$ concentrations with the longest ssDNA molecule (generated from the ~ 14 kb hairpin, H$_{13680}$). The experimental FECs obtained by varying NaCl and MgCl$_2$ concentrations are shown in Fig. 7.5A and B, respectively. The data shows that at higher ionic strengths (represented by higher salt concentrations) the deviation of the FECs from the ideal elastic behaviour (represented as the WLC, black curves) increases. The secondary structure plateau reaches its maximum for the highest salt concentrations for which we have data ([NaCl]=1000mM and [MgCl$_2$]=10mM). On the other hand, at the lower salt concentrations (10, 25 and 50mM NaCl), the FECs show little deviation from the ideal elastic behaviour, with a measured extension that is just slightly longer than what is expected from the ideal case. This shows that, in this low-salt conditions, the secondary structure, whose presence would lead to shorter extensions, is minimal (giving rise to $\epsilon < 0$, see below). The small deviations observed between the measured elasticity and the ideal limit have been suggested to be due to excluded volume effects [61, 105–107].

The proposed helix-coil cooperative model successfully reproduces the FECs for both varying monovalent (NaCl) and divalent (MgCl$_2$) salt concentrations, as shown in Figure 7.5A and B, respectively. It reproduces the stabilization of ssDNA secondary structure by salt screening,
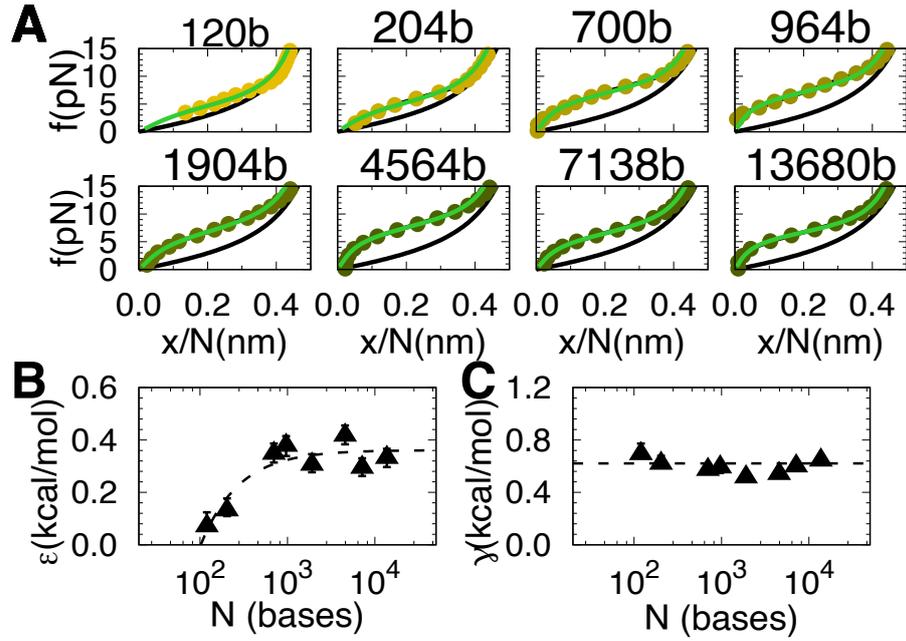
Figure 7.4: **ssDNA elasticity as a function of the molecular length**. (A) Experimental FECs (circles) for the different ssDNA molecules with lengths ranging from ∼120bs to ∼14000bs. The results from the fits to the helix-coil model, described in Eq. 7.3, are shown as a green continuous line. For comparison, the ideal elastic response (in absence of secondary structure) is represented in a black continuous line, using the WLC model (Eq. 4.13). (B) Average energy per basepair, ϵ, obtained from the fits shown in panel A as a function of the ssDNA molecular length (triangles). The dashed line corresponds to the fit to Eq. 7.7, with $\epsilon_0 = 0.36(2)$kcal/mol and $b = 37(7)$kcal/mol base. (C) Nearest-neighbour interaction energy, γ, obtained from the fits shown in panel A as a function of the ssDNA molecular length (triangles). The dashed line corresponds to the mean value from all fits, $\gamma = 0.61(2)$kcal/mol. The errors from the fits have been obtained as described in F.2. The experiments were performed at 25 °C in a buffer with 10mM $MgCl_2$, 10mM Tris-HCl pH7.5, 1mM EDTA, 0.01%NaN_3.

represented as an increase in the FECs plateaus with salt concentration. As shown in Fig. 7.5C, this stabilization is captured by the model as an increase on the average base-pairing energy, ϵ (NaCl (magenta) and $MgCl_2$ (blue)). The observed behavior is well described with logarithmic dependence on salt concentration, as reported for the hybridization of standard nearest neighbours Watson-Crick(WC) base-pairs(bps),[82, 108]:

$$\epsilon(C) = \epsilon_0 + m \log(C), \qquad (7.8)$$

with C being the salt concentration (in M units), ϵ_0 the free energy of base-pairing at 1M and m the slope of the salt correction. The fits of Eq. 7.8 to the ϵ values obtained for each salt concentration are shown in Fig. 7.5C as dashed lines. The results from the fits are $\epsilon_{0,\text{NaCl}} = 0.28(3)$kcal/mol, $m_{\text{NaCl}} = 0.13(2)$kcal/mol for the NaCl data and $\epsilon_{0,\text{MgCl}_2} = 0.74(15)$kcal/mol, $m_{\text{MgCl}_2} = 0.10(2)$kcal/mol, for the $MgCl_2$ data. The data for low NaCl conditions (10, 25 and 50mM) do
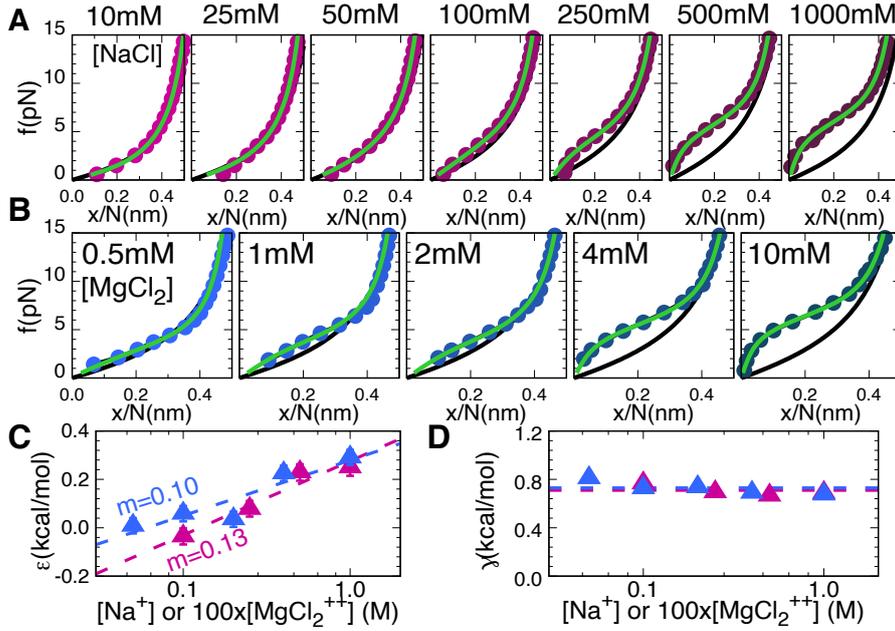
Figure 7.5: **ssDNA elasticity as a function of the ionic strength**. (A) ssDNA experimental FECs (magenta circles) obtained for the H_{13680} hairpin varying [NaCl] concentration. The results from the fits to the cooperative interacting model model, described in Eq. 7.3, are shown as a green continuous line. For comparison, the ideal elastic response (in absence of secondary structure) is represented in a black continuous line, using the WLC model (Eq. **??**). (B) Same as in panel A but for varying [$MgCl_2$] concentration (experimental points shown in blue). (C) Average energy per basepair, ϵ, obtained from the fits shown in panels A and B as a function of the ionic concentration (magenta triangles for Na^+, blue triangles for Mg^{++}). As shown in (A), the 10mM, 25mM and 50mM [NaCl] data do not show significant deviation from the ideal elastic response and are not included. The dashed lines correspond to the fit to Eq. 7.8, with $\epsilon_{0,NaCl} = 0.28(3)$kcal/mol and $m_{NaCl} = 0.13(2)$kcal/mol for Na^+ data and $\epsilon_{0,MgCl_2} = 0.74(15)$kcal/mol and $m_{MgCl_2} = 0.10(2)$kcal/mol for Mg^{++} data. (D) Nearest-neighbour interaction energy, γ, obtained from the fits shown in panel A and B as a function of the ionic concentration (colors as in panel D). The dashed lines correspond to the mean value from all fits, $\gamma_{NaCl} = 0.70(3)$ kcal/mol (magenta) and $\gamma_{MgCl_2} = 0.72(3)$ kcal/mol (blue). The $MgCl_2$ buffers contain 10mM Tris-HCl pH7.5, 0.01%NaN_3, while the NaCl also add 1mM EDTA to it. All experiments were performed at $T = 25\,°C$. The errors from the fits have been obtained as described in F.2.

not show secondary structure formation (i.e, their FECs throughout the force range is well captured by the WLC (Eq. **??**)) and their curves can be fitted well for any $\epsilon < 0$. Therefore the values of the fits obtained for NaCl of 10mM, 25mM and 50mM are not represented in Fig. 7.5C-D.

The values obtained for ϵ_0 are ~ 5 times smaller than the average Watson-Crick(WC) base-pairing energy and almost 3 times smaller than the most unstable WC-pair (AT/TA, ~ 0.85kcal/mol) [82, 108]. This is compatible with the fact that the studied molecules have random sequences, lacking regions of full complementarity, and the generated secondary structure might consist mostly of non-specific base-pair aggre-

gates, including loops and mismatches. Therefore, the energies associated to the secondary structure observed are expected to be much smaller than those of WC bps. In agreement with that, the unfolding forces of hairpins (with WC bps) measured in force-spectroscopy experiments [109] are of the order of 15pN, compared to $\sim$ 5pN force plateau associated to the secondary structure formation. Interestingly, the slope $m_{\text{NaCl}} = 0.13(2)$kcal/mol (Eq. 7.8) obtained for varying NaCl concentration is compatible with the homogeneous salt correction parameter of the unified oligonucleotide data given by [110] and mechanical unsipping [82] ($m = 0.114$kcal/mol and $m = 0.104$kcal/mol, respectively). Hence, the measured non-specific secondary structure salt dependence is similar to the WC-basepairing, suggesting that the underlying phenomenon is similar in both cases. For the MgCl$_2$ case, however, the obtained slope $m_{\text{MgCl}_2} = 0.10(2)$kcal/mol is about twice the value used in Mfold [102] ($m = 0.55$kcal/mol), where the assumption that the salt correction for the monovalent ions is exactly twice the correction for the divalent ones holds ($m_{\text{NaCl}} = 2 \cdot m_{\text{MgCl}_2}$). A definition for the equivalent monovalent concentration is proposed in [83]:

$$[Mon_{eq}] = \left[Na^+ \right] + \beta \log \left(\left[Mg^{++} \right]^\alpha \right), \tag{7.9}$$

with concentrations being given in units of M. The values for DNA hybridization used by Mfold [110] are $\beta = 3.3$ and $\alpha = 0.5$, whereas those for sequence-specific hybridization obtined from unfolding experiments [83] range from $13 \leq \beta \leq 254$ and $0.46 \leq \alpha \leq 0.87$. With the values obtained for m and ϵ_0 from our NaCl and MgCl$_2$ data, the corresponding monovalent equivalence factors for the secondary structure formation are $\beta = 55$ and $\alpha = 0.85$. This shows that the stabilizing effects of divalent and monovalent ions are quantitatively different regarding secondary structure formation.

Finally, the cooperativity term γ remains constant throughout all the salt concentrations studied, both monovalent and divalent (Fig. 7.5D). The averaged values obtained for the cooperativity are $\gamma_{\text{NaCl}} = 0.70(3)$kcal/mol and $\gamma_{\text{MgCl}_2} = 0.72(3)$kcal/mol, for monovalent and divalent salt concentrations. The values are compatible between them and comparable with the one obtained for the fits of varying length (see Sec. 7.3). The free energy of formation of the secondary structure would only depend on ϵ, if all the bases at zero force were to form secondary structure. However, this is not the case, and hence, this free-energy of formation will be also γ-dependent.

7.5   MODEL PREDICTIONS

The model developed in Sec. 7.2 allows us to make predictions about how the bases are distributed between the ideal ssDNA chain and the secondary structure domains. From the model we can compute the fraction of bases forming secondary structure at each force f, $\phi_{-1}(f)$. The fraction $\phi_{-1}(f)$, computed for the different experimental conditions studied, is shown in Fig. 7.6A for varying the molecular length (top),

NaCl concentration (middle) and $MgCl_2$ concentration (bottom). Insets of Fig. 7.6A show the fraction ϕ_{-1} for each condition at two different forces, $f = 0$ and 6pN. The model predicts that the fraction of bases forming secondary structure, $\phi_{-1}(f)$, start at a maximum value $\phi_{-1}(0)$, and decreases until a characteristic force is reached, f_c, above which the secondary structure vanishes ($\phi_{-1}(f_c) \approx 0$) and the experimental FEC collapses with the WLC. This characteristic force depends on the molecular length and salt concentration. For long molecules ($N \geq 500$b), this force is around ~ 10pN, whereas for shorter molecules the disruption of secondary structure takes place at lower forces ($f_c \approx 7 - 8$pN). The same trend is observed for varying NaCl and $MgCl_2$ concentration (Fig. 7.6A, middle and bottom panels). At high salts the stability of secondary structure is increased and a large force of $f_c \approx 10$pN is required to destabilize it, while at lower salt concentrations (100mM NaCl, $0.5MgCl_2$) the secondary structure vanishes at a lower force of $f_c \sim 5$pN.

The fraction $\phi_{-1}(0)$ quantifies the amount of secondary structure in DNA in the random coil state at zero force. This fraction cannot be directly measured with the present technique but could be measured by other means, such as flourescence or NMR. The model predicts that for $N \gtrsim 500$b at high salt (1M NaCl) 85% of the bases form secondary structure. This percentage of bases decays to $70 - 75\%$ for shorter lengths $N = 100 - 200$b. A more drastic effect is observed with salt. At the lowest salt concentrations where secondary structure is detected ([NaCl] $= 100$mM and [$MgCl_2$] $= 0.5$mM) the percentage of bases forming secondary structure is $\sim 45\%$, whereas this percentage increases to $\sim 85 - 90\%$ for the highest salt conditions studied ([NaCl] $== 1$M and [$MgCl_2$] $= 10$mM). Note that the values of $\phi_{-1}(0)$ obtained for the higher salt concentrations (~ 0.86 for 1M NaCl and ~ 0.89 for 10mM $MgCl_2$) are very similar and in agreement with the 100/1 salt rule [98, 99]. Furthermore, they are also compatible with those obtained for long molecules (with experiments performed at 10mM $MgCl_2$), where $\phi_{-1}(0) \approx 0.85$, as shown in Fig. 7.6A(top). Note that the model predicts the presence of secondary structure at zero force ($\phi_{-1}(0) \neq 0$, dashed line, inset of Fig. 7.6A middle) for the lower NaCl salt concentrations (10-50 mM NaCl), for which the measured FECs follow the ideal elastic response (Fig. 7.5A). This is a consequence of the secondary structure appearing at low forces (below 5pN) for these low salt conditions, which results in a very small, non-measurable, change in the molecular extension and the absence of the force plateau.

Another issue that can be investigated with the proposed model is how the bases are grouped into different secondary structure domains. In particular, the average number of bases per domain, $n_{av}(f)$, can be estimated using Eq. 7.6. As shown in Fig. 7.6B, $n_{av}(f)$ follows a similar behavior to $\phi_{-1}(f)$: it is maximum at zero force ($n_{av}(0)$), and decreases as the force increases, until a characteristic force is reached, f_c, for which $n_{av} \rightarrow 1$ (the minimum value allowed by the model). The value of f_c depends on salt concentration, following the same trend as for ϕ_{-1}, i.e. $f_c \approx 10$pN for high salt ([NaCl]$=1000$mM), while it drops to $f_c \approx 6$pN
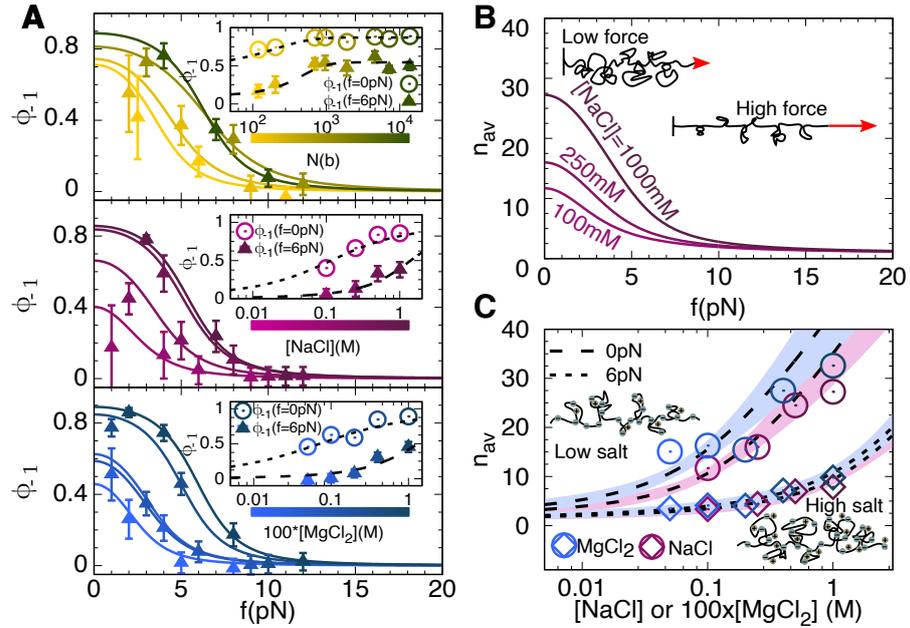
Figure 7.6: **Predictions from the model** (A) Fraction of bases forming secondary structure, ϕ_{-1} , as a function of the force for: (top) 4 different molecular lengths (120, 204, 1904 and 13680 bases, from light to dark color), (middle) 4 different NaCl concentrations (100, 250, 500 and 1000mM, from light to dark color) and (bottom) 5 different $MgCl_2$ concentrations (0.5, 1, 2, 4 and 10mM, from light to dark color). The insets show the dependence of the fraction ϕ_{-1} with respect to the molecular length and salt concentration at two different forces, 0 and 6pN. The dashed lines are obtained from Eq. 7.7(top) and 7.8(middle and bottom) using the fitted values of ϵ and the average γ (shown as dashed lines in Figs.7.4C and 7.5D). (B) Average number of base-pairs per secondary structure folded domain, n_{av}, as a function of the force for 3 different NaCl concentrations (100, 250 and 1000mM), as predicted by the model. The schematic depictions show a highly compacted form of the polymer at low forces that is stretched at higher forces, decreasing the size of the folded domains. (C) Average number of base-pairs n_{av} as a function of the salt concentration (magenta for NaCl and blue for $MgCl_2$) at two different forces, 0 and 6pN. Dashed lines are obtained from Eq. 7.6, taking ϵ from Eq. 7.8 and a constant γ (Figs. 7.5C, 7.5D). The shadow area shows the uncertainty region of n_{av}, considering the errors in ϵ and γ. The schematic depictions show the negatively charged DNA screened by the cations in the solvent facilitating stable secondary structure domains.

for low salt ([NaCl]=100mM). For $f \gtrsim f_c$ secondary strucutre formation is unstable and only a few bases domains are formed. The model shows that increasing the applied external force destabilizes the formation of secondary structure, by reducing both the fraction of bases that are forming secondary structure and the size of the secondary structure domains (sketch in Fig. 7.6B). Regarding the dependence of n_{av} with respect to salt concentration, the model predicts a marked increase in n_{av} when increasing [NaCl] or [$MgCl_2$], as shown in Fig. 7.6C. At zero force, the values for n_{av} range from ~ 15b, for the lowest salt conditions, to ~ 30b, for the highest ones. This range of values decreases as the force increases. At $f = 6$pN, the lower limit $n_{av} = 1$, is reached at low
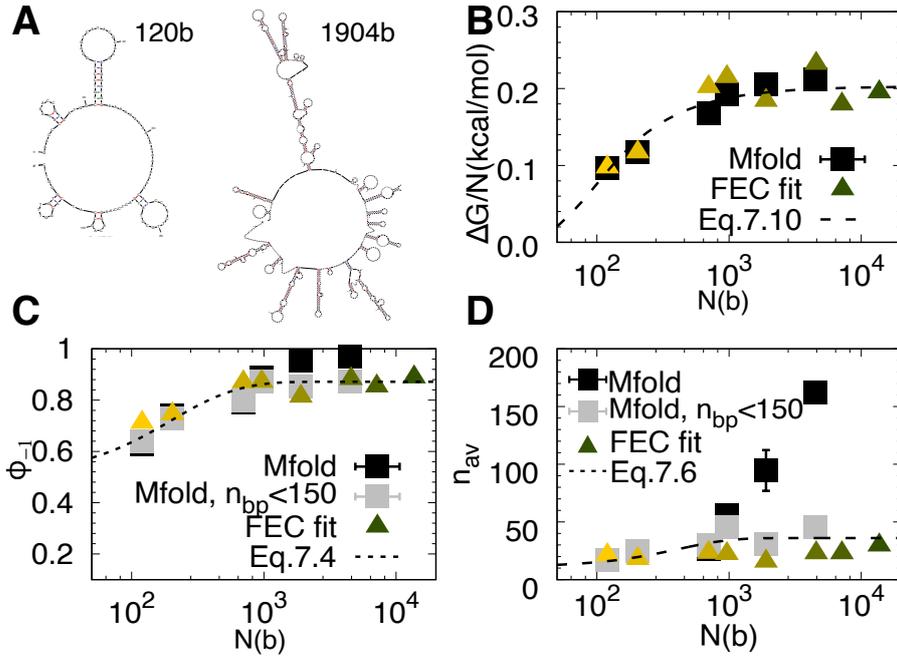
Figure 7.7: **Comparison between our model and Mfold.** (a) Native structures given by Mfold for a short (120b) and long hairpin (1904b). (b) Comparison of the free energy at zero force per base, $\Delta G/N$, given by Mfold and the helix-coil model. (c) Comparison of the fraction of bases that are in secondary structures (i.e. inside hairpins) for Mfold and the helix-coil model. (d) Average number of bases per hairpin, n_{av}. Same color code than in (c). All the Mfold uncertainties are standard errors, while for the values of the FEC fit correspond to the uncertainty propagation of γ and ϵ for Eqs. 7.10, 7.4 and 7.6.

salt concentrations, while at high salt concentrations ([$MgCl_2$]=10mM and [NaCl]=1000mM) $n_{av} \approx 8$b.

7.6 COMPARISON OF THE MODEL WITH MFOLD PREDICTIONS

Mfold is a computational method that is used to predict secondary structures of nucleic acids from its sequence, in the absence of an external force. In order to further validate the developed model, we compare out model predictions at zero force with those obtained for Mfold for the 120b, 204b, 700b, 964b, 1904b and 4454b sequences. The two longer sequences were not studied since the required computation time for determining the structures with Mfold is proportional to N^3 [111]. Mfold returns the set of structures that minimize the free-energy of formation (that is the free energy difference between the predicted structure and the ideal ssDNA without any secondary structure) of a given ssDNA sequence (Fig. 7.7A). Besides the expected value for the free energy of formation of the predicted structures, ΔG, the parameters ϕ_{-1}, n_{av} can also be obtained from analyzing the structures, by counting the amount of bases inside the secondary structure motifs and the average length of the motifs formed, respectively.

The free-energy of formation (at zero force) given by Mfold can be compared with the one ptedicted by the model. To do so, we derive an analytical expression for the difference in free-energy between the equilibrium secondary structure state and the totally unfolded configuration (App. C):

$$\frac{\Delta G}{N} = -\frac{\epsilon}{4} - \frac{1}{\beta} \log \left[\cosh \left(\frac{\beta\epsilon}{4} \right) + \sqrt{e^{-4\beta\gamma} + \sinh^2 \left(\frac{\beta\epsilon}{4} \right)}) \right]. \quad (7.10)$$

We use the free-energy normalized over the total number of bases, N, for comparing different DNA lengths. Notice that the theoretical expression for $\frac{\Delta G}{N}$, only depends on ϵ and γ.

As shown in Fig. 7.7B, there is a very good agreement between the free energy of formation given by the model (Eq. 7.10) – both using the values of ϵ and γ for FEC fit (colored triangles) and the average value of γ and ϵ from Eq. 7.7 (dashed line)–, and the values predicted by Mfold (black squares).

Regarding the fraction of bases in hairpins at zero force, ϕ_{-1}, and the average number of bases per hairpin, n_{av} the agreement between the model and the Mfold predictions for short molecules ($N < 1000b$) is also good (Fig. 7.7C,D). However, for long molecules differences appear: whereas Mfold predicts that almost all bases form secondary structure, with n_{av} increasing strongly with N, our model shows a significant amount of bases that are in the ssDNA form (more than $\sim 10\%$) and an average number of bases per hairpin that saturates around $\sim 40b$.

This discrepancy can be a consequence of the mean-field nature of our model, that does not consider sequence heterogeneity and has a constant ϵ and γ for all possible base-pairs. Taking into account specific sequences (as Mfold does), might favour complementarity between regions that are far apart in the ssDNA chain, greatly increasing the size of the secondary structure domains, as well as the number of bases forming them. However, Mfold is also a mean-field like approximation, somehow, as tertiary structures are not included (i.e. only planar ones are calculated).

To test wether the origin of the discrepancies found between Mfold and the model are due to the formation of such large domains, we constrained the computed structures by Mfold to be a maximum of 150 bases apart. Indeed, as it is shown in grey in Fig. 7.7C and D, the obtained values for both ϕ_{-1} and n_{av} are in agreement with those given by the model, which supports the interpretation given above.

7.7 CONCLUSIONS

Secondary structure formation can be detected in force spectroscopy assays, in which a ssDNA molecule is stretched by applying a mechanical force. The formation of secondary structure leads to a condensation of the ssDNA chain that is observed as a deviation of the measured FEC from the ideal polymer response (modelled as a WLC) leading to a force plateau at low forces[61, 104] (Figs. 7.4A and 7.5A). Several

factors affect the stability of secondary structure, such as the specific DNA sequence and molecular length, and the ionic strength of the environment. Here we use optical tweezers to measure the elasticity of ssDNA molecules along a wide range of conditions (lengths ranging from 100b to 14000b in different monovalent and divalent salt concentrations) and propose a minimal cooperative interacting model to describe the secondary structure formation that reproduces all experimental results. Compared to more complex theoretical frameworks [97, 104], the model proposed includes only two parameters: ϵ, which represents the average energy of secondary structure motifs per base, and γ, which accounts for the energy of coupling between adjacent bases. These two parameters are enough to capture the phenomenology observed in the wide range of experimental conditions tested. Typical values of the average energy ϵ are ~ 5 times smaller than those of canonical WC base-pairing, which is in agreement with the lower stability of the expected structures, compared with perfectly hybridized strands. Non-specific secondary structure might contain both WC and non-WC base-pairing, as well as loops and bulges. Finite-size effects are observed for ssDNA molecules $N \lesssim 500$b as a decrease in the values of ϵ, described by an inverse dependence of ϵ with the ssDNA molecular length (Eq. 7.7). On the other hand, ϵ presents a logarithmic dependence on the salt concentration (Eq. 7.8), consistent with salt effects being predominantly entropic. The slope obtained for the logarithmic relation is in agreement with that reported for dsDNA base-pairing [82, 110] for the case of NaCl($m_{\text{NaCl}} = 0.13(2)$kcal/mol), while it is substantially larger (~ 2 times) for the MgCl$_2$ case ($m_{\text{MgCl}_2} = 0.10(2)$kcal/mol), which a value of ~ 0.055kcal/mol was expected. Interestingly, the model requires some cooperativity-inducing coupling parameter, γ, to reproduce the experimental results, which value is independent of salt concentration ($\gamma \approx 0.65$kcal/mol in all conditions studied), indicating a coupling interaction of enthalpic origin. The model can also be used to compute the distribution of bases forming secondary structure, as well as the size of the secondary structure motifs at different forces. In particular, the model predicts that at zero force $\sim 50 - -85\%$ of the bases are forming secondary structure regions of average size of $\sim 15 - -30$ bases, depending on the molecular length and the salt concentration used.

The model is in good agreement with the free energy of formation of all the molecules studied as given by Mfold. However, it fails to reproduce the distribution of bases (information contained in the parameters ϕ_{-1} and n_{av}) for long bases, with sequences longer than ~ 1000b. This is due to the limitation given by a homogeneous base pairing energy, ϵ, and therefore the heterogeneous nature of the sequence could be incorporated to the model by including some variation of ϵ. Nevertheless, the helix-coil model captures the distribution of the bases for relatively short molecules $N < 1000$b, being compatible with those predicted by Mfold.

Both the number of bases forming secondary structure and the average size of the secondary structure motifs are maximal at zero force and decrease until vanishing above a characteristic force f_c ($\simeq 10$pN for long

molecules at high salt concentrations). Interestingly, the model predicts the presence of some secondary structure at very low salt conditions (below 100mM NaCl) where the measured FECs follow the ideal polymer elastic response throughout the force range. This is due to the low values obtained for the characteristic force f_c for low salt concentrations (2-3pN), giving rise to very small changes in the measured extension, which are undetected in our measurements. These model predictions about the amount of secondary structure in different conditions are difficult to measure using force-spectroscopy techniques, but they might be explored by imaging or fluorescence.

Part III

COLLABORATIONS

UNWINDING DNA BY RECQ HELICASE

Every time the information contained in DNA needs to be accessed (for DNA replication, reparation or transcription), dsDNA is unzipped. This reaction is catalyzed by a specific family of proteins (introduced in Chapter 3), helicases, that convert the chemical energy from ATP hydrolisis to the mechanical energy required to translocate along ssDNA while unwinding the dsDNA helix.

RecQ helicases, which belong to the Superfamily II group of helicases, help to maintain stability of the genome and suppress inappropriate recombination in organisms ranging from bacteria to humans. They function by catalyzing ATP and driving the unwinding of paired DNA, while translocating in the 3'-5' direction [112, 113].

Deficiencies and/or mutations in RecQ family helicases display aberrant genetic recombination and/or DNA replication, which leads to chromosomal instability and decreased ability to proliferate. These result in the appearence of several diseases, such as the autosomal recessive diseases Bloom syndrome, Rothmund-Thomson syndrome, and Werner syndrome.

Here we study the E. coli RecQ, which plays a central role in different DNA repair and recombination pathways. The helicase activity is assayed in a DNA hairpin, which is manipulated by an optical trap and indirectly measured through the changes in the measured molecular extension (see below). We perform an analysis of the efficiency of the motor, for which the correct characterization of the elastic response of ssDNA is critical, since it influences the energetics. The work presented in this chapter has been done in collaboration with Vegard Børve Sørdal, from the Department of Physics of the University of Oslo.

8.1 FORCE-CONTROLLED OPTICAL TWEEZERS EXPERIMENTS TO TEST HELICASE ACTIVITY

We use a 480bp hairpin, which is tethered between two beads: one, held in the tip of the micropipette and the other optically trapped. The molecule is stretched under a constant externally applied force, using a force-feedback protocol, described in [5]. Briefly, the force signal is read by the protocol, which then translates the signal to a movement in the wiggler, regulating the trap position to keep the force constant to a preset value. The wiggler then moves upwards(downwards) if the measured force is lower(higher) than the preset value, which is translated into a displacement in the optical trap position, λ.

When applying a high enough force ($\sim$ 15pN) the hairpin unzips. Below that force the hairpin is folded, and the helicase unwinding reaction can be followed by monitoring the changes in the molecular extension, which, in the constant force mode is directly given by the change in the
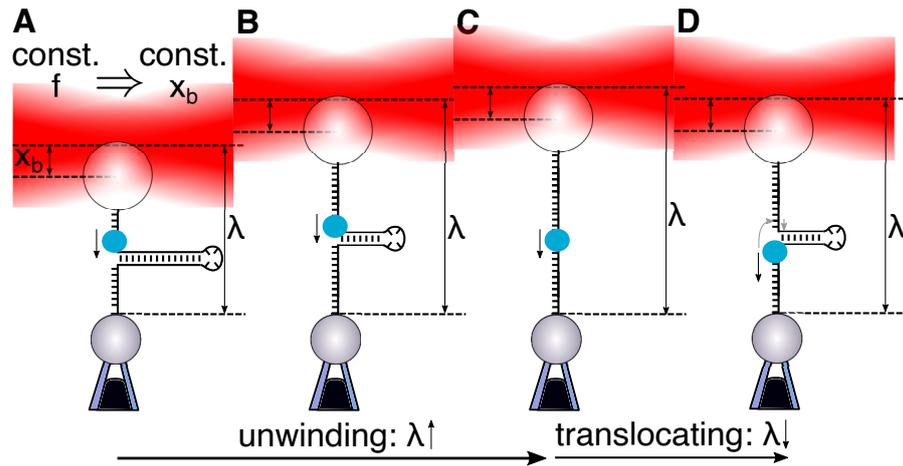
Figure 8.1: **Experimental setup**. (a) Initial condition, with the helicase (blue circle) bound to a segment of exposed ssDNA, unwinding the 480bp hairpin. The optical trap exerts a constant force by moving λ in order to keep x_b constant. (b) As the helicase unwinds dsDNA, the trap distance λ increases. (c) After dsDNA is fully unzipped, all the measured extension corresponds to ssDNA. (d) After passing the loop, the hairpin starts refolding again. However (gray), RecQ can switch strands, which permits the helicase to repeat the unwinding of the loop region several times.

measured λ. Since RecQ requires a ssDNA segment to bind and start unwinding, the experiments starts by mechanically unzipping the first $\sim$ 50bps. After injecting RecQ and ATP into the chamber, the helicase binds to the DNA and starts translocating, unwinding the hairpin, as shown schematically in Fig. 8.1A. The unwinding increases the amount of released ssDNA and, at constant force, this is translated to an increase in the measured trap distance λ. The measured λ becomes maximum when the molecule is fully unzipped(Fig. 8.1C). After the helicase fully unwinds the dsDNA, it keeps translocating forward along the ssDNA strand, while the hairpin rezips in its wake(Fig. 8.1D). This allows to study both the dsDNA unwinding as well as the ssDNA tranlocating helicase activities. Note that RecQ helicase translocates on ssDNA in a well defined 3'-5' direction. However, RecQ can switch strands during the unwinding, which leads to repetitive unwinding/translocating events with the same helicase [114], as shown in Fig. 8.2.

8.1.1 *Kinetics*

In Fig. 8.2 a typical experimental helicase unwinding trace is shown. When the helicase binds to the initial ssDNA segment and starts to unwind the hairpin, the position of the optical trap is moved away from the bead in the micro-pipette in order to accomodate the released ssDNA and to keep a constant applied force on the DNA. This is observed as a rising edge in the extension trace, as shown as a red box in the figure. After some time, the helicase succesfully unwinds the whole DNA hairpin, reaching a plateau at maximum extension (Fig. 8.2).
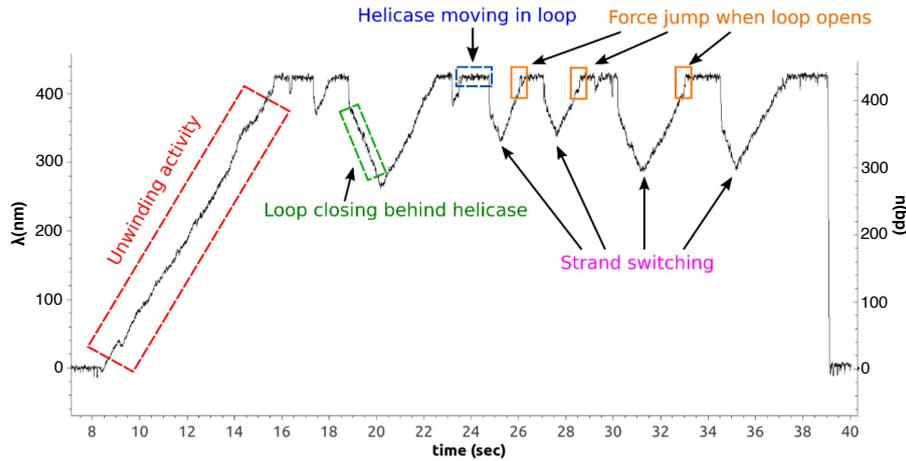
Figure 8.2: **Trace of a typical measured event**. A single helicase starts to unwind the molecule, which translates into an increase in the measured extension, λ. After a few seconds, the hairpin is fully unzipped, and it starts to rezip in the wake of the helicase. Finally it switches strand several times until the hairpin folds again. Figure from [115].

Besides this forward unwinding process, RecQ helicase show other behaviours: it can detach from DNA, as well as switch strands. Furthermore, the helicase can keep on translocating along the ssDNA, while the loop reforms behind it. The present study is only focused on the analysis of the forward unwinding part of the traces of the RecQ unwinding DNA process.

The experiments are performed at forces limited by:

- The unzipping force: at forces above $\sim$ 15pN, the hairpin spontaneously unzips, so no dsDNA is present to test the helicase unwinding activity. For this reason 12pN is the maximum force used in this set of experiments[1].

- Hairpin reformation at ssDNA at low forces: for $f < 5$pN, the hairpin reforms, even in the presence of the helicase because the flexible ssDNA wraps around the enzyme at low forces inducing hairpin rezipping (as has been previously described [116]).

Due to the above mentioned limitations, the experiments were performed in the range of forces $5 - 12$pN and $T = 25°$C. The buffer used contained 20mM Tris, 25mM NaCl, 3mM $MgCl_2$ and 100µM ATP.

As previously discussed, the ssDNA molecular extension depends on the force. In order to measure the helicase unwinding rates (in bp/s) or the helicase processivity (average number of bases unwound before dissociating), the measured extension at each force needs to be converted to number of unwound basepairs, Δn_{bp}, as:

$$\Delta n_{bp}(f,t) = \frac{\Delta x(t)}{2x_{WLC}(f)}, \tag{8.1}$$

---

1 The force could be increased up to $\sim$ 15pN by changing the sequence. However, sequence effects may arise from strongly varying its GC content.

where $x_{WLC}(f)$ is the obtained extension for a single base (hence the factor of 2, in order to obtain basepairs), using the WLC model at a given force, f. Note that, at constant force, the change in molecular extension is equivalent to the trap-displacement, i.e. $\Delta x = \Delta \lambda$.

The elastic parameters that are used to model the ssDNA behaviour are $l = 0.70$nm and $p = 0.75$nm, which correspond to the the ones reported in [61], and agree with the maximum distance change corresponding to the total number of basepairs in the hairpin to unwound, $\Delta x_{max} \approx 430$bps (see Fig. 8.2 right axis).
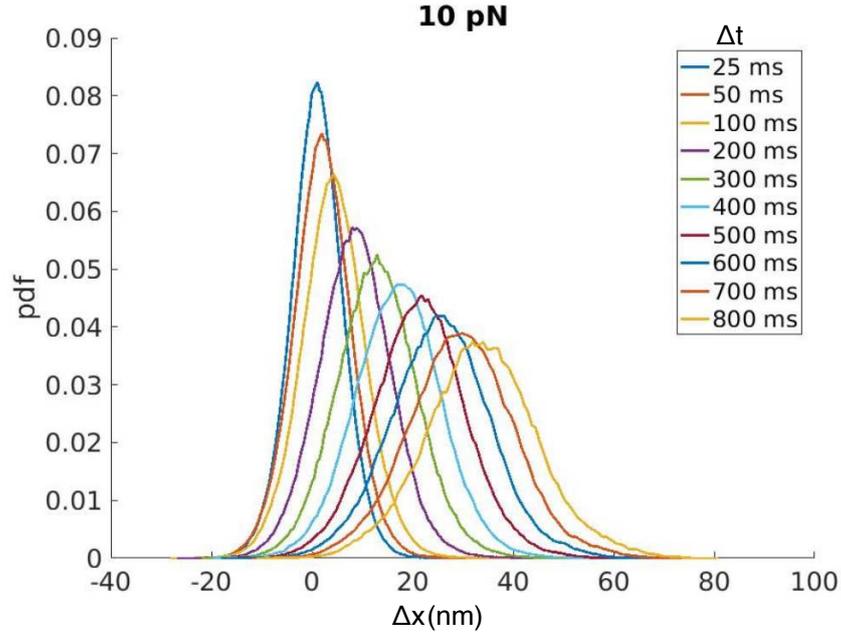
Figure 8.3: **Probability distributions of distance at 10pN**. The probability distributions of the helicase moving a distance Δx in a given time interval (Δt, shown in the legend). The error bars are given by the standard deviation of the velocity (Fig. 8.4). Figure adapted from [115].

From the experimental forces we compute two magnitudes: the change in the molecular extension in a time interval Δt, and the helicase unwinding rate given by $\Delta n_{bp}(\Delta t)/\Delta t$. For the first magnitude we compute the probability density functions (pdf) of Δx, for different Δt. As shown in Fig. 8.3, the obtained distributions are Gaussian, regardless of Δt, with larger standard deviations for longer Δt. For the latter, we compute the average helicase dsDNA unwinding rate as a function of the force. As shown in Fig. 8.4, the rate increases from ~ 20nm/s at 5pN to ~ 50nm/s at 12pN. However, this is mainly a consequence of the elastic properties of ssDNA (larger forces are translated to longer extensions). Indeed, the unwinding rate in bp/s remains practically constant around ~ 50bp/s, independently of the applied force, which shows that RecQ is an active helicase, as previously shown [117, 118].
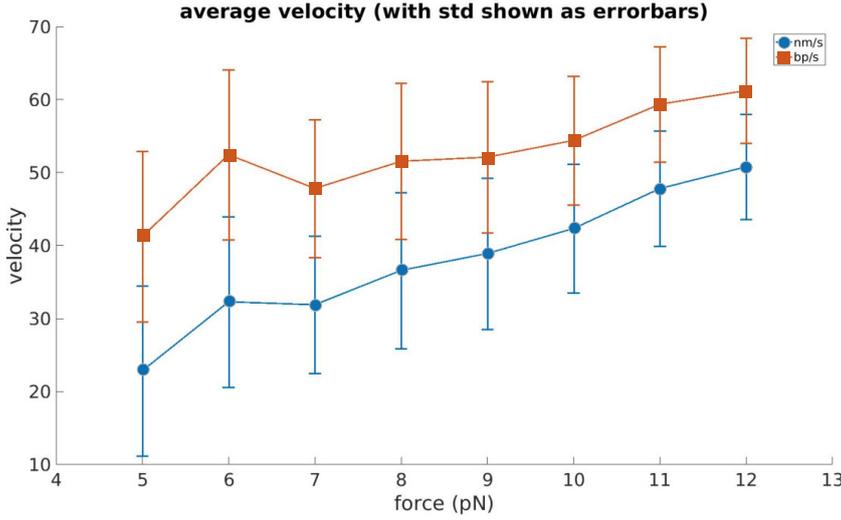
Figure 8.4: **Measured velocity as a function of the force**. The blue line is the measured velocity of the trap position, in nm/s, while the orange line corresponds to its conversion to bp/s using Eq. 8.1 in bp per second. Figure adapted from [115].

8.2 ENTROPY PRODUCTION

The conservation of energy in our system can be written as:

$$\Delta G_t^{ATP} + W_t^{OT} = \Delta G_t^{DNA} + Q, \tag{8.2}$$

where the first term accounts for the free energy of ATP hydrolisis. After a time interval t, the ATP hydrolisis provides a total energy:

$$\Delta G_t^{ATP} = n_{ATP}\Delta n_{bp}\Delta\mu, \tag{8.3}$$

where Δn_{bp} is the total number of basepairs opened in time t, n_{ATP} is the number of ATP hydrolized per basepair and $\Delta\mu$ is the chemical potential difference when converting one ATP to an ADP and an orthophosphate. Relation 8.3 assumes a tight coupling between chemical energy and the mechanical work. Typical values for $\Delta\mu$ are $13 - 20k_\mathrm{B}\mathrm{T}$ [119]. On the other hand, several studies have predicted a $n_{ATP} = 1$ for RecQ [120–122].

The second term in Eq. 8.2 accounts for the work exerted by optical tweezers:

$$W_t^{OT} = f\Delta x \tag{8.4}$$

were Δx is the trap displacement.

The third term in Eq. 8.2 accounts for the energy associated to the helicase unwinding that includes the bond breaking and strand stretching as:

$$\Delta G_t^{DNA} = \Delta G_{bp} \Delta n_{bp} - \int_0^f x_{WLC} =$$
$$= \Delta G_{bp} \Delta n_{bp} - f x_{WLC} + \int_0^{x(t)} f_{WLC}(x') dx', \tag{8.5}$$

where ΔG_{bp} is the free energy change due to opening a basepair, and the last two terms correspond to the work of stretching a chain of $2\Delta n_{bp}$ bases, under controlled force (using the WLC model introduced in Chapter 4).

Finally, Q accounts for the dissipated heat to the thermal bath. From Eqs. 8.2-8.5 we can compute the entropy production, Q/T, as:

$$\Delta S = \frac{\Delta x}{2 T x_{WLC}} (n_{ATP} \Delta \mu - \Delta G_{bp}) +$$
$$+ \frac{1}{T} \left(f \Delta x - \int_{x_0}^{x_t} f_{WLC}(x') dx' \right), \tag{8.6}$$

By substituting the expression of the free energy of stretching a WLC chain, with a persistence length $p = 0.75$nm, and a contour length per base, $l = 0.70$nm, we can write Eq. 8.6 as:

$$\Delta S = A \Delta x, \tag{8.7}$$

where the slope A is defined as

$$A = \frac{1}{T} \left[\frac{n_{ATP} \Delta \mu - \Delta G_{bp}}{2 x_{WLC}} + f - \right.$$
$$\left. - \frac{k_B T}{p} \left(\frac{l^2}{8 x_{WLC} (l - 2 x_{WLC})} - \frac{1 + \frac{l}{2 x_{WLC}}}{4} + \frac{x_{WLC}}{l} \right) \right]. \tag{8.8}$$

8.2.1 Fluctuation theorem

Starting with the assumption that the entropy production obeys the standard fluctuation theorem [123]

$$\frac{p(\Delta S)}{p(-\Delta S)} = e^{\Delta S / k_B} \tag{8.9}$$

and that $p(\Delta S)$ and $p(-\Delta S)$ only vary in time due to variations in the extension Δx, which is the only variable that depends on the time, follows that:

$$\frac{p(\Delta x)}{p(-\Delta x)} = e^{A \Delta x / k_B}. \tag{8.10}$$

With A defined in Eq. 8.8. For non-equilibrium steady states (NESS), the fluctuation theorem takes the form [124]

$$\lim_{t\to\infty} \frac{1}{\langle\Delta x\rangle} \log\left(\frac{p(\Delta x/\langle\Delta x\rangle)}{p(-\Delta x/\langle\Delta x\rangle)}\right) = A\frac{\Delta x}{\langle\Delta x\rangle} \quad (8.11)$$

Combining this equation and Eq. 8.8 allow us to obtain the average number of basepairs opened per ATP hydrolized, n_{ATP}. In our case, the obtained value at $f = 12$, is $n_{ATP} \sim 1/14\text{ATP/bp}$, i.e. the helicase unwinds 14bp with the the hydrolisis of a single ATP. This is unconsistent with ATP hydrolisis providing about $\sim 20k_\mathrm{B}T$ per ATP hydrolized: considering that $\Delta G_{bp} \sim 2k_\mathrm{B}T$, the maximum number of unwound bps per ATP consumed would be ~ 10 basepairs (assuming 100% efficiency). Moreover there is strong evidence in the literature for $n_{ATP} \sim 1\text{ATP/bp}$ for RecQ helicase.

This unconsistency can be due to the wrong assumption of tight coupling, which implies that the existence of other reaction coordinates (such as other conformational changes) have not been considered. Typically, the activity of molecular motors, such as RecQ, requires of successive, set of conformational changes. In order to account for it, a hidden entropy source could be added. However, this new term is no longer proportional to Δx, and hence the above proposed model is no longer valid in order to analyze the data.

8.3 SIMULATIONS

In order to better understand the discrepancy between our results for n_{ATP} and the values reported in the literature, giving $n_{ATP} \approx 1\text{ATP/bp}$ [120–122], we also performed simulations based on a brownian ratchet model [125, 126]. The model considers the DNA as a discrete chain, with the position of the helicase tightly coupled with that of the DNA fork. This is a simplification of more sofisticated models such as the Betterton-Jülicher model [127, 128], that consider two coupled dynamics: one for the fork and the other for the helicase.

The free energy difference of a helicase unwinding a single basepair, after hydrolizing n_{ATP} ATP, and releasing 2 ssDNA bases, can be written as

$$\Delta G = n_{ATP}\Delta\mu_{ATP} - \Delta G_{bp} + 2fx_{WLC} - 2\int_0^ f_{WLC}(x')dx', \quad (8.12)$$

where the first term accounts for the hydrolized ATP, the second term considers the free energy of basepairing and the third and fourth terms correspond to the free energy of stretching the 2 ssDNA bases unwound after the helicase-DNA fork advances one bp.

The probabilities of the helicase of moving forward (F), unzipping a basepair of the dsDNA chain, P_F and that of moving backwards (B), allowing for the rezipping of a basepair in the chain, P_B, during a time interval dt, are given by

$$P_F = k_0 e^{\frac{\Delta G}{k_\mathrm{B}T}} dt \quad (8.13)$$

and

$$P_B = k_0 e^{\frac{-\Delta G}{k_\mathrm{B}T}} dt. \quad (8.14)$$

Here we perform a Monte-Carlo simulation of the proposed model. We simulate a large number of traces ($N = 1000$), for each one of the forces used in the experiments described in Sec.8.1 ($5 - -12$pN). The elasticity of ssDNA is modelled as a WLC with $p = 0.75$nm and $l = 0.69$nm [61]. The following values for the parameters are considered for all the simulations presented: $\Delta\mu = 20k_{\mathrm{B}}T$ ([119]), $\Delta G_{bp} = 2k_{\mathrm{B}}T$ (as a reasonable average for the free energy reported in [82, 91]), $n_{ATP} = 1$ATP/bp ([120–122]), $k_0 = 10^6\mathrm{s}^{-1}$ ([129]), $dt = 0.5/3k_0$.
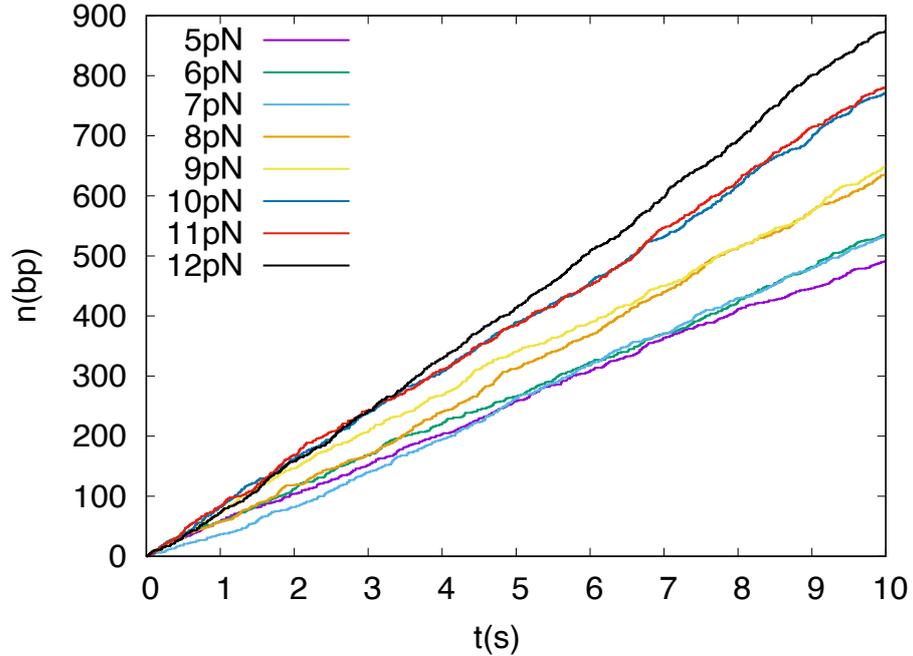
Figure 8.5: **Simulated traces for forces ranging 5-12pN**. In order to obtain the averaged values, $N = 1000$ traces for each force have been obtained.

Typical traces are shown in Fig. 8.5, which are remarkably symilar to the experimental ones (Fig. 8.5, red square, forward unwinding part). Then, we performed the same analysis described in Sec. 8.2. First, we compute the probability density function of the helicase displacement (in bp), considering various time intervals, as shown in Fig. 8.6. There is a qualitative agreement between the experiments and the obtained results: the distributions are all Gaussian, with standard deviation that increases linearly with time. Furthermore, for longer time intervals, the distance the helicase translocates increases proportionally to the velocities shown in the inset of Fig. 8.6. These features are characteristic of an advection-diffusion process

With the obtained traces, we can also compute the velocities by linear fitting each one of the traces. In the inset of Fig. 8.6, the obtained average values are shown. A trend of increasing speed for higher forces is observed, contrary to what is observed in our experiments. This is due to the fact that we have first simulated a Brownian ratchet with a free energy difference that is force-dependent, which translates to a force-dependent velocity.
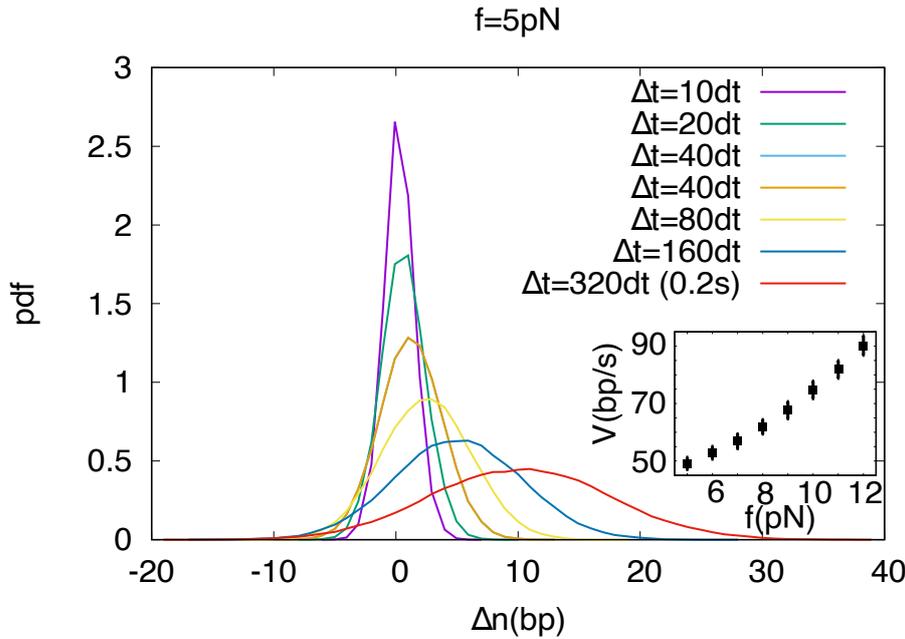
Figure 8.6: **Probability distributions of the helicase of moving** Δx **basepairs in a time interval** Δt, under a constant force of $f = 5$pN. The standard deviation of the Gaussians increases with the time window, Δt. The mean of the Gaussian distribution moves at a velocity of ~ 50bp/s, in agreement with the inset. (Inset) Averaged velocity at forces 5-12pN. Error bars are standard errors.

Finally, we can check the NESS fluctuation theorem for the simulated data, by plotting the left hand side of Eq. 8.11 vs $\langle \Delta x \rangle$, as shown in Fig. 8.7.

The long time limit, $t \to \infty$, corresponds to low values of $\Delta x / \langle \Delta x \rangle$, data for different Δt collapse. The linear fit returns a slope of $A = 0.39 k_\mathrm{B} T / nm$, which corresponds to the exact value from Eq. 8.8, after substituting the values for the elastic properties of ssDNA and the energetics used in the simulation.

8.4   CONCLUSIONS

By using optical tweezers we have investigated the dsDNA unwinding activity of the RecQ helicase on a DNA hairpin substrate. From the experimental data we have measured the unwinding rate whish is about ~ 50bp/s, independently of the applied force, in agreement with the literature [117, 118]. We have also computed the distribution of the helicase displacement after a time interval Δt, which follows a Gaussian distribution. The variance increases linearly with time, showing that RecQ behaves following an advection-dissusion process.

From analyzing the energetics of the whole system (helicase+DNA+trap), we have obtained an expression for the entropy production, which is a function of the average number of hydrolized ATP molecules per unwound basepair, n_{ATP}. The n_{ATP} is an important parameter that is
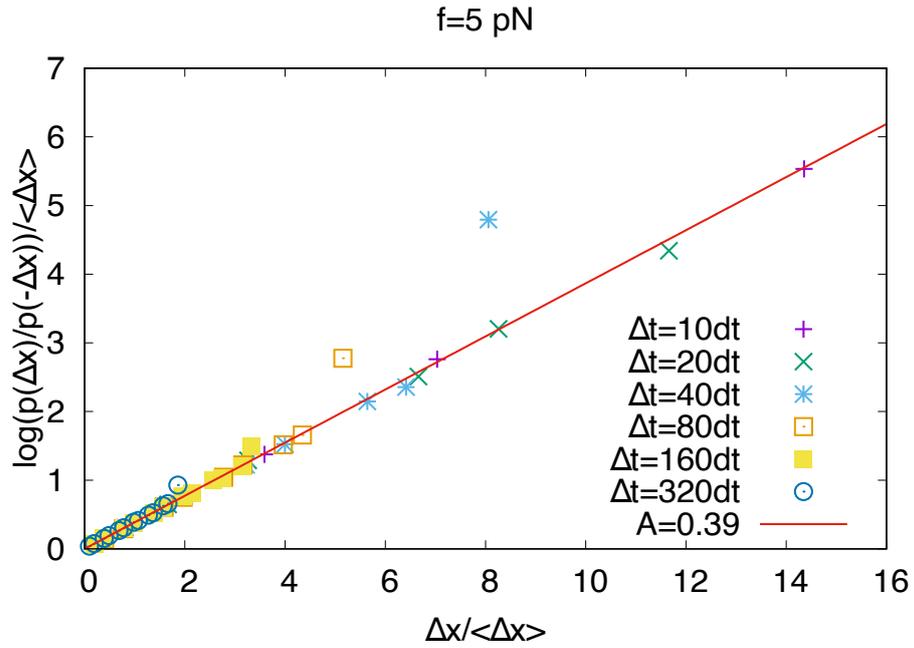
Figure 8.7: **Fluctuation theorem, at** $f = 5\textbf{pN}$. The fitted value for the slope, A, is in agreement with the theoretical one obtained from Eq. **??** with the imposed values in the simulations.

related with the motor efficiency. By assuming a tight coupling between the chemical and mechanical energies and fitting the entropy production to experimental data we obtain $n_{ATP} \approx 14\text{bp/ATP}$, which is much larger than that reported in the literature [120–122]. The obtained value is even larger than that predicted for a 100% efficiency motor.

We have also implemented Monte-Carlo simulations based on a Brownian ratchet model, aiming for a deeper understanding of the physics of the RecQ experiments. The simulations are based on very simple assumptions, but qualitatively reproduce main characteristics of the experimental results.

As expected, applying the entropy production analysis into the simulated model allows to recover the correct value for n_{ATP}. Further work regarding the simulations would be adding several features such as the dynamic of the DNA fork and the helicase (using a Betterton-Jülicher model), the optical trap or the force feedback or the sequence heterogeneity. We expect them to help us to ellucidate which factors affect the estimation of n_{ATP}.

DETECTION OF SINGLE DNA MISMATCHES

Defects such as base mismatches, missing bases, crosslinks and so on, occur in DNA with high frequency, and must be efficiently identified and repaired to avoid dire consequences such as genetic mutations. Identification of defective DNA structures is a difficult task, since small differences in base-pair bonding are hidden in the local structural variability of a generally random base-pair sequence. Here we focus on the detection of base mismatches, that is local deviations from the ideal Watson-Crick pairing rule, which may typically originate from DNA replication process, foreign chemical attack, or ionizing radiation. We introduce different mismatches in short DNA hairpins (10 or 20 base pairs plus a 4-base loop) sandwiched between dsDNA handles, to be used in single-molecule force spectroscopy with optical tweezers. Experimental detection of a mismatch defect demands the ability to measure slight deviations in the free energy difference between the folded molecule (hairpin) and the unfolded (ssDNA). For this reason, the correct characterization of the elasticity of ssDNA performed along the thesis is critical in detecting DNA mismatches. We perform both hopping and force-pulling experiments to measure the excess free energies and deduce the characteristic kinetic signatures of the mismatch from the force-distance curves. All-atom molecular dynamics simulations lend support to the detailed interpretation of the experimental data. Such measurements, at the lowest sensitivity limits of this experimental technique, demonstrate the capability of identifying the presence of mismatches in a random complementary dsDNA sequence, and provide lower bounds for the ability to distinguish different structural defects.

The work presneted in this chapter has been performed in collaboration with Fabio Landuzzi, from the Department of Physics of the University of Lille, and published recently [41].

9.1 DNA MISMATCHES

A DNA mismatch (MM) is a structural defect occurring when two non-complementary bases are aligned in a sequence of duplex DNA [130]. A MM is defined as a *transduction* when formed by non-complementary purine-pyrimidine (pur-pyr) bases, and *transversion* in the case of pur-pur or pyr-pyr pairs. Compared to DNA strands with the canonical (Watson-Crick) pairing rules, MMs are expected to produce alterations in the structure and stability of the DNA helix, especially in the proximity of the MM site [131–133]. MMs can appear during replication of DNA [134], heteroduplex formation [135], as well as by action of mutagenic chemicals, ionizing radiation, or spontaneous deamination [136]. MMs are efficiently corrected in DNA by mismatch repair (MMR) proteins.
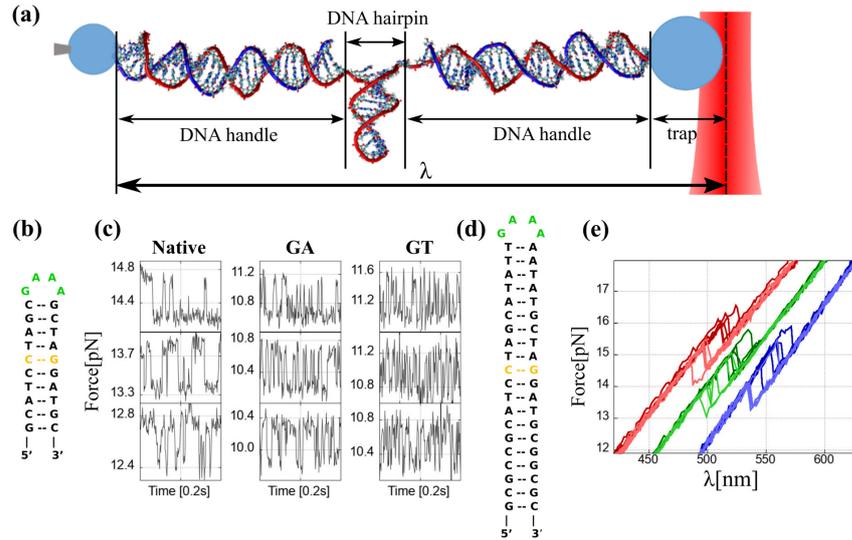
Figure 9.1: **Experimental setup and DNA sequences.** (a) Schematic representation of the LOT experimental setup. A DNA hairpin is attached between two micron-sized beads via dsDNA handles. One bead is held by air suction in the tip of a glass micropippette while the other is captured in the optical trap. λ stands as the distance between the tip of the pipette and the center of the optical trap. (b)Sequence of the 10bp hairpins with the 4-loop (green) and the single mismatch location (yellow). (c) Experimental force-time traces for the 10pb hairpins hopping experiments.(d) Sequence of the 20bp hairpins with the 4-loop (green) and the single mismatch location (yellow). (e) Several FDC cycles for native (red), G-A mismatch (green), G-T mismatch (blue) 20bp hairpins. In dark (light) colours are shown the stretching (releasing) curves.

Failures in detecting or correcting the lesion give rise to genetic mutations [136, 137] in fact, MMs have been associated with 10-30% of spontaneous cancers in various tissues [137, 138]. In particular, G-A and G-T defects are of great interest to the cancer biology community, since such type of MM can be formed efficiently during oxidative stress, both by endogenous processes and following chemo- or radiotherapy. A prominent example is the so-called 8-oxoG lesion [139], which differs from ordinary guanine in that a H atom is replaced by an O at the C8 position, and the N7 nitrogen becomes protonated. Such defect can be formed by several reactive oxygen species (ROS) that are able to attack the guanine. In the subsequent replication stage, the 8-oxoG displays a higher affinity for adenine than cytosine, thereby leading to a 8-oxoG-A mismatch and, ultimately, to a G-C→T-A mutation. Another common mismatch is the G-T formed with high probability by polymerases such as β and Taq, during DNA replication, thus being about a thousand fold more frequent than other MMs; this is mainly due to its strong thermodynamic stability, which makes its identification by repair enzymes quite difficult [140].

Powerful approaches are available to detect specific locations of DNA sequence variation, such as rolling circle amplification [141], molecular inversion probes [142], and MM ligation with bioluminescence detection

[143], all based on DNA ligase or nuclease activities; Taqman probes, molecular beacons, and related assays have also been used to detect specific targeted alleles [144, 145]. All of these methods require *a priori* knowledge of the reference sequence, and involve the construction of sophisticated probes. By contrast, MM-detection assays rely on the base-pairing quality of DNA, and subsequent enzymatic detection of mispaired bases [146, 147], and are thus independent of the exact identity of the underlying mutation. All such methods are suitable for high-throughput screening of unknown sequences, and work on large populations of identical DNA sequences, properly amplified by molecular engineering techniques. On the other hand, the detailed molecular structure modifications of individual DNA defects have been assessed by nuclear magnetic resonance (NMR), and x-ray diffraction [148–151] most notably for the typical G-T mismatch. However, the thermodynamic and kinetic changes in the stability of the DNA structure induced by such defects remain inaccessible to such techniques. Such details are indeed crucial to understand the extremely selective ability of repair enzymes (glycosylases), which are capable of rapidly scanning the DNA sequence at rates of 20-30 bp/s [152], while applying small forces to the backbone in a sort of highly dynamical "interrogation" process aimed at detecting differences of just a few kcal/mol [153].

In recent years many efforts have been devoted to experimentally characterize the biomolecular free-energy landscape of folding [154]. Dynamic force spectroscopy experiments are well-suited to study the folding/unfolding transitions of one molecule at a time with high spatial and temporal resolution [11, 24, 51]. From the experimentally measured force-dependent unfolding and folding kinetic rates it is possible to characterize the position of the transition states and the height of the corresponding kinetic barriers [14]. Notably, force spectroscopy by laser optical tweezers (LOT) has been successfully applied also to measure short DNA hairpin sequences, thereby allowing to detect elastic properties of ssDNA [63], and transition states during the repeated folding/unfolding pullings [155]. In a recent study [40], LOT were used to characterize four different MMs, by positioning pairs of identical defects at a close distance in 29-bp DNA hairpins with a 6-bases loop. AFM-based force spectroscopy studies have demonstrated the possibility of detecting single DNA mismatches by measuring irreversible rupture force distributions in oligonucleotide microarrays [156] and DNA origami platforms [157]. However the question remains whether mismatch free energy differences in a DNA duplex can be determined in pulling assays.

Here, we use force spectroscopy with LOT to demonstrate the capability to measure the excess free energy of individual MMs in DNA at the single molecule level. To this end, we have synthesized very short DNA hairpins, of 10 and 20 bp stem length, containing single G-A or G-T mismatches. Hairpins are tethered between two double-stranded (ds) DNA handles of 29 bp each. We investigated two different stem lengths to ensure a sufficiently high signal-to-noise ratio measurements to derive the excess free energy in all constructs. We find that the identification of single mismatches in pulling assays is facilitated using

different protocols depending on the length of the hairpin. While for hairpins shorter than 10-20bp equilibrium hopping experiments are more convenient, above 20bp nonequilibrium pulling experiments give a better signal. The study of two constructs with two different protocols (hopping versus pulling) allows us to monitor the internal dynamics of the native (i.e the canonical hairpin with Watson-Crick bp only) and defective molecules. It also permits to test the robustness of the measured excess energy values. Our results demonstrate the ability of the method to clearly detect the presence of single MM defects in the DNA sequence; defect energetics and dynamics can be qualitatively and quantitatively characterized within the lower limits of the experimental resolution. A set of Molecular Dynamics simulations supports the interpretation of the experimental data, and provides useful indications about the molecular details of the unfolding process, over a finer time scale that escapes the experimental resolution. The experimental method combined with numerical simulations offers a promising route to the characterization of defects in DNA sequences, and their interaction with damage-signaling and repair proteins.

9.2   MATERIALS AND METHODS

9.2.1   *DNA substrates*

The DNA constructs studied consist of a dsDNA segment (called stem) of either 10 or 20bp, inserted between two 29bp segments called handles. The two strands of the stem are connected via a loop of 4 bases. The sequences of the stem and loop region are shown in Fig.9.1b and 9.1d. Each handle is tailed either with biotin and digoxigenin in order to attach it specifically to streptavidin- or antidigoxigenin-coated beads. The synthesis of the hairpins is described in detail in Appendix A. Shortly, an oligo with the stem sequence (10 or 20bp), sided by 29b (corresponding to the handles) on each side (Sigma-Aldrich) is digoxigenin-tailed and annealed to the handle-region complementary 29b oligo. All our experiments are performed in TE buffer (Tris 10mM and 1mM EDTA) pH 7.5, 1 M NaCl at room temperature (298 K).

9.2.2   *Hopping experiments*

In hopping experiments, the pipette and optical trap relative distance, λ, is kept fixed, while the hairpin executes transitions between the folded and unfolded states. An example of experimental force-traces for hopping experiments is shown in Fig.9.1c. We monitored the force at 1kHz for 15-40s, for a time sufficient to register at least 20 conformational transition events and then moved slightly the distance λ and repeated the measurement. Notably, once λ is fixed, the force in the unfolded state is lower than that in the folded state, because the bead experiences a much larger displacement from the center of the optical trap when the hairpin is formed. As a consequence, if such a force jump between the

two states is larger than the thermal noise, monitoring the instantaneous force is equivalent to collecting information on the state (folded vs. unfolded) of the hairpin. By working at the vicinity of f_c, at which the two states have the same occupation probability, the kinetic rates of unfolding (k^+) and folding (k^-) fall in a timescale that allows to observe several hopping events. Each variation of the trap position, λ, produces a consequent variation of the probability to observe the system in the folded (w_N) or unfolded (w_U) state. Upon repeating the measurements at different λ's, we could measure the relative variation of the occupation probability of the two states. Typical time-series and probability histogram data for 10bp hairpins, with native sequence or including a GA and GT mismatch, are presented in Fig.9.2a and as multi-panel figures in the Appendix.

9.2.3 *Pulling experiments*

In a ramping protocol the trap position is moved at a constant speed, while force is measured. Typical force-distance curves (FDC) are shown in Fig. 9.1e (more detailed plot for the mismatch GA,GT and different pulling velocities are reported in Appendix). At low forces hairpins are in the folded state, with the stem forming a double helix; whereas at large forces they unfold to a stretched conformation, where the stem is found as ssDNA [158]. Transitions between both states are viewed in the FDC as a sudden jump in force. Forces at which such transitions take place change upon repeating the same experiment due to thermal fluctuations. Typically, two force branches are observed: the upper force branch shows the elastic response of the whole molecular construct when the hairpin is folded, whereas the lower force branch shows the response for the unfolded hairpin. By analyzing the folding/unfolding trajectories it is possible to detect the force value at which the hairpin unfolds/refolds for the first time along the trajectory; this is called the first-rupture force, and appears (see Fig.9.1d and Appendix) as a force rip along the FDC. For the 10bp hairpins, the rupture force distribution is not clearly detectable since the kinetic rates are comparable with the recording data frequency and the force fluctuations are comparable to the force jump, as shown in the trajectories of Appendix. For this reason, only the 20bp hairpin pulling data (shown in Appendix) has been used. For each sequence (native, GA and GT mismatches) we averaged measurements on at least three molecules at 100nm/s. Force-pulling experiments typically covered 50-100 cycles.

9.2.4 *Kinetic rate theory*

Information about the molecular free energy landscape (mFEL) can be obtained by analyzing the data using kinetic models, such as the Bell-Evans (BE) model [159, 160]. The BE model predicts an exponential dependence with force of the folding and unfolding kinetic rates:
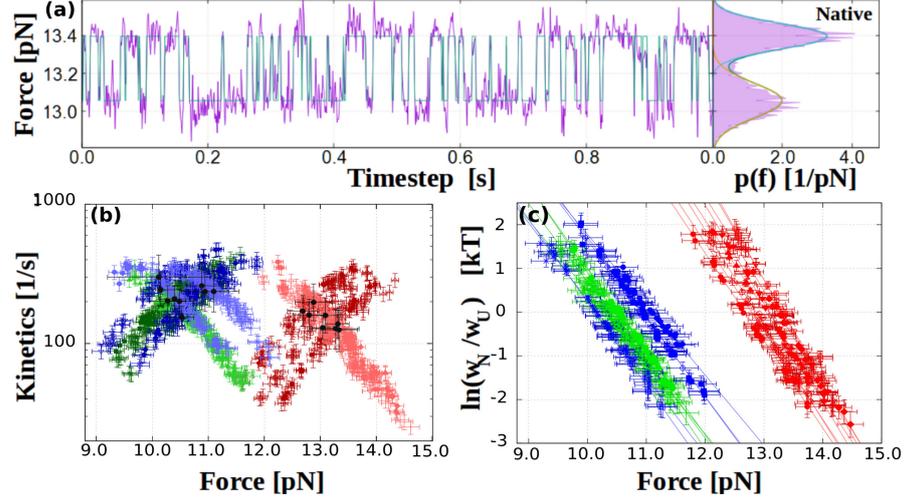
Figure 9.2: **10bp hairpin hopping results.** (a) Example of a hopping trace of the 10bp native hairpin (left panel), the green curve is the optimal trajectory obtained from the HMM. Force distributions (right panel), together with the best-fit from the HHM (blue curve) and the two-states Gaussian fit from Eq.(9.2) (green and yellow curves for each Gaussian component). (b) Unfolding (dark color) and refolding (light color) kinetic rates measured for the native (red), G-A mismatch (green) and G-T mismatch (blue) hairpins. The validity of the BE approximation is shown as a straight line in the semi-log scale. (c) Plot of $\ln(w_N/w_U)$ as a function of force from Eq.(9.3) for the three hairpins. Color code as in (b).

$$
\begin{aligned}
k^+(f) &= k_0 \exp[-\beta(B_{TS} - f \cdot x_{N \to TS})] = \\
&\quad k_m \exp[\beta f \cdot x_{N \to TS}], \\
k^-(f) &= k_0 \exp[-\beta(B_{TS} - \Delta G_{NU} + f \cdot x_{TS \to U})] = \\
&\quad k_m \exp[-\beta(f \cdot x_{TS \to U} - \Delta G_{NU})], \quad (9.1)
\end{aligned}
$$

where ΔG_{NU} is the free-energy difference between N and U extrapolated to zero force, and $x_{N \to TS}$ ($x_{U \to TS}$) are the distances between the x_{TS} and the folded (unfolded) state that are taken as force independent. B_{TS} is the kinetic barrier extrapolated to zero force and has been included in the prefactor $k_m = k_0 \exp(-\beta B_{TS})$. At the coexistence force f_c , $k^+(f_c) = k^-(f_c)$, and therefore $\Delta G_{NU} = f_c(x_{N \to TS} + x_{TS \to U}) = f_c x_{NU}$ with x_{NU} the released molecular extension at f_c. Bell-Evans kinetic rates are a valid approximation close to f_c, deviations are expected at sufficiently low or high forces [161]. From these phenomenological equations we could fit the kinetic rates obtained by the analysis of the pulling and hopping data with the force and extract the free-energy of formation, the molecular extension and the kinetic rate at the coexistence force. The BE approximation is valid in a small range of forces not too far from the coexistence force, f_c.

Table 9.1: **10bp results.**Parameters obtained from the hopping experiments for the 10 bp hairpins from the Bell-Evans model and compared with the theoretical values obtained from the nearest-neighbour model [132] with the contribution of the tetraloop (see Appendix). The theoretical f_c is obtained from Eq.9.3 imposing $\Delta G_{NU}(f_c) = 0$.

Hairpin	f_c	x_{NU}	$x_{N \to TS}$	$x_{TS \to U}$	ΔG_{NU}	ΔG_{NU}^0
	(pN)	(nm)	(nm)	(nm)	($k_{\mathrm{B}}T$)	($k_{\mathrm{B}}T$)
10bp nat.	13.0(4)	8.7(5)	4.1(4)	4.6(3)	27(4)	21(2)
Th. nat.	14.7	8.8	/	/	24.2	
10bp GA	10.4(3)	7.1(7)	3.4(2)	3.8(6)	18(2)	16(2)
Th. GA	11.8	8.4	/	/	18.1	
10bp GT	10.6(6)	6.1(5)	3.0(5)	3.1(4)	16(3)	16(2)
Th. GT	11.6	8.5	/	/	17.6	

9.2.5 *Molecular dynamics simulations*

For the molecular dynamics (MD) simulations we used the GROMACS 5.1 computer code [162, 163]. The hairpin plus handles molecular construct (see Fig.9.1a) was built as a continuous ssDNA chain spanning from the $5'$ to the $3'$ ends: the first and last groups of 29 bases were matched to two complementary 29-long ssDNA strands, to make up the two dsDNA handles; the central $10 + 4 + 10$ bases represented the hairpin, perfectly folded in the initial configuration. The end-to-end distance between the C1$'$ atoms of the first and last bp (to be used as reference length in the foregoing) is λ_0=23 nm. The structure of hairpin plus handles, with the same native base sequence used in the experiments, was assembled in a water box of size $50{\times}9{\times}12$ nm^3 with periodic boundary conditions in the three directions, containing about 174,000 TIP3P water molecules, plus 625 Na$^+$ and 488 Cl$^-$ ions, to ensure neutralization of the phosphate backbone charge, and physiological salt concentration around 0.15 M.

Equilibrium MD simulations were carried out at temperatures ranging from 300 to 360 K and pressure of 1 atm, at constant-$\{NVT\}$. Coulomb forces were summed by shifted particle-mesh Ewald electrostatics, with real space cut-off set at 1 nm; long-range dispersion forces were also cut-off at 1 nm. We used rigid bonds for the water molecules, which allowed to push the time step to 1 fs for both the thermal equilibration runs, and the force-pulling simulations. Typical preparatory constant-$\{NPT\}$ MD runs lasted between 10 and 20 ns; force-pulling simulations were carried out for 50 ns; thermal equilibrium simulations at constant-$\{NVT\}$ lasted typically 100 ns.

Steered molecular dynamics (SMD) were performed on the fragments with the constant-force pull code available in GROMACS. We applied a constant displacement/force parallel to the direction x by means of a harmonic-spring fictitious potential attached to the center of mass of the last base pair of the DNA handles. After some tests, the spring constants were set at 100 and 75 kJ mol^{-1} nm^{-2}, which is $\sim$ 1000 times higher than typical values of experiments performed with OT. Pulling

speeds of about 20 cm/s were used for most SMD simulations. Forces and displacements were recorded at intervals of 10 time steps.

We also consider the possibility to study this system using a coarse-grained model [164, 165]. We have performed simulation tests using the oxDNA2 model with sequence dependent stacking interaction implemented in LAMMPS [166]. This model has been used by other researchers to estimate the FEL of a short hairpin in melting [167]. Despite the increased simulation time (anyway not sufficient to obtain a proper kinetics characterization), these models have been tested only for the native sequence and do not account for relevant structural differences between purine and pyrimidine for mismatched basepairs. Moreover the implicit solvent do not completely account for the hydrodynamic effects during the pulling protocol. It is for these reasons that we decide to focus on the all-atoms model.

9.3 FREE ENERGY MEASUREMENTS FROM HOPPING EXPERIMENTS IN 10BP DNA HAIRPINS

On the left of Fig.9.2a it is shown an example of an experimental hopping trace for the 10 bp native hairpin, while on the right it is shown the histogram of the force distribution that can be well fitted to a sum of two Gaussians,

$$
P_\lambda(f) = \frac{w_{\lambda,U}}{\sqrt{2\pi\sigma_{\lambda,U}^2}} e^{-\left(\frac{f-\langle f_{\lambda,U}\rangle}{4\sigma_{\lambda,U}}\right)^2} + \frac{w_{\lambda,N}}{\sqrt{2\pi\sigma_{\lambda,N}^2}} e^{-\left(\frac{f-\langle f_{\lambda,N}\rangle}{4\sigma_{\lambda,N}}\right)^2} \tag{9.2}
$$

where $\langle f_{\lambda,U}\rangle, \langle f_{\lambda,N}\rangle$ represent the average force in the unfolded and folded state, and $\sigma_{\lambda,U}, \sigma_{\lambda,N}$ are the standard deviations of the force. A more accurate statistical analysis of the hopping traces based on Hidden Markov Models reveals only two underlying molecular conformations N and U (a detailed explanation of Hidden Markov Models is found in the Appendix).

It must be noted that for any fixed value of λ, the two kinetic rates in Eq.(9.1) are evaluated at different force values: namely, $k^+(f)$ $(k^-(f))$ is measured at force f_N (f_U) where the molecule hops from N to U (U to N). Therefore, we can separately plot $k^+(f_N)$ and $k^-(f_U)$, with λ acting as a parameter. These rates measured in the passive mode where the force is not kept constant are called apparent rates [81]. These data, plotted in Fig. 9.2b, mark two important points: first, for each type of hairpin the coexistence force, f_c, is readily identified by the value at which $k^+ = k^-$; second, the exponential dependence of the rates on the force indicates that the BE model [159, 168] should be applicable in this case.

Moreover, the molecule is found to follow exponential kinetics [14]. The probability distribution of the residence times in the folded/unfolded conformations is well described by an exponential function whose width is equal to the inverse of the unfolding/folding kinetic rate. The dependence
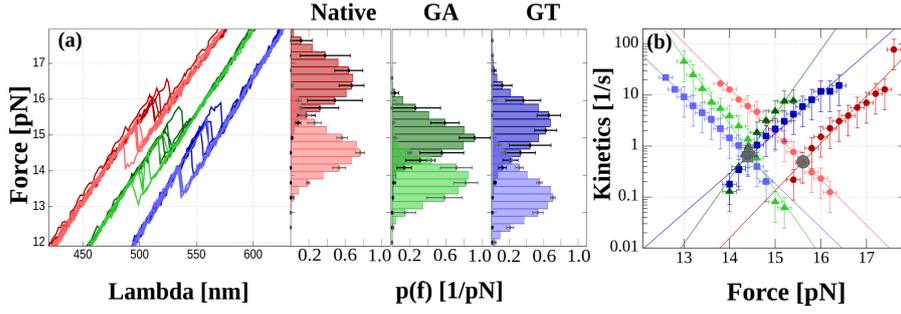
Figure 9.3: **20bp hairpin pulling results.** (a) Example of folding/unfolding pulling curves for native (red), G-A mismatch (green) and G-T mismatch (blue) 20bp hairpins. (b) Histograms of unfolding (dark colors) and folding forces (light colors). (c) Average kinetic rate as a function of the applied force. The black symbols indicate the coexistence force. Color code as in (a).

of these rates on the applied force provides accurate information about the height and position of the kinetic barrier.

From Eq.(9.1) we can express the ratio between the kinetic rates in the form

$$k_B T \log\left(\frac{k^+(f)}{k^-(f)}\right) = k_B T \log\left(\frac{w_U(f)}{w_N(f)}\right) =$$
$$-\Delta G_{NU} + f \cdot (x_{N \to TS} + x_{TS \to U}) = (f - f_c) \cdot x_{NU} \qquad (9.3)$$

with $\Delta G_{NU} = f_c x_{NU}$ the free energy difference extrapolated to zero force. In Fig. 9.2c it is shown the $\ln(w_N/w_U)$ versus force for all the 10-bp hairpins studied.

We can also estimate the free energy between N and U at zero force using the following formula (see Appendix),

$$\Delta G_{NU}^0 = \int_0^{f_c} x_{ssDNA}(f')df' - \int_0^{f_c} x_d(f')df' \qquad (9.4)$$

where f_c is defined by $k^+(f) = k^+(f)$ or $w_N(f_c) = w_U(f_c)$, see Eq.(9.3). While the estimation $\Delta G_{NU} = f_c x_{NU}$ in Eq.(9.3) is an extrapolated value using the BE model, Eq.(9.4) is an exact thermodynamic relation for ΔG_{NU}, which however requires accurate knowledge of the elastic response of ssDNA and the hairpin diameter.

Table 9.1 shows the experimental parameters obtained using the BE model and the theoretical ones predicted by the unified oligonucleotide thermodynamic data set incorporating mismatches [132]. The theoretical predictions of the FEL are obtained using the mfold Web Server [102] and take into account the contribution coming from the tetraloop (see Appendix). It is worth noting that the values of ΔG_{NU} and ΔG_{NU}^0 (Eqs.9.3,9.4) are compatible with each other. However the theoretically predicted value of f_c overestimates the experimental result by $\simeq 1.5$pN. Also the theoretical and experimental values of x_{NU} agree for the native hairpin while for the hairpins with mismatches the experimental value is underestimated by 1-2nm.

Table 9.2: **20bp results.** Parameters obtained from the pulling experiments for the 20 bp hairpins using the BE model and the theoretical ones obtained from the nearest-neighbour model [132].

Hairpin	F_c (pN)	x_{NU} (nm)	$x_{N \to TS}$ (nm)	$x_{TS \to U}$ (nm)	ΔG^0_{NU} $(k_B T)$
20bp nat.	15.6(2)	17.7(8)	9.3(4)	8.4(5)	52(3)
Th. nat.	16.0	18.2	/	/	54.0
20bp GA	14.4(2)	21(3)	10(2)	11.0(12)	48(3)
Th. GA	14.6	17.9	/	/	48.0
20bp GT	14.4(2)	16.9(12)	8.3(9)	8.7(3)	48(3)
Th. GT	14.5	17.9	/	/	47.4

9.4   FREE ENERGY MEASUREMENTS FROM PULLING EXPERIMENTS IN 20BP DNA HAIRPINS

Figure 9.3 shows the results of pulling experiments. From the FDCs (Figure 9.3a) we extract the rupture force distributions $\rho_{N(U)}(f)$ and survival probabilities $P_{N(U)}(f)$ which are related by $\rho_{N(U)}(f) = -P'_{N(U)}(f)$. Kinetic rates can be extracted from the rate equations,

$$\frac{dP_N(f)}{df} = -\frac{k^+(f)}{r} P_N(f) \Rightarrow k^+(f) = \frac{r\rho_N(f)}{P_N(f)}$$

$$\frac{dP_U(f)}{df} = -\frac{k^-(f)}{r} P_U(f) \Rightarrow k^-(f) = \frac{r\rho_U(f)}{P_U(f)} \quad (9.5)$$

with $r = k_{eff} v_{pull}$ the loading rate and k_{eff} the effective stiffness (equal to the slope of the FDC).

Figure 9.3b shows the results for the kinetic rates versus force in semi-log scale obtained from Eq.(9.5) and fitted to exponential functions as predicted by the BE model, Eq.(9.1). The coexistence force is shown as the crossing point of the fits (black points in the Fig. 9.3b). From the fitting parameters and using Eq.(9.3) we extract ΔG_{NU}. Results are shown in Table 9.2.

9.5   SIMULATION RESULTS

We realized a series of molecular dynamics (MD) simulations on the 10bp hairpin with native sequence, as described in the Materials and Methods. Given the extremely large difference between the experimental and MD time-scales, a direct comparison between the theoretical and experimental results is not possible, notably as far as estimates of defect energies are concerned. Therefore, we focused the simulations only on the native hairpin sequence, aiming to provide additional insight on the molecular-scale phenomena that could be relevant to analyze and interpret the experimental results.

In a first type of *non-equilibrium* MD simulations, we performed simulated force-pulling experiments on the 10bp hairpin. This was

achieved by fixing the center of mass of the first base-pair at one end of the dsDNA handles, and by moving at constant velocity the center of mass of the first base-pair at the opposite end of the other handle. In this way, the opening λ between the two opposite ends of the dsDNA handles is linearly increasing with time, from the zero-force value $\lambda_0 =$23.0 nm with the hairpin in the folded state. For a pulling velocity in the range of a few cm/s, this translates into a similarly linear opening of the hairpin, as measured by looking at the relative distance between the sugar C1$'$, or the backbone P atoms of the first base pair (namely, the GC pair directly linked to the two dsDNA handles). The molecular extension has been chosen the natural reaction coordinate to directly compare simulations and experiments, since the experiments allow us to measure, almost directly, the molecular extension. This choice is justified by the relative simplicity of the hairpin, as compared to complex molecules such as proteins, and the particular direction of the pulling protocol, different from the one used in DNA overstretching [169]. The unfolding of DNA hairpins occurs sequentially starting from the stem along the secondary structure of the hairpin. These considerations allow us to avoid the complexities of a multi-dimension FEL as, for example, used in proteins [170].

In Fig. 9.4 are shown the result of these simulations. The simulated force-pulling experiments are followed up to the complete opening of the hairpin, over trajectories of typical duration of $\sim$ 50 ns. An example of the results is shown in Fig. 9.4a, where the relative distance ("opening") between the C1$'$ atoms of each base pair in the hairpin is shown as a function of the simulation time (bp are numbered from 1, next to the loop, to 10, next to the dsDNA handles, see the inset). A first observation that clearly emerges from such a plot is that the unfolding of the hairpin does not seem to proceed by an ordered and progressive snapping of each base pair in sequence, but rather follows a kind of collective process, in which groups of bp open up simultaneously, e.g. at times $\sim$ 17 and $\sim$ 43 ns. Furthermore, the outermost bp (n. 10 in Figure 9.4) seems to be already opened from the beginning of the simulation, and the second one follows almost immediately, after just a few ns. This observation may have interesting consequences, to be discussed in the foregoing.

To bypass at least in part the time-scale limitations of MD, we performed a second set of simulations by picking a few "interesting" configurations from the trajectory shown in Fig. 9.4a, at times $t \simeq$7.5, 18, 22, 28, 35, 45 ns, corresponding to a relative opening between the opposite ends of the dsDNA handles of $\lambda \simeq$23.6, 26.1, 26.4, 27.6, 29.0, 31.1 nm. Each of these configurations were then run in a 100-ns MD simulation at constant-$\{NVT\}$ with the fixed-λ external constraint (SHAKE-LINCS algorithms [163, 171]). The hairpin equilibrium dynamics in such conditions may be seen as an approximation to a quasi-reversible pulling experiment but on the much faster MD timescale.

Fig. 9.4b shows the equilibrium probability distributions of the C1$'$-C1$'$ distances for each bp of the hairpin, with color codes corresponding to those of Fig. 9.4a. The four panels correspond to four progressive opening values, $\Delta\lambda = \lambda - \lambda_0 =$0, 0.6, 4.4 and 8 nm. It may be noticed

that at zero opening (i.e., zero average external force) the equilibrium distribution confirms the above observation that the first base pair is constantly opened up, with a C1′-C1′ distance of about 1.9 nm. At $\Delta\lambda$=0.6 nm (corresponding to about t=7.5 ns in Fig.9.4a) the first two bp are spread open, and the third one is just broadening its equilibrium width. At $\Delta\lambda$=4.4 nm (corresponding to $t \simeq$28 ns) the outermost six bp are widely opened, while the four inner ones are still closed at their equilibrium C1′-C1′ of 1.02 nm. Eventually, at $\Delta\lambda$=8 nm (corresponding to $t \simeq$45 ns) two more bp start opening, and only the two bp closer to the hairpin loop are still in the closed state. Notably, at openings larger than about 3-4 nm, most base pairs display a doubly-peaked distribution, very likely indicative of the rotational flipping in-and-out of these bases about the backbone, and roughly parallel to the main hairpin axis.

In a last set of equilibrium MD simulations, we wanted to test the excitation dynamics of the hairpin. These simulations were run for 100 ns at constant-$\{NVT\}$, for all the values of fixed $\Delta\lambda$ above, by increasing the temperature of the simulation in steps of 10 K above room temperature. In Fig. 9.4c we present a subset of the results, namely the data for one particular value of the opening, $\Delta\lambda$=4.4 nm, and for three temperatures T=300, 320, 340 K. At such opening, each 20 K increase in temperature corresponds to an applied force $\Delta f \simeq$18.5 pN, that is a value comparable to the experimental coexistence force f_c. Therefore, by running MD at 320 and 340 K under fixed opening, we are simulating the effect of injecting once/twice an amount of energy $\Delta f \cdot \Delta\lambda$: for a fixed $\Delta\lambda$, this would correspond to proportionally increasing the effective stiffness of the optical trap. Figure 9.4c shows the probability distributions for the bp 10 to 4, the innermost 3 bp remaining closed at any temperature; the traces in color correspond to T=300 K (blue), 320 K (black) and 340 K (red). It is noted that at room temperature, the energy injected to keep the hairpin at the opening of $\Delta\lambda$=4.4 nm is unevenly distributed among the outermost bp (6 to 10, with the 5 only slightly excited), which display average amplitudes decreasing towards the loop end; the width of the distribution is nearly double for the outermost bp 10, compared to the other ones. As the injected energy (i.e. temperature) is increased, the distributions tend to become more even (centroid of the distribution decreasing for bp 8-10, and increasing for bp 4-7), while at the same time each distribution covers a very broad range of bp opening, spread over about 2 nm for the outermost bp. We interpret these results as the effect of increased cooperativity in the unfolding transition, at high values of force: while at low forces the bp tend to open individually or in small batches, at high applied forces the unfolding appears to occur by a simultaneous opening of the entire hairpin.
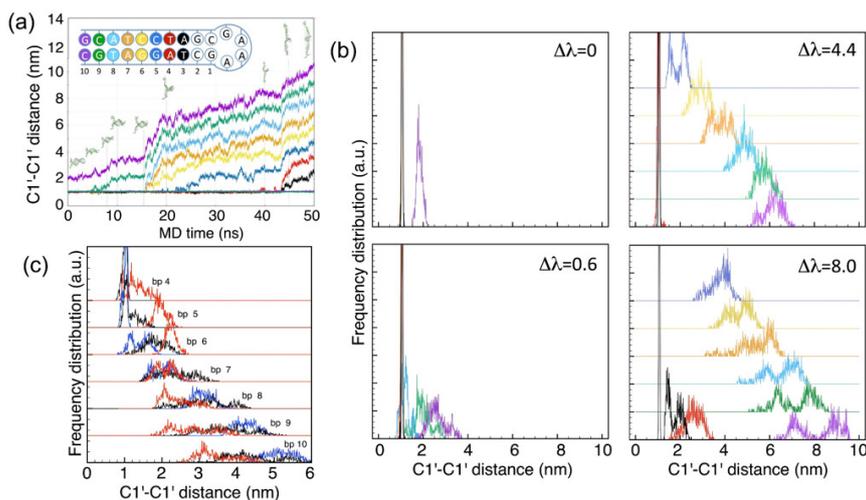
Figure 9.4: **MD simulations.** (a) Constant pulling speed (about $10^6 - 10^7$ times larger than in the experiments) and T=300 K for the 10bp hairpin with native sequence (shown in the inset above, with bp numbering and color codes). The 50-ns long time traces, starting from the folded state and going to the fully unfolded, display the relative distance between the sugar C1′ atoms of each base pair, as indicated in the legend (the bp number 10 being the one close to the dsDNA handles, and the bp 1 being the one adjacent to the loop). Above the plots, schematic snapshots give a visual indication of the average molecular configuration of the hairpin at approximately the time corresponding on the x-axis. (b) Equilibrium probability histograms for the C1′-C1′ distance of each bp in the native hairpin, from 100-ns long MD $\{NVT\}$ trajectories at $\Delta\lambda$=0, 0.6, 4.4, 8 nm. The histograms are colored according to the bp scheme of panel a; data for the two larger $\Delta\lambda$ are shifted upwards on the y-axis for better clarity. (c) Equilibrium histograms for the C1′-C1′ distance of each bp in the native hairpin (numbered according to the inset scheme in panel a), from 100-ns long MD $\{NVT\}$ trajectories at fixed $\Delta\lambda$=4.4 nm, and different temperatures: T=300 K (blue), 320 K (black) and 340 K (red).

9.6 COMPARISON OF THE RESULTS

The ability to detect point defects along the sequence of a DNA molecule is a relevant problem, both for molecular biology studies of DNA replication and damage repair, and for the many technological applications exploiting base-pair complementarity of nucleic acids (e.g. PCR, DNA self-assembly, etc.). However, differences in base pairing between the native, Watson-Crick A-T and G-C purine-pyrimidine pairs, and the defective ones, typically including mismatches (MM), are very subtle and escape a direct experimental determination. The effect of a MM in a random position of a DNA sequence may reveal itself as a small local difference in free energy, which however shows up only when the two DNA strands are split apart, e.g. during replication; or as a small variation in the elastic and mechanical properties of the molecule to an applied stress, e.g. by a repair enzyme.

We have used single-molecule force spectroscopy with LOT, to demonstrate the ability of this experimental method to detect the small free-energy and force differences associated with a single, isolated MM in a

DNA sequence. As a test system, we adopted a short self-complementary hairpin undergoing a reversible folding/unfolding transition, with the aim of extracting information about the differences induced in this dynamical transition by the presence of a G-A or G-T mismatch, inserted in the middle of the hairpin. This same system was used in a recent work [40] in which longer hairpin constructs including pairs of MM were used. With respect to that work, here we tested the lower limits of the technique, to detect a single defect in the shortest possible sequence. To this end, we used hairpins of length 10 and 20 bp, in hopping and pulling experiments respectively. Both methods showed pros and cons, namely: (a) the hopping method is more adapted to the shorter 10bp hairpins, since the folding/unfolding rate is exponentially larger, therefore a large number of folding/unfolding transitions can be observed in this case, with much better statistics; (b) the force-pulling method is more adapted to the longer 20bp hairpins, since the difference between the folding/unfolding force path is proportional to the hairpin length, and for the shorter ones the two force branches are so close to be nearly indistinguishable.

Table 9.3: Folding free energies for the 10bp and 20bp DNA hairpins. Experimental data (second column) are compared with estimates obtained from the nearest-neighbor model (NN) using the energies from SantaLucia [132] or from unzipping experiments [82] .

Hairpin	ΔG_0 (kcal/mol)	NN [132] (kcal/mol)	Unzipping [82] (kcal/mol)
10bp native	12.5 ± 2	14.3	12.0
10bp GA	9.5 ± 1	10.7	-
10bp GT	9.5 ± 1	10.4	-
20bp native	31.5 ± 1	32.0	29.5
20bp GA	28.5 ± 1	28.4	-
20bp GT	28 ± 1	28.1	-

In both sets of experiments, we could clearly demonstrate the detection of a single MM in the hairpin, both in the form of a lower free energy, ΔG_0, and a lower coexistence force, f_c (see Tables 9.1-9.3). However, the difference in free energy between the hairpins with G-T and G-A mismatches obtained using both methods, pulling and hopping experiments, show similar and compatible values (see table 9.4). An improved estimate of the relative free-energy differences $\Delta \Delta G$ between the native and mismatched sequences for the 10bp hairpin can be obtained by subtracting the two independent values ΔG and ΔG_0 obtained from kinetics (Eq.9.3) and thermodynamics (Eq.9.4). We obtain $\Delta \Delta G_{GA} = 4 \pm 2$ kcal/mol and $\Delta \Delta G_{GT} = 4.5 \pm 2$ kcal/mol .

As already noted, the absolute values of free-energy for the 10bp hairpin in Tables 9.1 and 9.3 appear to be lower (by $\sim 15\%$) than the nearest-neighbor (NN) model prediction [172]. This agrees with the trend observed for the lower values obtained for f_c and x_{NU} as compared to the theoretical predictions (Table 9.1) Yet, for the native

10bp hairpin we find a value of ΔG_0 that falls closer to the unzipping prediction. The opposite situation is found for the 20bp native hairpin where the NN prediction works better than the unzipping prediction (Tables 9.2 and 9.3). Overall this suggests that, for the 10bp and 20bp hairpins, the measured and predicted energy numbers agree within errors.

Table 9.4: Folding free-energy differences $\Delta\Delta G_0$ between the native and mismatched sequences obtained from different experiments, compared with NN model predictions. $\Delta\Delta G_0$ is obtained from the results shown in tables 9.1-9.3

hairpin	$\Delta\Delta G_0$ [kcal/mol]	NN model [kcal/mol]
10bp GA	3 ± 3	3.6
20bp GA	3 ± 2	3.7
10bp GT	3 ± 3	3.9
20bp GT	3.5 ± 2	3.8

However, we cannot exclude the possibility that the apparently lower thermodynamic stability (and lower f_c and x_{NU}) of the 10bp hairpin arises from fraying effects. Note that the 10bp hairpin has only two GC bp at the beginning of the stem followed by two AT bp which render the fork less stable. MD simulations, show that the first base-pair in the hairpin stem appears to be always opened, even at zero applied force, decreasing the folding free energy of the hairpin as observed. This same effect was observed also in the longer hairpins of the study by McCauley et al. [40]. The MD simulations suggest that the first bp of the hairpin could be already open even before the hairpin unfolds. To test this possibility we have compared the experimental results with the NN model ignoring the contribution of the first G-C pair in the 10bp hairpin (NN*). The new NN* energies show a lower value than the experimental ones for the 10bp case, making the experimental values fall in between the two predictions (NN and NN*). The energy numbers are (in kcal/mol):

- 10bp native 11.8[NN*]< 12.5 [exp] < 14.3 [NN];

- 10bp G-A 8.2[NN*]< 9.5 [exp] < 10.7 [NN];

- 10bp G-T 7.9[NN*]< 9.5 [exp] < 10.4 [NN].

Fraying might be tested in future experiments and simulations as a possible explanation by studying 10bp hairpins with more stable stems. Overall the results show the power of combining computer simulations and experiments to improve the models used to interpret the experimental data. Note that for the 20bp, the NN values and the experimental ones are compatible (Table 9.3), therefore application of a NN correction in this case might be unnecessary. There are two possible explanations

for such a difference. First, while the 10bp hairpin starts with two G-C bp at the beginning of the stem, the 20bp has seven G-C bp. Second, in the 10bp hairpin the distance between the fork and the MM position (4-5bp) is well below the 10-11 nucleotides (a full turn of the double helix), which is generally considered as the limit for two defects to interact. For the 20bp hairpin such distance (9-10 nucleotide) is larger. Future studies might consider addressing the fraying problem with more detail.

In the Bell-Evans model the folding/unfolding process is described by a thermally activated diffusive process across an energy barrier, [173] separating two independent conformations of the molecule. The combined Bell-Evans and HMM analysis of the hopping data for the 10bp defective hairpins did not show evidence of an intermediate state possibly caused by the mismatch (shown in Appendix B). The same was found in preliminary hopping data for the 20bp defective hairpins. The MD simulations demonstrated a variable degree of cooperativity in the folding/unfolding transition of the 10bp native hairpin, in that at lower forces the hairpin appears to unfold in a sequential way, but with groups of bases opening up together, while at higher forces the hairpin tends to open up in one collective snapping of the bonds between bps. For a given displacement λ, a larger force translates into a stiffer optical trap, and such an increase in cooperativity with stiffer coupling is in agreement with a previous experimental study [174] and theoretical models [175, 176]. The transition from an additive to a collective unfolding may also be interpreted in terms of an increasing "friction" effect that builds up between the closed bps, which must overcome a twist elastic barrier, at the same time as the chemical bond-breaking barrier [177]; however in the present case this effect would be driven by a variable force, rather than by variation in the polymer physical length.

9.7 CONCLUSIONS

Single-molecule force spectroscopy with LOT permits to identify the presence of a single mismatched base-pair in short 10-20 bp DNA hairpins. The mismatch decreases the thermodynamic stability of the hairpin (by lowering its free-energy of formation and coexistence force). It also speeds up the kinetics of folding of the hairpin to a larger extent, by lowering the height of the barrier.

Passive hopping experiments allow us to detect a single mismatch in 10bp hairpins, while pulling experiments produce a better signal in longer sequences (20bp).

The folding/unfolding transition in the 10bp and 20bp defective hairpins is well described by a two-state model, without evidence for intermediate states. This result is based on the analysis of hopping experiments and a Hidden-Markov model.

MD simulations of the hairpin unfolding transition under external force support this view and provide important clues for the analysis

of experimental data; simulations also suggest a variable degree of cooperativity in base-pair opening during the forced unfolding.

The folding/unfolding transition in short hairpins is therefore a relevant test bed for studying defects in DNA. Further work should concentrate in refining the experimental set up, in order to arrive at a better quantitative characterization of point defects, and move towards the study of the role of signalization and repair proteins in defect dynamics. MD simulations can play a relevant role in assisting and guiding the experimental analysis.

Part IV

FINAL CONCLUSIONS

FINAL CONCLUSIONS

The interest of physics in studying biological systems started more than two centuries ago. However, the actual concept of biophysics would require a the development of a lot of theoretical and experimental breakthroughs. At the beginning of the 20th century, physics was experiencing a great revolution fueled by the advent of relativity and quantum theory. Regarding quantum physics, the experimental implications of the theory would be immense, leading, among others, to the creation of the lasers. The invention of a monochromatic and coherent source of light would become a milestone of modern physics. One of the most relevant and broadly used applications of the laser, optical tweezers, would be discovered by Arthur Ashkin. Laser optical tweezers (LOT) would provide the capability of performing new experiments, allowing the advent of single-molecule biophysics.

Single-molecule experiments have emerged as a powerful tool to study biomolecules with an unprecedent resolution. The ability to manipulate individual molecules and monitor the changes in their mechanical properties has proven attractive for its biological applications, since it allows to study the individual unfolding of a molecule, without averaging the information over a large number of molecules, which effectively hides some of the information regarding their properties. LOT experiments have in common the relevant role played by thermal fluctuations. The measured forces and distances are in the range of nanometers and picoNewtons, respectively. Therefore, the energies explored are on the order of the thermal fluctuations at room temperature, $1k_{\mathrm{B}}T \sim 0.25$pNnm. Consequently, each repeated experiment may yield different outputs, which makes them a wonderful tool to check and develop new physical laws in non-equilibrium physics.

In this thesis, single-molecule experiments using LOT are employed to extract accurate information about the thermodynamics and kinetics of various molecular systems, with special emphasis on the elastic properties of single-stranded DNA (ssDNA). The distance and force ranges provided by LOT ($\sim$nm) allow to study the elasticity of ssDNA molecules over four decades of length: from $\sim 10^1$ to $\sim 10^3$ bases.

The first part of the thesis provides a general description of the research field as well as the main theoretical framework for the basic concepts that will be developed throughout the thesis. In chapter 2, we introduce the miniTweezers and the experimental setup used throughout the thesis is described, as well as the physical basis of its working mechanisms, introducing the phenomenon of optical trapping. A brief introduction of the biomolecules of study in this thesis is included in chapter 3, with an explanation of their historical discoveries, as well as their structure and function. The main focus of this chapter is on ssDNA, which is the main object of study of the thesis. Chapter 4 is about the description of

the elasticity of ssDNA, where polymer models have been introduced. Particularly, the Freely-Jointed Chain and Worm-Like Chain models are used to reproduce the behaviour of ssDNA molecules throughout the thesis.

In chapter 5, we study the elasticity of ideal ssDNA chain, by fitting the elasticity of different DNA sequences to the WLC model. In past studies, ssDNA elasticity was described by a broadly dispersed range of parameters, with contour lengths varying up to 50% and persistence length ranging $\sim 0.5 - 3$nm. The blocking-splint oligo technique is described, which allows us to study the elasticity of short (tens of bases) DNA molecules in a wider range of forces. The obtained data is only successfully reproduced by using the extensible WLC. In the literature, the elastic parameters were obtained by using different sequences and techniques, which lead to the use of different ranges of forces to fit the curves. The need of the extensibility in order to successfully characterize the elasticity of ssDNA explains some of the previous previous discrepancies regarding the elastic parameters (e.g. persistence length) obtained in different studies.

The extensibility of a polymer chain can be understood by means of increasing its contour length by increasing the applied force. This can be easily understood for molecules that have a certain degree of structure that can be stretched, such as for the case of the dsDNA double helix. For the ssDNA case, however, the interphosphate distance is limited by covalent bonds in the molecules, which makes harder to provide a physical interpretation of the extensibility. The explanation we provide for the required extensibility of the model requires of a transition experienced at the nucleotide level: for example, a change in DNA sugar pucker conformation. A simple two-states model is introduced and preliminary results regarding its energetics are presented.

Some DNA sequences have been shown to experience a stacking-unstacking transition. In chapter 6, we study several molecules, with different degrees of stacking, by obtaining their force extension curves (FECs) using the blocking-splint oligo method. The degree of stacking given in each one of the studied molecules is a function of the consecutive number of adenines present in them, as was expected from previous works. We have also demonstrated that for short random DNA sequences without consecutive adenines, the behaviour corresponds to an ideal ssDNA, without any signatures of the S-U transition, well captured by an extensible WLC. We obtain the elastic parameters describing the FECs of the ideal ssDNA behaviour over 3 decades of salt concentration ($10^{-3} < [\mathrm{NaCl}] < 10^{0}$M).

A cooperative helix-coil model including heterogeneity is developed and used to fit the obtained FECs, allowing to obtain elastic parameters that describe the stacked chain. The salt dependence of the stacking-unstacking transition is also measured by varying the salt concentration over two decades. The free energy of formation of dsDNA duplexes depends on the salt concentration. The obtained salt dependence on the stacking free-energy of ssDNA provides a possible explanation for the observed salt dependence of duplex formation.

The model also allows to obtain the elasticity of the fully-stacked ssDNA state, which is characterized by fitting the model to experiments with different sequences and varying salt concentration, from 50mM NaCl to 1M NaCl.

Further work could be done, by varying the $MgCl_2$ concentration in the experiments, characterizing the elastic response of the unstacked and stacked states regarding the difference between monovalent and divalent ions. Also, a deeper study of the stacking formation in short molecules could be performed by performing experiments varying the loop sequence. Then, by varying the motifs present in the loop and repeating the analysis presented in this chapter one could investigate more in detail which motifs act as nucleation sites for the stacking of the sequence.

One of the main limitations in order to extract the elasticity of ssDNA of long (> 100bases) molecules at high salt concentration (NaCl $>$ 100mM) is the formation of non-specific secondary structures at low forces ($f < 10$pN). This gives rise to several questions such as: how is it formed? Does it depend on salt conditions? How does the sequence length affect it? In chapter 7, we studied the formation of secondary structure that appears in the FECs obtained by pulling ssDNA molecules longer than ~ 100 bases. In order to reproduce the obtained FECs, a helix-coil model with cooperativity is proposed and used to extract some mean-field characteristics of these structures. 8 different sequences are studied, for lengths comprising 120 to ~ 14000 bases. Their elasticity and deviation from the ideal elastic behaviour is characterized. We obtain a length dependence in secondary structure formation, that we interpret it is due to finite-size effects, which can be reproduced in our model.

The model can be used to predict the formation of secondary structures at zero force. This allows to compare our predicted structures from the model and those obtained from Mfold. The agreement between the predicted free energy of formation and the distribution of the bases in the hairpin-like structures of the secondary structure for both models is remarkable. However, discrepancies start to appear for long molecules (> 1000 bases), for which Mfold predicts the formation of large secondary structure domains, that our homogeneous model does not capture.

In order to explore the salt dependence of the secondary structure, we focused on the results for a ~ 14kb molecule for 3 decades of varying NaCl and $MgCl_2$. All experimental FECs are fitted to the helix-coil model. As expected, the free-energy of formation of the secondary structure is salt dependent, which the model captures by incorporating a logarithmic dependence of the free energy of base pairing, with a salt-correction parameter that matches that reported for DNA hybridization.

The last part of the thesis is about two different applications for which it is key to correctly determine the ssDNA elasticity. In Chapter 8, we study the interaction between the RecQ helicase from E. coli and DNA, i.e. how the RecQ unwinds double-stranded DNA molecules, releasing single-stranded DNA. We obtain some of its kinematic properties, such as the unwinding rate, which we find to be about ~ 50bp/s, independently

of the applied force, in agreement with the literature. The variance increases linearly with time, showing that RecQ behaves following an advection-dffusion process.

From analyzing the energetics of the whole system (helicase, DNA and trap), we have obtained an expression for the entropy production of the system using the Fluctuation Theorem. By assuming a tight coupling between the chemical and mechanical energies and fitting the entropy production to experimental data we obtain $n_{ATP} \approx 14\text{bp/ATP}$ for the number of translocation bp per hydrolized ATP, which is much larger than the value reported in the literature [120–122].

We have also implemented Monte-Carlo simulations based on a Brownian ratchet model, aiming for a deeper understanding of the physics of the RecQ experiments. The simulations are based on very simple assumptions, but qualitatively reproduce main characteristics of the experimental results.

In Chapter 9, the effect of DNA mismatches, i.e. non complementary base pairing, on the stability of DNA is studied. To do so, two types of experiments on several DNA sequences are performed: stretching and releasing the molecule by moving the optical trap (pulling experiments) and monitoring the folding/unfolding of the molecule passively (hopping experiments).The mismatch decreases the thermodynamic stability of the hairpin (by lowering its free-energy of formation and coexistence force). It also speeds up the kinetics of folding of the hairpin to a larger extent, by lowering the height of the kinetic barrier.

The results are compared to MD simulations, which provide important clues for the analysis of experimental data; simulations also suggest a variable degree of cooperativity in base-pair opening during the forced unfolding.

The correct characterization of ssDNA elasticity is key to extract thermodynamic properties from force-spectroscopy techniques involving ssDNA: its applications go from extracting energetic parameters from a helicase unwinding dsDNA to the ability to detect mismatches in DNA hairpins. An accurate description of ssDNA elasticity requires to take into account complex and, most-likely, coupled phenomena described throughout the thesis. The stacking of bases as well as the formation of non-specific secondary structure formation are dependent on the sequences. For this reason, further work regarding varying the sequences of study needs to be performed in order to get a wider and more comprehensive picture of ssDNA. Unsolved questions such as the paper of stacking in the secondary structure formation, the origins of the extensibility in the models required to reproduce their FECs or the sequences that act as nucleation sites regarding stacking will require to further develop, improve and combine current experimental techniques.

Part V

APPENDIX

THE MINITWEEZERS BEGINNER'S GUIDE

This section contains a brief guide aiming to contain all of the minimum information required to start performing the first experiments in the miniTweezers. We describe in detail how optical tweezers can be used to investigate molecular folding using short DNA hairpins (a double-helix stem of a few tens of base pairs terminated by a loop) as a model systems. We will describe how to synthesize a short DNA hairpin, how to unzip it by stretching it while holding it between a micropipette and an optical tweezers, and how to analyze the obtained data. This experiment provides an ideal example of the use of single-molecule mechanics to determine elastic behaviour of biopolymers, to measure the intramolecular weak interactions that stabilize native molecular structures, and to extract the free energy of biomolecule formation from out-of-equilibrium pulling experiments. The text is an adapted version of the Chapter 4.1 of [178], which has been submitted to publish in Advances in Optics and Photonics. The experimental setup we use is based on the miniTweezers setup developed by Smith, Cui and Bustamante [22], described in Sec 2.2.

A.1 SINGLE-MOLECULE MECHANICS

A.1.1 *Microfluidics chamber*

The schematic of the experimental setup is presented in Fig. A.1(a). The experiments are performed in a microfluidics chamber placed vertically within the setup. As schematically shown in Figs. A.1(b) and A.1(c), this microfluidics chamber has three channels: we will refer to the three channels as the upper, central and lower channel. The propagation of both laser beams is perpendicular to the chamber surface (z-axis). The experimental area is restricted to the central channel, where the object of study (biomolecule or cell) is held between two beads, one held by the optical trap and the other held by a glass micropipette, as shown in Fig. A.1(d). The upper and lower channels are used to supply the two types of coated beads used in the experiments.

The microfluidics chamber is realized by sandwiching two layers of parafilm (Parafilm M, Bemis) between two coverslips (No. 2, dimensions $24\,\text{mm} \times 60\,\text{mm} \times 200\,\mu\text{m}$), as shown in Fig. A.1(c). In detail, the steps to prepare the chamber are:

1. The entrance and exit holes are drilled in one of the glass coverslips using a laser cutter. The coverslips are cleaned with a solution of 70% ethanol.[1]

---

1 Although 100% ethanol can also be used, a 70% solution is preferred because it prevents the sporulation of some microorganisms.
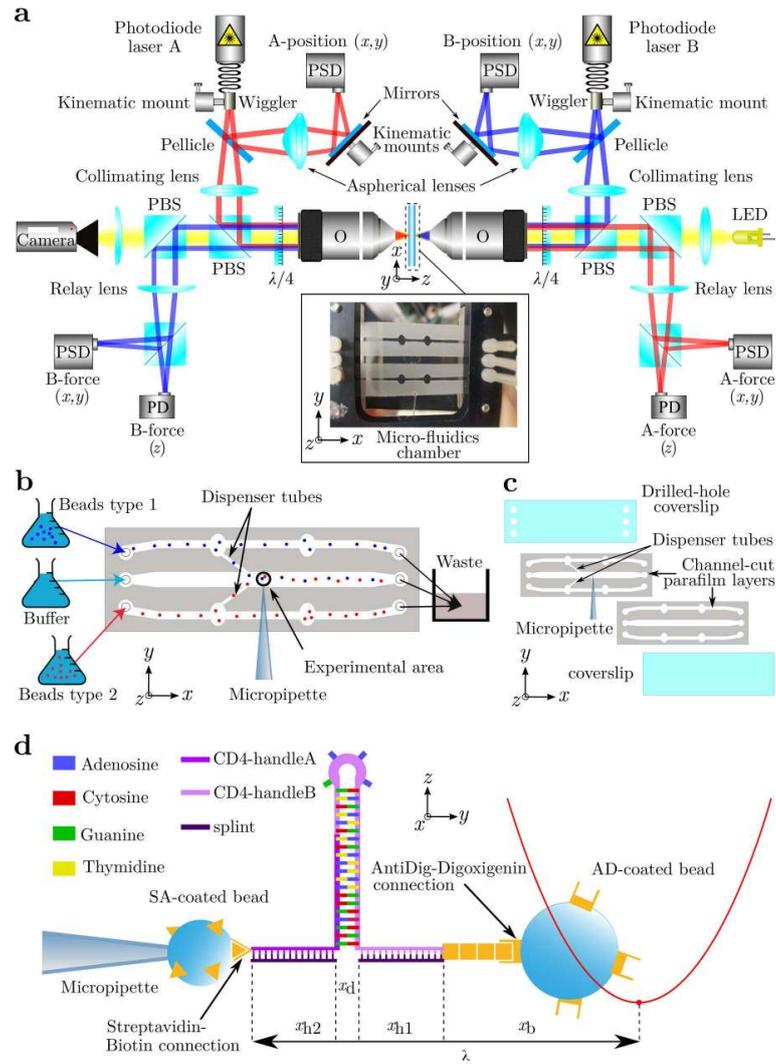
Figure A.1: **Single-molecule pulling experiments: miniTweezers setup.**
Schematic representations (a) of the miniTweezers setup, (b) of the microfluidics
chamber (the flow goes from left to right, and the laser beams propagates
perpendicularly to the surface of the chamber), (c) of the assembly procedure
for the construction of the chamber, and (d) of the molecular configuration
for the DNA hairpin pulling experiments (where the DNA hairpin is held
between a bead held by a micropipette and a bead captured in the optical
trap, depicted as an harmonic potential). Note that x_{h2}, x_d, x_{h1}, and x_b are
distances along the y-axis.

2. The three channels are drawn in the two parafilm layers using a
   laser cutter.

3. One parafilm layer is attached on the drilled glass coverslip.

4. A glass micropipette with a diameter of $\sim 1\,\mu m$ is produced by
   heating and pulling a glass tube (King precision glass, Inc., inside
   diameter=0.04 mm, outside diameter=0.08 mm, length=6.00 mm,
   glass type KG-33), as described in section A.1.3.

5. This glass micropipette is placed on top of the parafilm layer perpendicular to the channels with the tip positioned in the central channel, as shown in Figs. A.1(b) and A.1(c).

6. The upper and lower channels are connected to the central one via glass dispenser tubes (King precision glass, Inc., inside diameter=0.04 mm, outside diameter=0.10 mm, length=6.00 mm, glass type KG-33) cut using a diamond-tip cutter to obtain a clean cut.

7. The second parafilm layer is placed on top and the chamber is closed using the second glass coverslip (previously cleaned with 70% ethanol), as shown in Fig. A.1(c).

8. The chamber is sealed by heating it on a hot plate at 120°C while exerting a pressure of about 1 Kg in all the chambers' surface (either by placing a weight on top of the microfluidics chamber or, more simply, by exerting pressure by hand). To prevent the glass from breaking, it is recommended to sandwich the chamber between two thicker glass slides to homogenize the pressure applied to the surface.

9. The chamber is placed in the metallic mount, as shown in Fig. A.1(a) (zoomed picture). The correct alignment between the holes drilled in the glass coverslip and the holes in the mount is critical to ensure the flow in the chamber. It is also very important to tighten the screws to avoid buffer losses between the plastic tubes and the glass chamber, since this would cause flows inside the chamber. Do this carefully, because tightening the screws too much could break the glass coverslip.

10. The side of the micropipette coming out of the chamber is cut to get rid of the excess tube, leaving only about ~ 3 cm, and is introduced into a polyethylene tube (polyethylene tubing, Warner Instruments, PE-10/100; outside diameter $= 0.61$ mm, inside diameter $= 0.28$ mm). This tube is then fixed to the mount with tape. The pipette is easily breakable, so the tube needs to be placed as straight as possible. The connection is then sealed by using a special glue (Norland, NOA-61; UV Curing Optical Adhesives). The glue is placed between the the glass tube of the pipette and the polyethylene tube to fill the void outer space between tube and pipette via capillarity. After observing that the glue has entered into the tube, the UV-curation is performed by leaving the whole chamber for ~ 25 min under a UV lamp. Since the glue could reach the end of the glass tube of the micropipette and block it, the chamber has to be placed under the UV-radiation as fast as possible. After the curation, a syringe (1 ml Luer Lock, HSW SOFT-JECT U100 Insulin Henke Sass Wolf) is inserted using a needle (BD Microlance 30G $\times$ 1/2" $- 0.30$ mm $\times$ 13 mm) at the end of the tube to create suction at the tip of the micropipette.

11. The holes of the mount are connected with silicone-rubber tubes
    (Tygon 3350, Saint-Gobain, .031 ID X .093 OD X 50 FT TY-
    GON 335) via nylon socket screws (Nylon set $8 - 32 \times 3/8$",
    Product-Components, previously drilling a hole with a number
    45 drill bit, of 0.82"), as shown in Fig. A.1(c). A segment of
    ~ 20 cm polyethylene tube (Polyethylene tubing, Warner Instru-
    ments, PE-50/100: outside diameter $= 0.97$ mm, inside diameter
    $= 0.58$ mm) is inserted into the silicone-rubber tubes. The three
    exit tubes are connected to a trash (any small plastic container
    with a capacity volume ~ 100 ml), while the three entry channels
    are connected to a syringe using a polyethylene tube (Polyethy-
    lene tubing, Warner Instruments, PE-50/100: outside diameter $=
    0.97$ mm, inside diameter $= 0.58$ mm) which is inserted into the
    silicone-rubber tubes on one end, and a needle (HSW FINE-JECT
    $23G \times 1$" $- 0.6$ mm $\times 25$ mm).

A.1.2   *The miniTweezers setup*

The miniTweezers [22] (Fig. A.1(a)) consists of two focused counter-
propagating laser beams ($P = 200$ mW, $\lambda = 845$ nm) that create a
single optical trap [16]. It employs high-NA objectives (NA $= 1.2$),
but underfills them to be able to collect almost all scattered light to
measure the change of light momentum using position sensitive detectors
(PSDs). Before the objectives, a pellicle diverts $\sim 8\%$ of each laser beam
to a secondary PSD to determine the optical trap position, λ. The
remaining $\sim 92\%$ is collimated by using a lens and it is introduced into
the optical axis by using a Polarizing Beam-Splitter (PBS) that selects
the horizontally polarized light. Then the linearly polarized light of each
laser is circularly polarized by a quarter waveplate[2]. The laser beams
are focused by their corresponding objectives and form the optical trap,
which can hold dielectric objects with a refraction index bigger than
the aqueous surrounding medium (e.g. polystyrene bead) and exert force
to them. The exiting light is collected by the opposite objective and
it is converted to vertically polarized light by another quarter-wave
plate. The vertically polarized light can be extracted from the optical
path using two PBSs and one relay lens that redirect the light to the
PSD that measure the intensity of the beam (i.e. its exerted force). The
miniTweezers has a resolution of 0.1 pN and 1 nm at a 1 kHz acquisition
rate.

The chamber where the experiments are performed (section A.1.1) is
placed between the two objectives and held with a metallic mount with
three stages to permit movement along the x-, y- and z-directions. There
are two possible ways to manipulate the position of the optical trap with
respect to the chamber depending on the precision that is needed. For
large displacements (up to hundreds of micrometers), the whole chamber

---

2 The use of quarter-wave plates ensures that the light coming from the two laser
  beams do not interact with each other and guarantees that the light reflected from
  the particle is not returned to the laser but is reflected to the opposite PSD[5].

is displaced along the x-, y- and z-directions using stepmotors. For fine displacements (less than a few micrometers, but with a resolution of $\sim 1\,\text{nm}$ and rates of displacement up to $\sim 50\,\text{nm}\,\text{s}^{-1}$), the optical trap is displaced along the x- and y-directions by a 2D piezoelectric motor attached to the tip of the optical fiber (wiggler) used to inject the laser into the optical setup.

Before any experiment, the setup needs to be aligned going through the following steps:

1. Remove any trapped object.

2. Move the chamber with the stepmotors so that the micropipette is a few micrometers away from the optical trap, in the so-called "working zone".

3. Remove any voltage applied to the piezoelectric motor.

4. Turn off the LED and remove the light filter, to be able to see the optical trap on the camera.

5. Decrease the laser power until it is possible to separate both lasers.

6. Using the screws of the kinematic mount, move the B laser (the one that has the real image shown on the screen) until it is superposed on the A laser.

7. Put back on the light filter, turn on the LED, and increase the laser power to its working value.

8. Set the current values recorded by the PSD as the zero force baseline along the x- and y-directions (center of the light spot) and along the z-direction (size of light spot).

9. Trap a bead with the optical trap by moving the chamber with the stepmotors close to the appropriate dispenser tube.

10. Bring the trapped bead to the working zone.

11. Fine-tune the alignment of the B laser using the screws of the kinematic mount controlling the wiggler of its fiber so that the xy-force signals from both force-PSDs are as close as zero as possible, i.e., both lasers exert an xy-force as close to zero as possible.

12. Fine-tune the position of one of the objectives along the z-directions so that the z-force from both force-PSDs is as close to zero as possible.

13. For both lasers, move the kinematic mount of the mirror that deflects the light that gets towards the position PSD (between the aspherical lenses and the position-PSD) until its signal is zeroed. One way to put this mirror would be to displace the aspherical lens label downwards (right above the objectives), and place the position PSD next to the lasers, facing downwards. The mirrors

and their respective kinematic mount would be placed where the A and B position PSD are placed, reflecting light from horizontal to vertical direction.

A.1.3 *Pipette making*
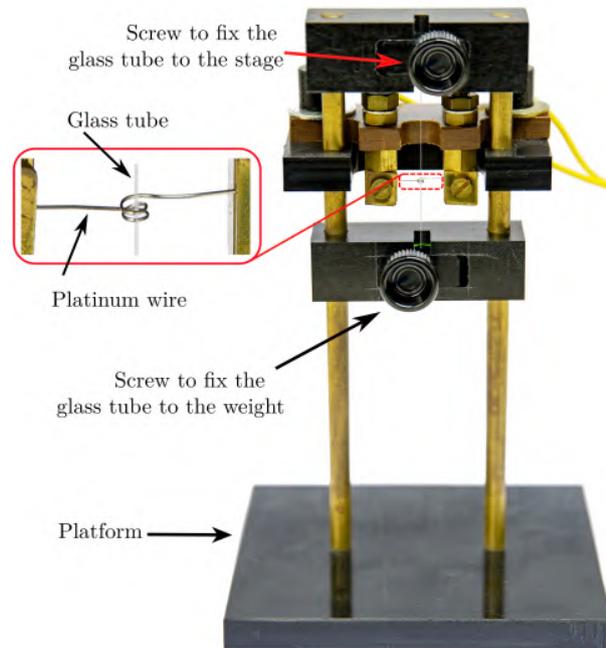
Figure A.2: **Pipette puller.** First, the glass tube is carefully centered within a coiled platinum wire, as shown in the inset. Then, one end of the tube is attached to the puller, while the other is attached to a weight that pulls the tube down. Finally, an electric current intensity ramp is applied through the platinum wire, heating the adjacent glass tube, which during the melting is pulled down, creating a micropipette.

A micropipette is used to hold a bead by air suction, as shown in Fig. A.1(b). The tip of this micropipette needs to have an inside diameter $\sim 1\,\mu$m, large enough to exert sufficient suction and small enough not to let the beads flow inside it. One might use a commercial pipette puller to produce the pipettes. Here, we will explain how a glass tube can be pulled to produce such micropipette using the homemade device shown in Fig. A.2 [22, 23].

This homemade pipette puller consists of a plastic platform with two parallel metallic bars, which hold a stage and along which a plastic weight can slide up and down. The two ends of the glass tube are fixed with screws to the stage and the weight, respectively. Thus, the glass tube is held in tension due to the pull exerted by gravity on the weight. The central part of the glass tube passes through a platinum wire; the two ends of the wire are connected to an electric supply, which provides a current ramp from 0 to $\sim 6\,$A in $\sim 8\,$s. The exact maximum intensity and time of the ramp have to be tuned to get the appropriate diameter

and shape of the pipette (in general, longer times and higher intensities correspond to smaller tips). Specifically, the steps required to produce a micropipette are:

1. Unplug the wire from the electric supply and place the pipette puller horizontally.

2. Insert the glass tube through the stage, the platinum coil, and the plastic weight.

3. Carefully center the glass tube within the platinum wire.

4. Screw the glass tube at the stage and plastic weight.

5. Place the platform vertically (as shown in Fig. A.2)), carefully to avoid any sudden hit that may break the glass tube.

6. Apply the current ramp. The weight will drop into the platform. The geometry of the platinum wire is of a critical importance to obtain the proper shape and size of the micropipette. For that reason, when the weight drops into the platform, we have to ensure that the remaining glass tube of the stage is not touching the platinum wire, otherwise it would get stuck to it.

7. Unscrew the plastic weight and (with extreme caution and preferably with ethanol-cleaned tweezers) take the micropipette and place it into the chamber being built.[3]

A.1.4 *Beads preparation*

Biomolecules such as DNA, RNA and proteins are chains whose units are either nucleic bases or aminoacids [54]. These molecules can be conveniently manipulated through beads attached at their ends. Antigen-antibody connections can be used to attach a specific bead at each end, taking advantage of the fact that these connections are extremely specific [179]. In particular, avidin/streptavidin–biotin[4] and digoxigenin–anti-digoxigenin connections are often employed in single-molecule force-spectroscopy experiments, because they can hold forces up to $\sim 100\,\mathrm{pN}$ at typical optical-tweezers loading rates of $\sim 1\,\mathrm{pN\,s^{-1}}$ [181].

Each bead is coated with a different molecule, which specifically binds to the cognate tails. Streptavidin-coated (SA) beads can be directly purchased (SPHERO streptavidin – polystyrene particles, $0.5\%\,\mathrm{w/y}$, $2.17\,\mathrm{\mu m}$, 5 ml). Anti-digoxigenin-coated (AD) beads are purchased as G-protein-coated polystyrene beads (Kisker Biotechnologies – G-coated polystyrene particles, $0.5\%\,\mathrm{w/y}$, $3.18\,\mathrm{\mu m}$, 5 ml), which must then be activated with anti-digoxigenin. The difference in size between these beads permits one to easily distinguish them by microscopy.

---

3 The micropipette easily breaks or gets blocked by dust particles at the minimum contact of the tip of the glass tube with anything, so it needs to be handled with extreme care.

4 The interaction between streptavidin and biotin is one of the strongest non-covalent interactions in Nature [180].

A.1.4.1   *Buffers.*

The following buffers are required for the bead preparation:

PBS (PH 7.0) NaCl 0.14 M, KCl 2.7 mM, $K_2HPO_4 \cdot 3H_2O$ 61 mM, KH_2PO_4 39 mM, NaN_3 (sodium azide) 0.02%. To prepare 50 ml: fill ~ 40 ml of a 50 ml Falcon tube with Milli-Q water; add 0.406 g of NaCl, 0.01 g of KCl, 0.696 g of $K_2HPO_4 \cdot 3H_2O$, 0.265 g of KH_2PO_4, and 1 g of NaN_3; dissolve using a magnetic mixer; add Milli-Q water until reaching 50 ml; check the pH and add NaOH until the solution reaches a pH 7.0.

PBS (PH 7.4) NaCl 0.14 M, KCl 2.7 mM, $K_2HPO_4 \cdot 3H_2O$ 80.2 mM, KH_2PO_4 20 mM, NaN_3 (Sodium azide) 0.02%. For preparing 50 ml: follow the same procedure as for the previous buffer, adjusting to pH 7.4.

ANTIBODY CROSSLINKER BUFFER (PH 7.4) Na_2HPO_4 100 mM, NaCl 100 mM. To prepare 10 ml: fill ~ 7 ml of a 10 ml Falcon tube with Milli-Q water; add 0.142 g of Na_2HPO_4 and 0.058 g of NaCl; shake well until the salts have dissolved; add Milli-Q water until the 10 ml of total volume is reached; check the pH and add NaOH until the solution reaches pH 7.4.

All salts can be acquired from any chemical distributor (e.g., Sigma Aldrich). All products have to be of a biomolecular grade of purity. The water for preparing all buffers has to be Milli-Q water.[5] For the RNA experiments, use RNAse-free water for all preparations (and make sure the pHmeter is also cleaned with RNAse-free water).

A.1.4.2   *SA beads*

For the already coated SA beads, the protocol consists in exchanging the buffer where they are preserved with PBS (pH 7.4). The procedure (~ 30 min) for the preparation of 1 ml of SA beads (which should be sufficient to perform experiments for several months) is:

1. Homogeneously resuspend the SA beads. To do so, place the container with the purchased particles into a vortex mixer for several seconds. An additional step of several seconds of sonication further improves the resuspension.

2. Take 1 ml of the resuspended SA beads and place it in a new Eppendorf tube.

3. Centrifuge the Eppendorf tube at 10000 rpm for 5 min. The dense SA beads will precipitate, while the buffer will float above them.

---

5 Milli-Q water is obtained by filtering the source water (usually distilled water) first through mixed bed ion exchange and organics (activated charcoal) cartridges, and then through a filter which removes any intact organisms. Usually a UV lamp completes the purification process.

4. Extract the overnatant buffer and add 1 ml PBS (pH 7.4). Mix the SA beads and buffer. Sonicate a few seconds and centrifuge at 10000 rpm for 5 min, again. Extract the overnatant, resuspend in 1 ml PBS (pH 7.4) and sonicate during several seconds. This step is performed to exchange the buffer.

5. Aliquote in 20 tubes (50 µl each). This allows for an optimal sterile preservation of the beads.

6. Store at 4°C for up to ∼ 6 months.

A.1.4.3 *AD beads*

The protocol for the preparation of the AD beads consists of three steps: (1) exchange the preservation buffer; (2) attach the anti-digoxigenin to the G-protein; (3) exchange the crosslinking buffer. The procedure (∼ 90 min) for the preparation of 1 ml of AD beads (which should be sufficient to perform experiments for several months) is:

1. (The first time the anti-digoxigenin batch is dissolved). Add 200 µl of PBS (pH 7.4) to dilute the dry anti-digoxigenin (sheep poly-colonal anti-dig antibody, Roche 1333 089).

2. Prepare the dimethyl pimelimidate (DMP) crosslinker buffer by mixing 50 mg DMP and 200 µl antibody crosslinker buffer (pH 7.4). The DMP crosslinker buffer needs to be freshly prepared every time the AD beads are synthesized.

3. Homogeneously resuspend the G-coated beads. To do so, place the container with the purchased particles into a vortex mixer for several seconds. An additional step of several seconds of sonication improves the resuspension.

4. Centrifuge at 5000 rpm for 2 min. The dense beads will precipitate, while the buffer float above them.

5. Extract the overnatant buffer and add 1 ml antibody crosslinker buffer (pH 7.4). Mix the beads and the buffer, sonicate for a few seconds, and centrifuge at 5000 rpm for 2 min.

6. Repeat the previous step (washing) and resuspend within 1 ml antibody crosslinker buffer (pH 7.4).

7. Add 60 µl of anti-DIG antibody and 30 µl of freshly dissolved DMP crosslinker buffer.[6]

8. Tumble at room temperature for 60 min.

9. Centrifuge at 5000 rpm for 2 min.

---

6 The volume of added anti-DIG antibody will depend on the coating of the beads: it may need to be higher if the vendor supplies the beads with a higher density of coating. Nevertheless, the proportions of anti-DIG and DMP crosslinker buffer have to be preserved (2:1).

10. Wash twice with 1 ml PBS (pH 7.0), resuspend with the same buffer, and sonicate during several seconds.

11. Aliquote in 20 tubes (50 µl each).

12. Store at 4°C for up to $\sim$ 6 months.

A.1.4.4   *Molecular buffer*

A typical molecular buffer for DNA pulling experiments consists of 10 mM Tris pH 7.5[7] (Trizma, Sigma Aldrich), 10 mM EDTA[8] (EDTA, Sigma Aldrich), 1 M NaCl, and 0.01% NaN$_3$ (sodium azide, to avoid bacterial growth). To facilitate the molecular buffer preparation, it is convenient to prepare stocks of:

1 M TRIS PH 7.5 Fill $\sim$ 40 ml of a 50 ml Falcon tube with Milli-Q water. Add 6.05 g of Tris-HCl. Dissolve using a magnetic mixer. Add Milli-Q water until reaching 50 ml. Check the pH and add a solution of 25% HCl until the solution reaches pH 7.5. For longer storage, a final auto-cleavage can be performed.

EDTA 0.5 M PH 8.0 Fill $\sim$ 40 ml of a 50 ml Falcon tube with Milli-Q water. Add 7.306 g of EDTA. Add NaOH to the solution (the EDTA does not dissolve in water if pH<7.5). Dissolve using a magnetic mixer. Add Milli-Q water until reaching 50 ml. Check the pH and add NaOH until the solution reaches pH 8.0.

5 M NACL Fill $\sim$ 40 ml of a 50 ml Falcon tube with Milli-Q water. Add 14.49 g of NaCl. Dissolve using a magnetic mixer and heat the tube to facilitate it. Add Milli-Q water until reaching 50 ml.

1% NAN$_3$ Fill $\sim$ 40 ml of a 50 ml Falcon tube with Milli-Q water. Add 0.5 g of NaN$_3$. Dissolve using a magnetic mixer. Add Milli-Q water until reaching 50 ml.

After the stocks have been prepared, for preparing 50 ml of the molecular buffer: Fill $\sim$ 30 ml of Milli-Q water in a 50 ml Falcon tube. Add 10 ml of the 5 M NaCl stock, 1 ml of the 0.5 M EDTA pH 8.0 stock, 0.5 ml of the 1 M Tris pH 7.5 stock, and 0.5 ml of the 1% NaN$_3$ stock. Add Milli-Q water until reaching 50 ml of volume. Mix the molecular buffer and filter it (Sterile Syringe Filter, w/0.2 µm Cellulose, Acetate Membrane, VWR International) introducing the filtered solution in a new 50 ml Falcon tube.

---

7 TRIS is used in the formulation of buffer solutions in the pH range between 7.5 and 8.5.

8 EDTA is widely used for scavenging metallic ions, including divalent ones, which most enzymes need to be active. For this reason, it is widely used as a food preservative or stabilizer. In our case, it inactivates DNAses and RNAses, preventing nucleic acid degradation.

A.1.4.5 *Beads incubation*

To finalize the preparation of the beads, incubate in an Eppendorf tube (~ 1.5 ml) 1 µl of a solution of the biomolecule of interest[9] mixed with 14 µl of buffer (where the experiments are going to be performed) and 5 µl of the previously prepared AD beads. After ~ 25 min have passed, add 1 ml of the molecular buffer.

For the SA beads, no incubation is required, dilute 1 µl of SA beads in 1 ml of the molecular buffer in an Eppendorf tube.

A.1.5 *Synthesis of a short DNA hairpin*

As shown in Fig. A.1(d), a typical single-molecule experiment is performed using a short DNA hairpin of a few tens of basepairs (bp). A DNA hairpin is a single-stranded DNA molecule that forms a double helix stem ended by a loop. The hairpin is flanked by two double stranded helices that act as handles. Handles are molecular spacers used to manipulate the DNA hairpin. These spacers also avoid non-specific interactions between the DNA hairpin and the beads when performing the experiments. The handles are tagged with biotin and digoxigenin at one end to specifically attach them to the coated beads. The 29-bp handles are chosen because they provide higher rigidity than longer handles and higher signal-to-noise ratio measurements [81].

To synthesize short DNA hairpins with 29-bp dsDNA handles, the desired DNA sequence (that we will denote as CD4) is purchased in a series of oligonucleotides that self-assemble into the hairpin and handles[81]. Briefly, a 29-bp sequence is used to make the handles, with the hairpin sequence flanked by the handles at both sides (Table A.1). For very short hairpins (≤ 13 bp), the hairpin sequence can be ordered as a single oligonucleotide; however, for longer hairpins it is useful to split the molecular construct into two oligonucleotides that are annealed and ligated together. Finally, the dsDNA handles are created by annealing a third oligonucleotide that is complementary to the handle region ("splint"). The sequences of the oligonucleotides used for the CD4 DNA hairpin used in this section are shown in Table A.1. One oligonucleotide (Table A.1, CD4-handleA) is purchased with a biotin at its beginning (5'-biotinylated), whereas the oligonucleotide containing the opposite handle (Table A.1, CD4-handleB) is purchased 5'-phosphorylated and then tailed at its 3'-end with multiple digoxigenins. All three oligonucleotides can be purchased from companies such as Eurofins, Fisher Scientific or Sigma Aldrich.

To complete the DNA hairpin synthesis, three main reactions have to be performed:

1. Tailing reaction: As we have already noted, one oligonucleotide is already purchased with the biotin at its 5' end. However, to be able to stretch the hairpin, digoxigenins need to be added to

---

9 Typically, the molecule is concentrated and needs to be diluted 1:10 to 1:100.

Name	Sequence
CD4-handleA	5′-biotin-AGT TAG TGG TGG AAA CAC AGT GCC AGC GCG CGA GCC ATA AT-3′
CD4-HandleB	5′-Phos- CTC ATC TG**G AAA** CAG ATG AGA TTA TGG CTC GCG ACT TCA CTA ATA CGA CTC ACT ATA GGG A -3′
splint	5′-GCG CTG GCA CTG TGT TTC CAC CAC TAA CT-3′

Table A.1: **Oligonucleotides.** They are used to synthesize the hairpin by a tailing, annealing and ligation reaction. The splint is used to the create the 29-bp double-stranded DNA handles. The hairpin loop is highlighted in bold and the sequences corresponding to the handles are highlighted in blue.

the other end of the hairpin by performing the tailing reaction described in Table A.2.

Milli-Q water	$8\,\mu l$
5X CoCl$_2$ solution*	$4\,\mu l$
5X reaction buffer*	$4\,\mu l$
10 mM dATP (Sigma Aldrich)	$1\,\mu l$
1 mM DIG-dUTP (Sigma Aldrich)	$1\,\mu l$
100 μM CD4-HandleB	$1\,\mu l$
Terminal transferase ($400\,U\,\mu l^{-1}$, Sigma Aldrich)	$1\,\mu l$
Total Volume	$20\,\mu l$

Table A.2: **Oligonucleotide tailing reaction.** Mix all the items in an Eppendorf tube and keep it at 37°C for 15 min. After the reaction is finished, add 1 to $2\,\mu l$ of 0.5 M EDTA pH 8.0 to quench the reaction. (*) Included in the Terminal transferase kit from Sigma Aldrich.

After the tailing steps, the oligos are purified using the QIAquick Nucleotide Removal Kit (Qiagen) and eluted in $50\,\mu l$ Tris-Cl 10 mM, giving rise to a final concentration, assuming a 100% efficiency, of $\sim 2\,\mu M$.

2. Annealing reaction: The final construct is then assembled in an equimolar reaction. Since the sequences are complementary, at equal concentration for all the oligonucleotides, the equilibrium structure is the formed hairpin with dsDNA handles (Fig. A.1(c)). The annealing is performed in two steps : (a) CD4-handleA + splint and CD4-handleB + splint; (b) mix of the two previous reactions. The annealing protocol is described in Table A.3.

3. Ligation: The molecular assembly is almost completed. All the Watson-Crick bonds of the structure have been formed. However, the backbone of the two oligos of the stem of the hairpin are still not covalently bonded. The final step is to connect the phosphorylated 5' end of CD4-handleB oligo with the C3' end of CD4-handleB. To do so, the assembly is then ligated in an overnight reaction

	Reaction I	Reaction II
Milli-Q water	12 μl	8 μl
DIG-tailed CD4-HandleB ($\sim 2\,\mu$M)	-	5 μl
CD4-HandleA (5 μM)	1 μl	-
splint (100 μM)	1 μl	1 μl
1 M Tris pH 7.5	0.5 μl	0.5 μl
5 M NaCl	0.5 μl	0.5 μl
Total Volume	15 μl	15 μl

Table A.3: **Annealing reaction.** Mix the items for each reaction in a separate Eppendorf tube. Keep both reactions at 95°C for 1 min, 80°C for 10 min, decrease by 0.5°C every 10 min down to 40°C. Hold the temperature, mix both reactions and decrease by 0.5°C every 20 min down to 10°C. The decrease in temperature can also be achieved by letting the tubes inside a thermal bath cooling down with the heater off; however, the usage of a thermocycler is recommended, since it allows for a better temperature control.

using T4 DNA ligase (New England Biolabs) at 16°C. The ligase is heat inactivated at 65°C for 10 min, as shown in Table A.4.

Milli-Q water	52 μl
Annealing product	30 μl
10× T4 DNA ligase buffer	10 μl
10 mM ATP	5 μl
T4 DNA ligase (400 U μl^{-1})	3 μl
Total Volume	100 μl

Table A.4: **Ligation reaction.** Mix all the items in an Eppendorf tube and keep them at 16°C overnight. Afterwards, keep the Eppendorf tube at 65°C for 10 min to inactivate the ligase.

A.1.6 *Pulling experiments*

Paradigmatic manipulation experiments one can perform with a DNA hairpin are pulling experiments. The DNA hairpin is tethered between the two beads and the force it experiences is cyclically varied between two forces, a minimum and a maximum force value. Starting at a low force (typically 0 to 5 pN) with the hairpin folded, the trap is moved away from the pipette at a constant pulling speed and the exerted force steadily increased. At a given point, the hairpin will reach a force high enough to unravel the double-stranded DNA helix structure reaching the unfolded state. This mechanically induced denaturation process is called unzipping. The stretching of the unfolded hairpin keeps on until a maximum force is reached. Afterwards, the same reverse protocol is applied: the trap is moved backwards at the same pulling speed and the applied force decreased until reaching the minimum force value. In this reverse process (called rezipping) the DNA will refold to the hairpin

native state when the applied force is low enough. Once the applied force reaches the initial minimum value a new pulling cycle starts again. The CD4 DNA hairpin is a good molecular marker that can be used for force calibration (as described in Chapter 2). It reversibly unfolds and folds at 14.7 ± 0.3pN at standard conditions ($T = 298$K, 1M NaCl).
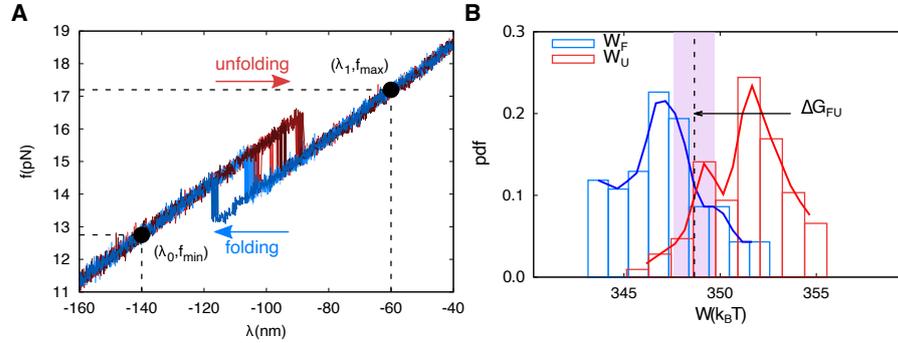
Figure A.3: **Free-energy of hairpin formation.** (a) Hairpin pulling trajectories shown as force-distance curves. The unfolding cycles are plotted in red and the folding cycles in blue. The black dots indicate the initial (λ_0, f_{min}) and final (λ_1, f_{max}) points considering for computing the work W_{FU}. Since the L values are obtained from the light position at the position PSD the λ_0 and λ_1 stand as relative trap positions with respect to an arbitrary zero (which corresponds to the center of the position PSD). (b) Probability density functions (PDFs) of the work, computed using equation (A.1), for the unfolding (W_U, red histograms) and folding (W_F, blue histograms) cycles. The value at which the probabilities cross (dashed vertical line) is the free energy difference between the folded and unfolded states ΔG_{FU}. This value is obtained as the crossing point between the lines interpolating the histograms (blue and red solid lines).

The detailed procedure to perform the pulling experiment is:

1. The three channels are filled with the molecular buffer described in section **??** through the syringes connected to the channels. The buffers have to be kept in the fridge and filtered using a 10 ml syringe and a filter (Sterile Syringe Filter, w/0.2 μm Cellulose, Acetate Membrane, VWR International) to prevent microorganism growth.

2. The chamber (already held in the metallic mount as described in section A.1.1) is placed between the objectives.

3. The buffer containing SA beads is flowed in the upper channel and AD beads in the lower channel (the channels can be interchanged).

4. The microfluidics chamber is moved using the stepped motors to bring the optical trap close to the dispenser tube of the upper supplier channel to capture a SA bead.

5. After a SA bead is captured by the optical trap, the microfluidics chamber is moved back to position the SA trapped bead on the

tip of the micropipette. The syringe is then pulled and the SA bead by air suction at the tip of the pipette.

6. The microfluidics chamber is then moved close to the dispenser tube of the lower channel supplying AD beads.

7. Once an AD bead is captured (paying special attention not to trap more than a single bead), the AD bead is brought close to the SA bead, i.e., the chamber is moved so that the optical trap is in the working zone.

8. The tether is formed by poking the SA bead held by the micropipette with the optically trapped AD bead, until a non-zero force is recorded upon retracting the AD from the SA bead.

9. Check that only one molecule has been attached to the bead. To do so, the easiest way is to move the optical trap away from the bead in the pipette and verify that a sudden decrease of the force by $\sim 1.5\,\mathrm{pN}$ occurs at $\sim 15\mathrm{pN}$ (red lines in Fig. A.3(a)), which indicates the unfolding of the DNA molecule. Then, by moving inwards the optical trap, a similar increase of the force should occur at $\sim 14\mathrm{pN}$ (blue lines in Fig. A.3(a)), which indicates that the hairpin is folding back. This equilibrium unzipping-rezipping force has been found to be $14.7 \pm 0.3\mathrm{pN}$ at standard conditions (see above).

10. Now the pulling experiment can start by cyclically varying the distance λ between the pipette-held particle and the center of the trap (as shown in Fig. A.1(d)). In this way repeated unfolding and refolding pulling cycles (~ 100) are performed over the same molecule while registering the force and the trap position. Some representative data are shown in Fig. A.3(a), with a typical pulling speed of $100\,\mathrm{nm\,s^{-1}}$. To extract the free energy of formation of the hairpin, as we will see in the next section, it is key that the hairpin is folded at $\lambda = \lambda_0$ and unfolded at $\lambda = \lambda_1$.

A.1.7 *Extracting the free-energy of hairpin formation*

A main application of single molecule experiments is the possibility to extract free energy differences from irreversible work measurements using fluctuation theorems. From the force-distance curves (Fig. A.3(a)) the work W exerted on the molecule between the starting trap position, λ_0, and the final one, λ_1 can be directly measured as the area below the curve:

$$W = \int_{\lambda_0}^{\lambda_1} f \, d\lambda. \tag{A.1}$$

For quasi static processes (i.e. sufficiently slow pulls) the work equals the free energy difference ΔG_{FU} between the initial λ_0 (folded) and final λ_1 (unfolded) trap positions. However this is not true for irreversible pulls which exhibit hysteresis between the stretching (forward) and releasing

(reverse) processes. The difference in the area between consecutive unfolding and folding cycles is the dissipated work by the system which is positive in average according to the second law of thermodynamics. Noteworthy, since we are not in the thermodynamic limit (as we are dealing with small systems), there may be trajectories that violate the second law of thermodynamics [182].

Fluctuation theorems state that the work distributions measured along the forward and reverse processes fulfil a symmetry relation [182]. A main consequence of such a relation is that forward and reverse work distributions, despite being different due to the hysteresis, they cross each other at the value of ΔG_{FU}, independently of how much irreversible the pulling process is. To extract the work distributions we use equation (A.1) to compute the work for all the forward (stretching) and reverse (releasing) cycles. The probability distributions of the work for the forward and reverse processes are shown in Fig. A.3(b) by the red and blue histograms, respectively. The crossing point of the forward and reverse work probability distributions corresponds to the difference in free energy ΔG_{FU} between the folded (λ_0) and unfolded (λ_1) states. The work histogram in Fig. A.3(b) is obtained by taking $\lambda_0 = -140\,\text{nm}$ ($f_{\text{min}} = 11.35\,\text{pN}$) and $\lambda_1 = -60\,\text{nm}$ ($f_{\text{max}} = 17.16\,\text{pN}$). Note that λ_0 and λ_1 stand as relative trap positions with respect to an arbitrary zero. The crossing point of the histograms gives $\Delta G_{\text{FU}} = 348.67\,k_{\text{B}}T$ (with $1k_{\text{B}}T = 4.114\,\text{pN\,nm}$ at $T = 25°\text{C}$).

The free energy difference between the folded and unfolded states can be written as:

$$\Delta G_{\text{FU}} = \Delta G_{\text{FU}}^0 + \Delta W_{\text{FU}}^{\text{ssDNA}} - \Delta W_{\text{FU}}^{\text{dipole}} + \Delta W_{\text{FU}}^{\text{handles}} + \Delta W_{\text{FU}}^{\text{bead}}, \quad \text{(A.2)}$$

where the term ΔG_{FU}^0 is the free energy of formation of the molecule, which is the quantity we are interested in obtaining, and the other contributions are the reversible work differences related to the different parts of the experimental setup (ssDNA, handles, bead, hairpin dipole, as shown in Fig. A.1(d)). These contributions can be computed as follows:

1. $\Delta W_{\text{FU}}^{\text{ssDNA}}$ is the reversible work needed to stretch the unfolded molecule, which is modeled with an inextensible Worm-like chain [50]:

$$f_{\text{ssDNA}}(x) = \frac{k_{\text{B}}T}{p}\left(\frac{1}{4\left(1 - \frac{x}{nl_0}\right)^2} - \frac{1}{4} + \frac{x}{nl_0}\right), \quad \text{(A.3)}$$

where f_{ssDNA} is the force, x is the molecular extension, n is the number of monomers of the molecule ($n = 44$ bases for our case), $l_0 = 0.59\,\text{nm}$ is the contour length per monomer, and $p = 1.35\,\text{nm}$ is the persistence length (parameters from Ref. [63]). The stretching contribution of this released ssDNA is then computed as

$$\Delta W_{\text{FU}}^{\text{ssDNA}} = \int_0^{x(f_{\text{max}})} f_{\text{ssDNA}}(x)dx. \quad \text{(A.4)}$$

The value we obtain is $\Delta W_{\text{FU}}^{\text{ssDNA}} = 19.0\,k_{\text{B}}T$.

2. $\Delta W_{\mathrm{FU}}^{\mathrm{dipole}}$ is the reversible work needed to stretch the folded molecule, which is modelled as a single dipole of length $d = 2\,\mathrm{nm}$[63], corresponding to the double-helix diameter[34]:

$$x_{\mathrm{dipole}}(f) = d\left[\tanh\frac{f\,d}{k_{\mathrm{B}}T} - \frac{k_{\mathrm{B}}T}{f\,d}\right] \tag{A.5}$$

and

$$\Delta W_{\mathrm{FU}}^{\mathrm{dipole}} = \Delta f \Delta x - \int_{f_{\mathrm{min}}}^{f_{\mathrm{max}}} x_{\mathrm{dipole}}(f')df', \tag{A.6}$$

where $\Delta f = f_{\mathrm{max}} - f_{\mathrm{min}}$ and $\Delta x = x(f_{\mathrm{max}}) - x(f_{\mathrm{min}})$. The value we obtain is $\Delta W_{\mathrm{FU}}^{\mathrm{dipole}} = 4.6\,k_{\mathrm{B}}T$.

3. Finally, the last two contributions to equation A.2 can be computed by considering the slope of the force-distance curves just before the rip, where the distance corresponds to the optical trap position (λ), as shown in Fig. A.3(a). The slope κ_{eff} can be easily obtained by linearly fitting the data points right before the unfolding event. This yields the effective stiffness of two serially connected springs: the trapped bead of stiffness κ_{B}, and the handles stiffness, $\kappa_{\mathrm{handles}}$: $\kappa_{\mathrm{eff}}^{-1} = \kappa_{\mathrm{B}}^{-1} + \kappa_{\mathrm{handles}}^{-1}$. Assuming that the hairpin orientation has a much higher stiffness than the handles and the optical trap, the combined work can be written as [63]:

$$\Delta W_{\mathrm{FU}}^{\mathrm{handles}} + \Delta W_{\mathrm{FU}}^{\mathrm{bead}} \approx \frac{f_{\mathrm{max}}^2 - f_{\mathrm{min}}^2}{2\,\kappa_{\mathrm{eff}}}, \tag{A.7}$$

where we have considered that the stiffness of the folded molecule (dipole) is very large and consequently does not contribute to κ_{eff}. The value we obtain is $\Delta W_{\mathrm{FU}}^{\mathrm{handles}} + \Delta W_{\mathrm{FU}}^{\mathrm{bead}} = 282\,k_{\mathrm{B}}T$.

From equations A.4, A.6 and A.7, and from the crossing point of the forward and reverse work probability distributions (ΔG_{FU}), we obtain all the terms needed to calculate the free energy of formation of the hairpin, ΔG_{FU}^0, from equation A.2. The free energy of formation at zero force we obtain for this molecule is $\Delta G_{\mathrm{FU}}^0 = 51.9\,k_{\mathrm{B}}T$, which is in agreement with that of the sequence of the CD4 hairpin, $\Delta G_{FU}^0 = 50.6\,k_{\mathrm{B}}T$, using the values provided in Ref. [91]. Typical errors of ΔG_{FU}^0 are of the order of $\sim k_{\mathrm{B}}T$, which correspond to a $5-10\%$ relative error.

DNA SYNTHESIS

B.1 SYNTHESIS OF HAIRPINS

The sequences of the oligos used for preparing the DNA hairpins are contained in tables of B.1.1. The synthesis of 120b and 204b was performed following the procedure described in [63] (with tailing, annealing and ligation steps). The procedure for the 964b and 700b is described detailedly in [183]. The synthesis of 4452b (2.2kbp hairpin) and 13680b (6.8kbp hairpin) were performed following the procedure described in [183] (adding preliminar digestion and phosphorilation steps to the previous ones).

The 1904b and 7138b synthesis were performed using the same protocol than the described in [183], only changing the first step of digestion. In the case of 7138b, the restriction enzyme used in the digestion was EcoRI (New England Biolabs) while for the case of 1904b, the restriction enzyme used is BspHI(New England Biolabs).

B.1.1 *Sequences of DNA oligos*

Name	Sequence
120b − A	5′-Biotin-AGT TAG TGG TGG AAA CAC AGT GCC AGC GCG AAC CCA CAA ACC GTG ATG GCT GTC CTT GGA GTC ATA CGC AA -3′
120b − B	5′-GAA GGA TG**G AAA AAA AAA AAA AAA AAA A**CA TCC TTC TTG CGT ATG ACT CCA AGG ACA GCC ATC ACG GTT TGT GGG TTC AGT TAG TGG TGG AAA CAC AGT GCC AGC GC-3′
204b − A	5′-Bio-AGT TAG TGG TGG AAA CAC AGT GCC AGC GCC GAC CTC T-3′
204b − B	5′-Pho-TAA CCT CCA AGC GTA CAG GGT GGA CTT TGC AAC GAC TTC CTA GAC CAA AGA CTC GCT GTT TAC GAA ATT TGC GCT CAA GCG AGA GTA TTG AA**T TTT** TTC AAT AC-3′
204b − C	5′-Pho-TCT CGC TTG AGC GCA AAT TTC GTA AAC AGC GAG TCT TTG GTC TAG GAA GTC GTT GCA AAG TCC ACC CTG TAC GCT TGG AGG -3′
204b − D	5′-Pho-TTA AGA GGT CGA GTT AGT GGT GGA AAC ACA GTG CCA GCG C -3′
splint	5′-GCG CTG GCA CTG TGT TTC CAC CAC TAA CT-3′

Table B.1: Oligonucleotides used in the hairpins that only need a tailing, annealing and ligation reaction. The splint is used to create double-stranded DNA handles (29bp). In bold, each hairpin's loop.

Name	Sequence
13680b − loop	5′-Pho-GAT CGC CAG TTC GCG TTC GCC AGC ATC CG**A CTA** CGG ATG CTG GCG AAC GCG AAC TGG C-3′
7138b − loop	5′-Pho-AAT TGC CAG TTC GCG TTC GCC AGC ATC CG**A CTA** CGG ATG CTG GCG AAC GCG AAC TGG C-3′
4452b − loop	5′-Pho-TGA TAG CCT **ACT A**AG GCT ATC ACA TG-3′
1904b − loop	5′-Pho-CAT GAC AGT CGT TAG TAA CTA ACA TGA TAG TTA C**TT TT**G TAA CTA TCA TGT TAG TTA CTA ACG ACT GT-3′
964b − loop	5′-Pho-GTC ACT TAG TAA CTA ACA TGA TAG TTA C**TT TT**G TAA CTA TCA TGT TAG TTA CTA A-3′
700b − loop	5′-Pho-GTC ACT TAG TAA CTA ACA TGA TAG TTA C**TT TT**G TAA CTA TCA TGT TAG TTA CTA A-3′
Bio − cosRshort	5′-Bio-GAC TTC ACT AAT ACG ACT CAC TAT AGG GAA ATA GAG ACA CAT ATA TAA TAG ATC TT-3′
cosRlong	5′-Pho-GGG CGG CGA CCT AAG ATC TAT TAT ATA TGT GTC TCT ATT AGT TAG TGG TGG AAA CAC AGT GCC AGC GC-3′
Bio − cosLshort	5′-Bio-GAC TTC ACT AAT ACG ACT CAC TAT AGG GAA ATA GAG ACA CAT ATA TAA TAG ATC TT-3′
cosLlong	5′-Pho-AGG TCG CCG CCC AAG ATC TAT TAT ATA TGA GTC TCT ATT AGT TAG TGG TGG AAA CAC AGT GCC AGC GC 3′
HandBio − SMFP	5′-Bio-GAC TTC ACT AAT ACG ACT CAC TAT AGG GAA ATA GAG ACA CAT ATA TAA TAG ATC TTC GCA CTG AC -3′
HandDig − SMFP	5′-Pho-AAG ATC TAT TAT ATA TGT GTC TCT ATT AGT TAG TGG TGG AAA CAC AGT GCC AGC GC -3′
splint3	5′-TCC CTA TAG TGA GTC GTA TTA GTG AAG TC-3′
inverted − splint	3′-AAA AA-5′-5′-GCG CTG GCA CTG TGT TTC CAC CAC TAA C(SpC3)-3′

Table B.2: Oligonucleotides used in the long hairpin synthesis with a digestion reaction involved. In bold, the 4 nucleotides that form the loop.

Name	Sequence
13680b − block − loop	5′-TAG TCG GAT GCT GGC GAA CGC GAA CTG GCG-3′
7138b − block − loop	5′-TAG TCG GAT GCT GGC GAA CGC GAA CTG GCG-3′
4452b − block − loop	5′-TAG TAG GCT ATC ACA TGC TGG CCA CCG GCT-3′
1904b − block − loop	5′-TTA CAA AAG TAA CTA TCA TGT TAG T-3′
964b − block − loop	5′-TTA CAA AAG TAA CTA TCA TGT TAG T-3′
700b − block − loop	5′-TTA CAA AAG TAA CTA TCA TGT TAG T-3′
204b − block − loop	5′-AAA ATT CAA TAC TCT CGC TTG AGC G-3′

Table B.3: Oligonucleotides used for blocking the loop.

MODELS DERIVATION

C.1 HELIX-COIL MODEL DERIVATION

A chain of N monomers in a thermal bath of temperature T is under tension by an external, controlled force, f. σ_i stands for the state (equivalently to the spin in the Ising formalism) of each one of the monomers. The two states to consider are hairpin-like state ($\sigma_i = -1$) or ssDNA ($\sigma_i = 1$). The hamiltonian of such a system is

$$\mathcal{H} = -\epsilon/2 \sum_i^N \sigma_i \delta_{\sigma_i}^{-1} - \int_0^f x_d(f')df' \sum_{hairpins} - \int_0^f x_1(f')df' \sum_i^N \sigma_i \delta_{\sigma_i}^1 - \gamma \sum_i^N \sigma_i\sigma_{i+1}.$$
(C.1)

Where ϵ is the free energy of basepairing base (which is only considered for the bases in the hairpin-like state), the second and third terms take into account the stretching energy contribution of all the hairpins within the secondary structure – each one contributing with an extension $x_d(f)$ – and in ideal ssDNA form –with an extension per base of $x_1(f)$–, respectively; and the last term, γ, gives the energy reward(penalty) of having neighbouring monomers in the same(different) state.

The total number of hairpins, at a given force, is equivalent to the number of domains in the $\sigma_i = -1$ state,

$$\sum_{hairpins} = \left(\frac{1 - \sigma_i}{2}\right)\left(\frac{1 + \sigma_{i+1}}{2}\right) = \frac{N}{4} - \frac{1}{4 unstacking model} \sum_i^N \sigma_i + \frac{1}{4}\sum_i^N \sigma_{i+1} - \frac{1}{4}\sum_i^N \sigma_i\sigma_{i+1},$$
(C.2)

where the two central terms cancel out, due to periodic boundary conditions, $\sigma_{N+1} = \sigma_1$. Hence, the total number of domains is:

$$\sum_{hairpins} = \frac{N}{4} - \frac{1}{4}\sum_i^N \sigma_i\sigma_{i+1}.$$
(C.3)

We then plug the expression of Equations C.3 into the hamiltonian of Equation C.2, and we obtain

$$\mathcal{H} = -\epsilon/2 \sum_i^N \frac{1 - \sigma_i}{2} - \int_0^f x_d(f')df' \left(\frac{N}{4} - \frac{1}{4}\sum_i^N \sigma_i\sigma_{i+1}\right) -$$
$$- \int_0^f x_1(f')df' \sum_i^N \frac{1 + \sigma_i}{2} - \frac{N\gamma}{4} + \frac{\gamma}{2}\sum_i^N \sigma_i - \frac{\gamma}{4}\sum_i^N \sigma_i\sigma_{i+1}.$$
(C.4)

Equation C.4 can be re-written, by grouping terms multiplying $\sum_i^N \sigma_i$ and $\sum_i^N \sigma_i \sigma_{i+1}$ in the following form:

$$\mathcal{H} = -A \sum_i^N \sigma_i - B \sum_i^N \sigma_i \sigma_{i+1} - C(f) - D, \tag{C.5}$$

where we split the terms that do not multiply any σ_i in the ones that depend on the force, $C(f)$, and the ones that do not, D. A, B, C and D can be written as:

$$A = -\frac{\epsilon}{4} + \frac{1}{2} \int_0^f x_1(f')df', \tag{C.6}$$

$$B = -\frac{1}{4} \int_0^f x_d(f')df' + \gamma, \tag{C.7}$$

$$C = \frac{N}{2} \left(\int_0^f x_1(f')df' + \frac{1}{2} \int_0^f x_d(f')df' \right), \tag{C.8}$$

$$D = \frac{N}{4}\epsilon. \tag{C.9}$$

The states contribute to the total extension as follows: a base with $\sigma_i = 1$ (ssDNA), contributes with a extension of $x_1(f)$, while all the bases in the hairpin conformation (secondary structure, $\sigma_i = -1$) do so by $N_{hairpins}x_d$. Note that This extension can be written as an extension per base (in secondary structure conformation) as $N_{hairpins}x_d/n_{-1}$, where n_{-1} is the number of bases in the -1 state. Doing so, the total extension per base, $x(f)/N$, can be written as:

$$\frac{x(f)}{N} = \frac{n_1(f)}{N}x_1(f) + \frac{N_{hairpins}(f)}{N}x_d(f), \tag{C.10}$$

where $n_1(f)/N$ is equivalent to the per base probability of being in $\sigma = 1$, $p_1(f)$, while $N_{hairpins}$ is the total number of segments with consecutive bases in $\sigma_i = -1$, i.e. the total number of domains of the chain.

Computation of $n_1(f)$

It can be shown that

$$n_1(f) = N\frac{1 + \langle \sigma_n \rangle (f)}{2}, \tag{C.11}$$

where $\langle \sigma_n \rangle (f)$ is the average state as a function of the force (equivalent to the magnetization in the regular Ising model). When $\langle \sigma_n \rangle (f) = 0$, half of the bases are in each of the two states. $n_1(f) \to 1$ when $f \to \infty$ (at high enough forces, no secondary structure is present).

Let V the transfer matrix of Equation C.5:

$$\langle \sigma_1 \rangle (f) = \frac{1}{Z} \sum_{\sigma_1...\sigma_N} \sigma_1 V(\sigma_1, \sigma_2)V(\sigma_2, \sigma_3)...V(\sigma_N, \sigma_1) = \frac{1}{Z} \sum_{\sigma_1...\sigma_N} V'V^{N-1},$$
$$\tag{C.12}$$

where

$$V' = e^{-\kappa/N} \begin{pmatrix} e^{-\beta(A+B)} & e^{\beta B} \\ -e^{\beta B} & -e^{\beta(A-B)} \end{pmatrix}, \tag{C.13}$$

with the definitions of A, B and C from Eqs. C.6, C.7 and C.8. Considering Pauli matrices, it can be shown that $V' = \sigma_z V$. Therefore we can write:

$$\langle \sigma_1 \rangle (f) = \frac{1}{Z} tr \left(\sigma_z V^N \right). \tag{C.14}$$

We then extend Equation C.14 to any site, n, using the cyclic property of the trace

$$\langle \sigma_n \rangle (f) = \frac{1}{Z} tr \left(\sigma_z V^N \right), \tag{C.15}$$

which can be written in the $|v_\pm\rangle$ basis, diagonalizing V:

$$\langle \sigma_n \rangle (f) = \frac{1}{Z} \left(\langle v_\pm | \sigma_z | v_\pm \rangle \lambda_+^N + \langle v_\pm | \sigma_z | v_\pm \rangle \lambda_-^N \right). \tag{C.16}$$

In the thermodynamic limit $Z = \lambda_+^N$ and the second term vanishes, and the magnetization can be written as:

$$\langle \sigma_n \rangle (f) = \langle v_\pm | \sigma_z | v_\pm \rangle. \tag{C.17}$$

From Equation C.17, we note that $\langle \sigma_n \rangle (f)$ only depends on the eigenvectors of the transfer matrix. Since our hamiltonian (Equation C.5) is equivalent to 1D-Ising, so are the magnetization and therefore the fraction of bases in single-stranded DNA conformation ($\sigma = 1$).

We can write the eigenvectors by defining an angle θ as

$$\tan \theta = \frac{e^{-2\beta B}}{\sinh (\beta A)}, \tag{C.18}$$

then the eigenvectors are:

$$|v_+\rangle = \begin{pmatrix} \cos (\theta/2) \\ \sin (\theta/2) \end{pmatrix} \tag{C.19}$$

and

$$|v_-\rangle = \begin{pmatrix} -\sin (\theta/2) \\ \cos (\theta/2) \end{pmatrix}. \tag{C.20}$$

By substituting Equation C.19 to Equation C.17, we obtain:

$$\langle \sigma_n \rangle (f) = \cos^2 (\theta/2) - \sin^2 (\theta/2) = \cos (\theta) = \frac{\sinh (\beta A)}{\sqrt{e^{-4\beta B} + \sinh^2 (\beta A)}}. \tag{C.21}$$

Finally, recovering the expression C.11 and plugging in Equation C.21:

$$p_1(f) = \frac{n_1(f)}{N} = \left(\frac{1}{2} + \frac{\sinh (\beta A)}{2\sqrt{e^{-4\beta B} + \sinh^2 (\beta A)}} \right), \tag{C.22}$$

where we have defined the probability per base of being in single-stranded DNA as $p_1(f)$. This probability fulfills that the equivalent one for being in the hairpin conformation, $p_{-1}(f)$ is:

$$p_{-1}(f) = \frac{n_{-1}(f)}{N} = \left(\frac{1}{2} - \frac{\sinh(\beta A)}{2\sqrt{e^{-4\beta B} + \sinh^2(\beta A)}} \right). \qquad \text{(C.23)}$$

Computation of $N_{hairpins}(f)$

The number of hairpins (i.e. the total number of domains), $N_{hairpins}$, is described in Equation C.3 and it can be written as

$$\langle N_{hairpins} \rangle = \frac{N}{4} \left(1 - \langle \sigma_i \sigma_{i+1} \rangle \right). \qquad \text{(C.24)}$$

In order to calculate $\langle \sigma_i \sigma_{i+1} \rangle$, we can proceed similarly to Section C.1. To do so,we can calculate, more generally, $\langle \sigma_n \sigma_{n+r} \rangle$:

$$\langle \sigma_n \sigma_{n+r} \rangle(f) = \frac{\langle v_+ | \sigma_z V^r \sigma_z | v_+ \rangle}{\langle v_+ | V^r | v_+ \rangle}. \qquad \text{(C.25)}$$

The computation gives

$$\langle \sigma_n \sigma_{n+r} \rangle(f) = \cos^2(\theta) + \left(\frac{\lambda_-}{\lambda_+} \right)^r \sin^2(\theta), \qquad \text{(C.26)}$$

where the second term vanishes for long distances($r \to \infty$) and the correlation goes like $\sim \cos^2(\theta)$. It is easy to prove that, after including this result for $r = 1$, Equation C.24 can be rewritten as:

$$\langle N_{hairpins} \rangle = \frac{N}{4} \sin^2 \theta \left(1 - \frac{\lambda_-}{\lambda_+} \right), \qquad \text{(C.27)}$$

which can be further simplified by using C.21. The simplification gives the final result for the total number of hairpins:

$$\langle N_{hairpins} \rangle = \frac{N}{2} \frac{\sinh(\beta A)}{\sqrt{e^{-4\beta B} + \sinh^2(\beta A)} + \cosh(\beta A)}, \qquad \text{(C.28)}$$

Computation of $n_{av}(f)$

Combining both Equations C.23 and C.28, we can write the average number of bases per hairpin, $n_{av}(f)$ as

$$n_{av} = \frac{n_{-1}(f)}{N_{hairpins}(f)} = \frac{N \, p_{-1}(f)}{N_{hairpins}(f)}. \qquad \text{(C.29)}$$

Substituting from Equations C.28 and C.22, we get:

$$n_{av} = \frac{e^{2\beta A}\sqrt{e^{-4\beta B} + \sinh^2(\beta A)} - \sinh(2\beta A)/2 + e^{-4\beta B} + \sinh^2(\beta A)}{\sinh(\beta A)\sqrt{e^{-4\beta B} + \sinh^2(\beta A)}}. \qquad \text{(C.30)}$$

Extension-Force Formula

Substituting Eqs. C.22 and Eq. C.28 to Eq. C.10, we obtain the extension per base as a function of he force:

$$\frac{x(f)}{N} = \left(\frac{1}{2} + \frac{\sinh(\beta A)}{2\sqrt{e^{-4\beta B} + \sinh^2(\beta A)}}\right) x_1(f) + \frac{1}{2}\frac{\sinh(\beta A)}{\sqrt{e^{-4\beta B} + \sinh^2(\beta A)} + \cosh(\beta A)} x_d(f),$$

(C.31)

with the parameters A and B, which are defined at the beginning of section **??**. Note that all force-dependency of Eq. C.31, relies on:

$$A = -\frac{\epsilon}{4} + \frac{1}{2}\int_0^f x_1(f')df',$$

(C.32)

$$B = -\frac{1}{4}\int_0^f x_d(f')df' + \gamma.$$

(C.33)

D

DETERMINING THE FREE ENERGY OF MISMATCHES IN DNA

D.1 SYNTHESIS OF DNA HAIRPINS
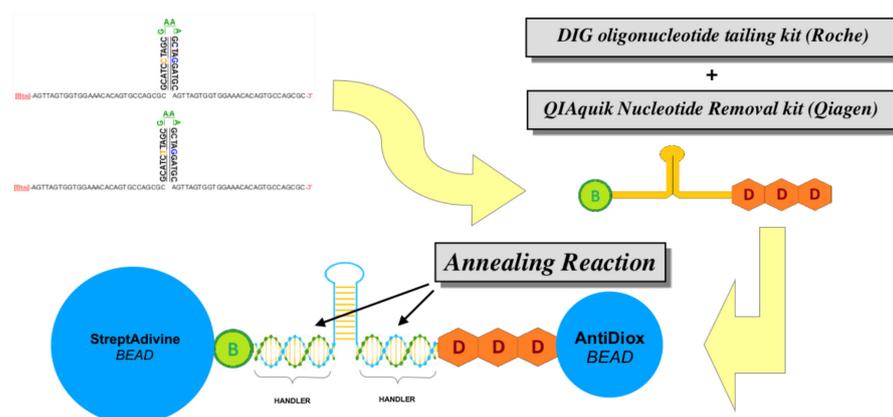
Figure D.1: **Hairpin synthesis.** Schematic representation of the procedure used to synthesize the DNA hairpins.

The synthesis of all the hairpins of 10bp and 20bp were performed in 2 general steps, represented in figure D.1:

1. Digoxigenin tailing of the 3' end: 1µl of the oligonucleotide (100µM) was tailed using Terminal Transferase from Calf Thymus, recombinant, E. coli (Roche), 1µl (100µM) of dATP (Roche), 1µl (100µM)Digoxigenin-dUTP (Roche), 5µl of 4x CoCl$_2$ solution, 4µl of reaction buffer and 8µl of water (total volume $V_T = 20$µl). The product is then purified using the Qiaquick Nucleotide Purification Kit (QIAGEN) and resuspended in 50µl of 10mM Tris pH 7.5 buffer, giving rise to a final concentration of the oligo of 2µM.

2. Annealing reaction: 10pmoles of the DIG-tailed oligo and 20pmoles of the splint were annealed in 50µl of buffer (100mM NaCl and 20mM Tris pH 7.5). The temperature of the dilution is kept 10 min at 70 °C, 10 minutes at 55 °C, followed by decreasing 1 °C every minute until 4 °C is reached.

Oligo	Sequence
10bp	5′-Biotin-AGT TAG TGG TGG AAA CAC AGT GCC AGC GCG CAT C**C**T AGC GAA AAG CTA GGA TGC AGT TAG TGG TGG AAA CAC AGT GCC AGC GC-3′
20bp	5′-Biotin-AGT TAG TGG TGG AAA CAC AGT GCC AGC GCG CGC CGC ATC **C**TA GCA TAT TAG AAA ATA TGC TAG GAT GCG GCG CAG TTA GTG GTG GAA ACA CAG TGC CAG CGC-3′
splint	5′-GCGCTGGCACTGTGTTTCCACCACTAACT-3′

Table D.1: DNA oligos used for the DNA hairpins synthesis. The base in black of the 10bp and 20bp oligos are kept for the unmodified hairpin, or changed by an Adenine (for the G-A mismatch) or by a Thymine (for the G-T mismatch).

D.1.1   *Obtaining the free energy difference between folded and unfolded states*

The free energy difference for the unfolding transition at non-zero force, $\Delta G_{NU}(f)$, is recovered from the zero-force one, ΔG_{NU}^0, by subtracting the reversible work done to orient the molecule along the pulling force axis. The contribution to this work in the force ensemble, due to the optical trap and the handles are equal for both the folded and unfolded state, hence the only free energy contributions are due to the hairpin. Depending on the hairpin state, folded or unfolded, the contributions are:

   (i) the orientation of the dsDNA of the hairpin about its main axis (Eq.A.5)

   (ii) the stretching contribution of the released ssDNA upon hairpin unfolding (Eq.4.13).

Both stretching contributions for hairpin orientation and ssDNA stretching rely on the force-extension relationships. The former is described by a freely-jointed chain (FJC) model with the Kuhn length taken equal to the hairpin diameter, $d = 2.0$nm. The latter is obtained by considering the ssDNA fragment of the number of nucleotides of the unfolded hairpin, described by a worm-like-chain (WLC) model, using the interpolation formula[50], with a persistence length $p = 1.35$nm and nucleotide unit length of 0.585nm (both parameters values coming from a best fit to the WLC curve for single-strand DNA in short hairpins[63]) The final expression for the non-zero force of the free energy of the folding-unfolding transition is then:

$$\Delta G_{NU}(f) = \Delta G_{NU}^0 - \int_0^f x_{ssDNA}(f')df' + \int_0^f x_{orient}(f')df', \quad \text{(D.1)}$$

where ΔG_{NU}^0 is obtained using the kinetic rate theory, described in the main text.
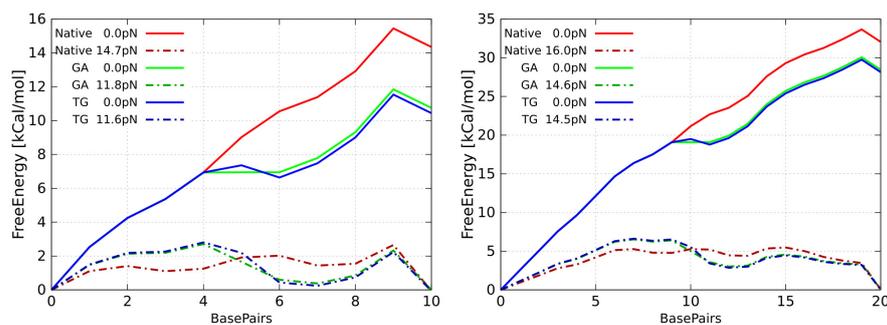
Figure D.2: **Free Energy Landscape** of the 10bp (left) and 20bp (right) DNA hairpin. Different color represent the native (red), GA mismatches (green) and GT mismatch (blue) sequence. The continuous line represent the free energy landscape obtained with the NN model at zero force considering the contribution of the tetraloop as obtained from the mfold Web Server. The dashes line represent the FEL at the coexistence force. The contribution of the unfolded bases is estimate using the WLC model for different contour length and subtracting (when at least one bp is still folded) the orientation of the dsDNA. Due to the different ΔG_{NU} the coexistence force is different for the native sequence and the mismatched sequences, this is also reflected in a different extension between the folded and unfolded state (see Tables 9.1-9.2)

D.2   HIDDEN MARKOV MODEL

In the hopping experiments analysis we extract information on the hairpin states and transition kinetics using the Hidden Markov Model (HMM)[184]. The hidden Markov model is a probabilistic method for the study of time series. It quantifies the likelihood of a model to interpret a series of data and using the iterative Baum-Welch algorithm it is possible to fit the parameters of the model and maximize the likelihood.

The initial parameters of the model are the number of hidden states, the initial probability $\{w_n^0\}$ associated with each state and the transition matrix T_{nm}. The matrix describes the probability that a state m after one step jumps to state n and it is taken as time homogeneous. The transition rate between states could then be read from the off-diagonal elements of the matrix (eventually dividing by the time-step if we want to switch from discrete to continuous time). Given the finite series of force observations $F = (f_1, f_2, f_3, ..., f_T)$ taken from the instrument at a fixed sampling rate, we start from the tentative hypothesis that the system can occupy two states with Gaussian distributed forces (Suppl.Fig.D.3-D.5). This introduces two additional parameters in the algorithm for each hidden state: the mean $\bar{f}_n$ and standard deviation $\bar{\sigma}_n$. By means of successive iterations, the maximum likelihood of the model can be computed, and the parameters describing the series F can be optimized.

The two-state model is validated *a posteriori* by looking at the experimental force distributions (e.g. in Fig.9.2a and Suppl.Fig.D.3-D.5) and comparing the likelihood of the two states model ($N = 2$) with a model with $N > 2$ states. In fact, no hidden states beyond the two folded/unfolded configurations emerged from this analysis. The optimized parameters give an estimation of the average force in the folded/unfolded state with the associated standard deviation, the kinetic rate coefficients k^+, k^- of the folding/unfolding process, and the probability w_U, w_N of either state in the hopping experiment, at a given trap position λ.

Repeating this analysis for different λ, see Suppl.Fig.D.3-D.5, one can extract the force dependence of the kinetic rates.
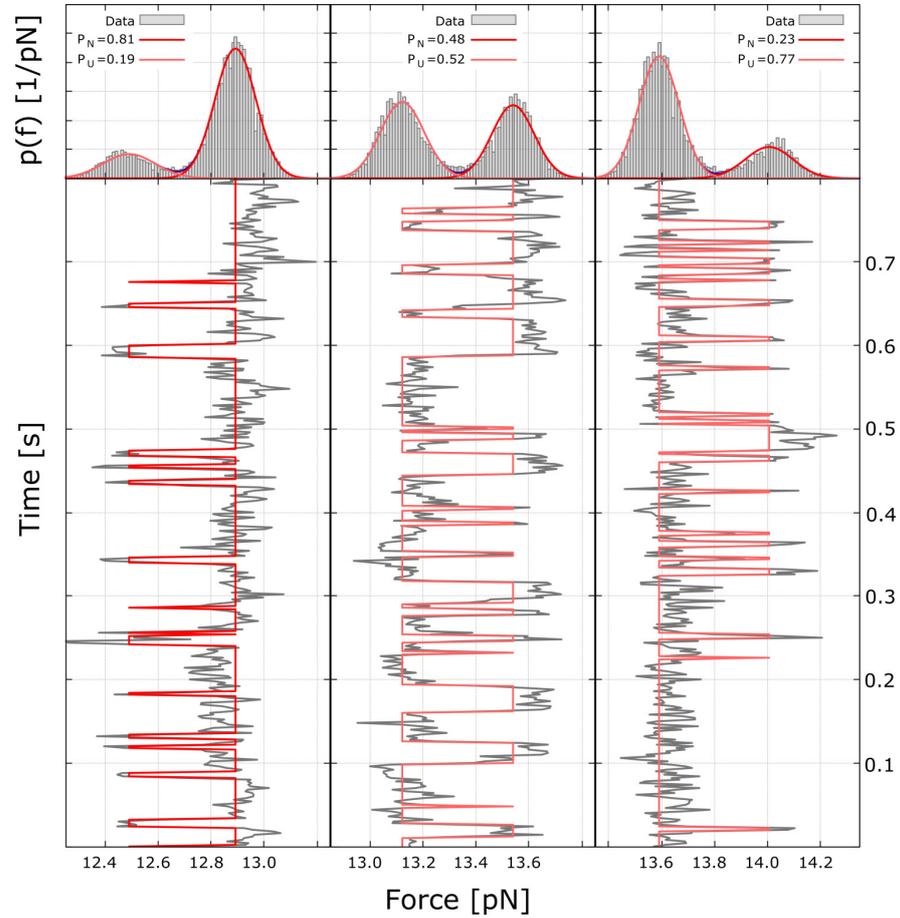
Figure D.3: **Example of HMM analysis in the 10bp native hairpin**.
Experimental force-time traces (gray) for the 10pb native hairpin during
hopping experiments with LOT and the corresponding force histogram (gray
bars). In the force time series, hidden by the noise, we observe multiple jumps
between the two states as determined by the two force levels evidenced in the
force histogram Gaussian peaks (upper panels). These two peaks correspond
to the folded (higher force) and the unfolded (lower force) states of the hairpin.
The HMM extracts information about the hopping kinetics by determining
the optimal trajectory for a given time series (shown in red overlapped to the
original grey signal). The two optimized Gaussian force distributions are shown
in the upper panels for the P_N folded (bright red) and P_U unfolded (pale
red), and compared with the experimental force histograms (grey bars). This
analysis is repeated for different values of λ (increasing from left to right). Here
is shown $0.8s$ of the whole 15-$30s$ force time series for three conditions close
to the coexistence force: $P_N > P_U$ (left); $P_N, P_U \simeq 0.5$ (middle); $P_N < P_U$
(right). The values of P_N, P_U are shown as legends in the upper panels.
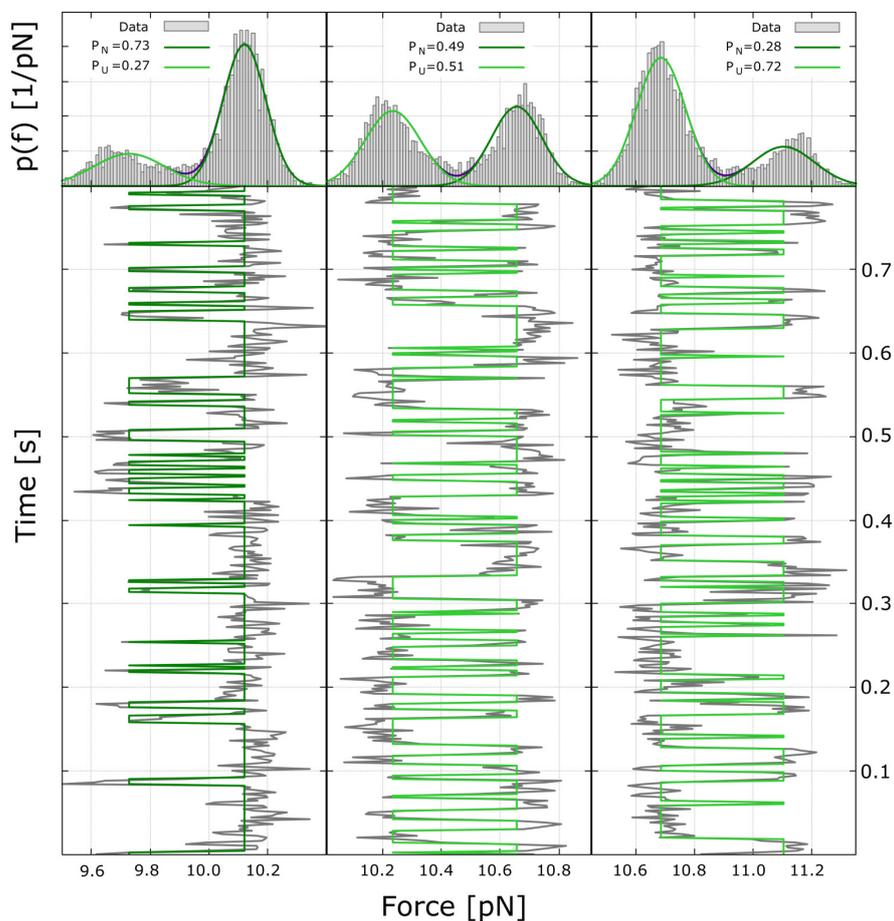
Figure D.4: **Example of HMM analysis in the 10bp hairpin with GA mismatch**. Same as in Suppl. Fig.D.3 for the 10bp hairpin with a GA mismatch, the main difference is in the color of the curves obtained from the HMM that are now green according to the color code used in Fig.9.2. Comparing these plots with the previous ones in Suppl.Fig.D.3 we observe that the value of the average hopping force at coexistence (middle panels) changes from $\sim 13.4\mathrm{pN}$ in the native sequence to $\sim 10.4\mathrm{pN}$ in the mismatch GA sequence.
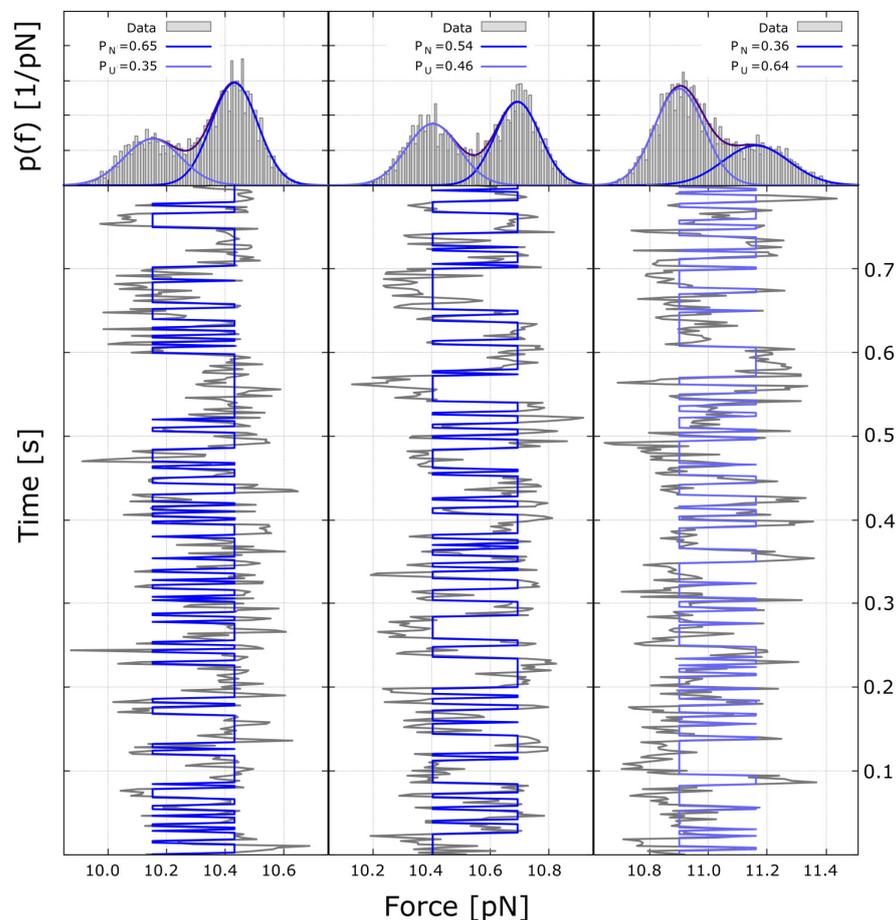
Figure D.5: **Example of HMM analysis in the 10bp hairpin with GT mismatch defect.** Same as in Suppl. Fig.D.3 for the 10bp hairpin with a GT mismatch defect, the color of the curves obtained from the HMM is blue in agreement with the color code used in Fig.9.2. Comparing these plots with the previous ones in Suppl.Fig.D.3 we observe that the value of the average hopping force at coexistence (middle panels) changes from ~ 13.4pN in the native sequence to ~ 10.6pN in the mismatch GT sequence. Such a difference cannot be appreciated comparing with the mismatch GA sequence (Suppl.Fig.D.4).

D.3   PULLING EXPERIMENTS 20BP HAIRPIN.

The data analysis of force pulling experiments on the 20bp hairpin is done in four steps:

(i) **Drift suppression.** The force pulling trajectories are aligned to eliminate drift effects. Drift effects are small in a single pulling cycle, however they are not along the entire measurement process (typically lasting for 50-100 cycles) as drift tends to accumulate with time leading to net shifts in FDC of a few nm. We used an alignment software that substantially compensates for such accumulated drift. Examples of aligned trajectories are plotted in the upper part of Suppl. Fig. D.6-D.8.

(ii) **Rupture force measurements.** Each unfolding trajectory is checked to identify the value of the force (first-rupture force) at which a force rip is observed for the first time along the unfolding trajectory (red trace in the upper panels of Suppl. Fig. D.6-D.8). Analogously, for each refolding trajectory we identify the value of the force (first-folding force) at which a positive force jump is observed for the first time along the refolding trajectory (blue trace in the upper panels of Suppl. Fig. D.6-D.8).

(iii) **Rupture force distributions.** First-rupture and first-folding forces are collected among the recorded trajectories to extract rupture-force distributions (central panels in Suppl. Fig. D.6-D.8) for unfolding (red bars) and folding (blue bars).

(iv) **Kinetic rates.** From the rupture force distributions and using Eq.9.5 we extract the kinetic rates for the unfolding and refolding process (bottom panels in Suppl. Fig. D.6-D.8). The coexistence force is identified by the value of the force at which the unfolding and folding rates are equal.

The analysis is repeated for different molecules of a given sequence (native, GA, GT) pulled at 100nm/s. Rupture force distributions and kinetic rates are averaged over the molecules, the results are shown in Fig.9.3 (main text). In figure Suppl. Fig. D.6-D.8 we present results obtained for three different molecules pulled at 100nm/s. Pulling experiments were repeated at different pulling speeds (100,200,300 nm/s, Suppl. Fig.D.9). While the hysteresis between unfolding and refolding trajectories increases with the pulling speed, the estimated values of the coexistence forces do not vary substantially much, as expected.
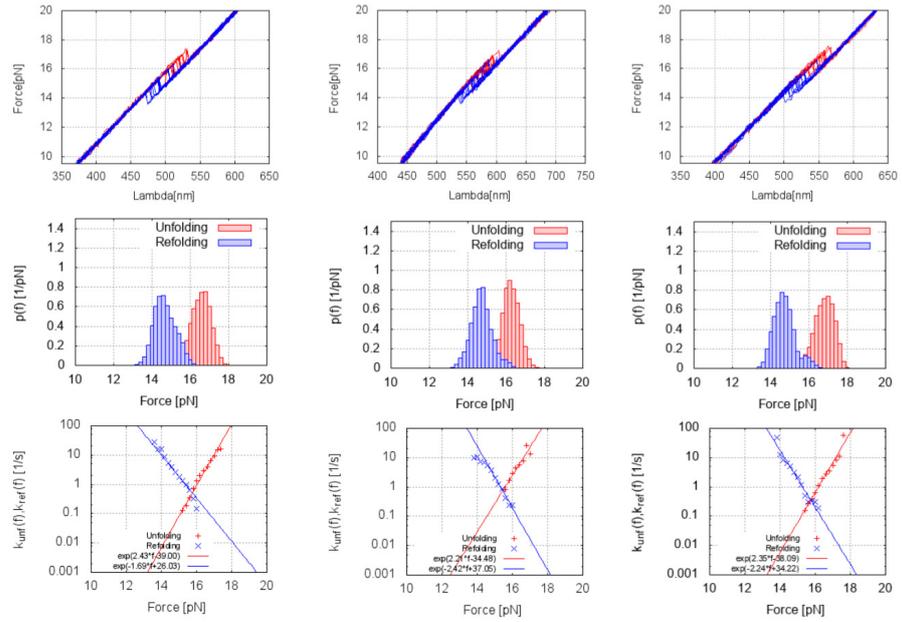
Figure D.6: **Pulling experiments on three molecules of the 20bp hairpin with native hairpin sequence, at pulling speed 100 nm/s**. Upper row: FDC (red trajectory for unfolding, blue trajectory for refolding). Middle row: the corresponding rupture force distribution histograms. Bottom row: force-dependent reaction rates $k^+(f)$ (unfolding, red) and $k^-(f)$ (folding, blue).
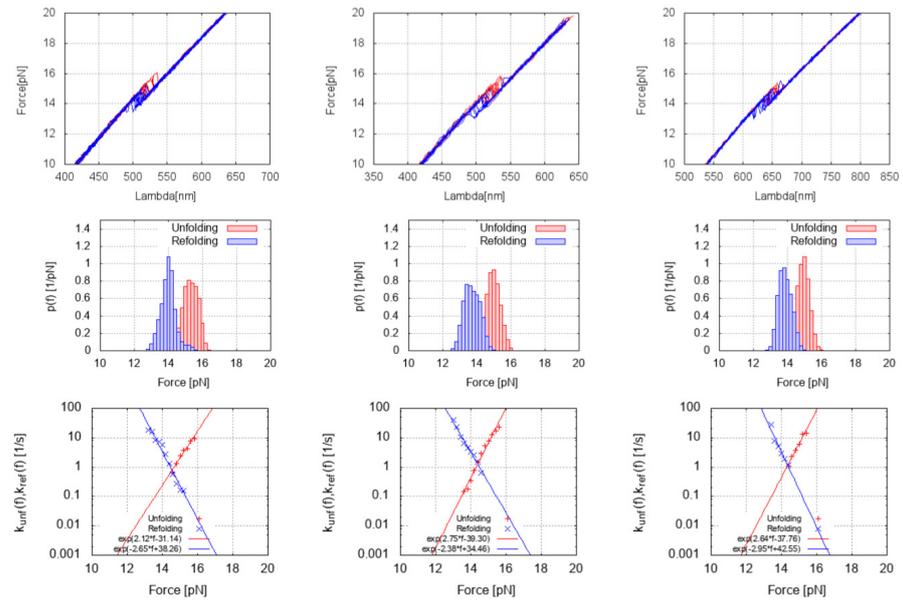

Figure D.7: **Pulling experiments on three molecules of the 20bp hairpin with GA mismatch hairpin sequence, at pulling speed 100 nm/s.** Rows and panels are defined as in Suppl. Fig. D.6.
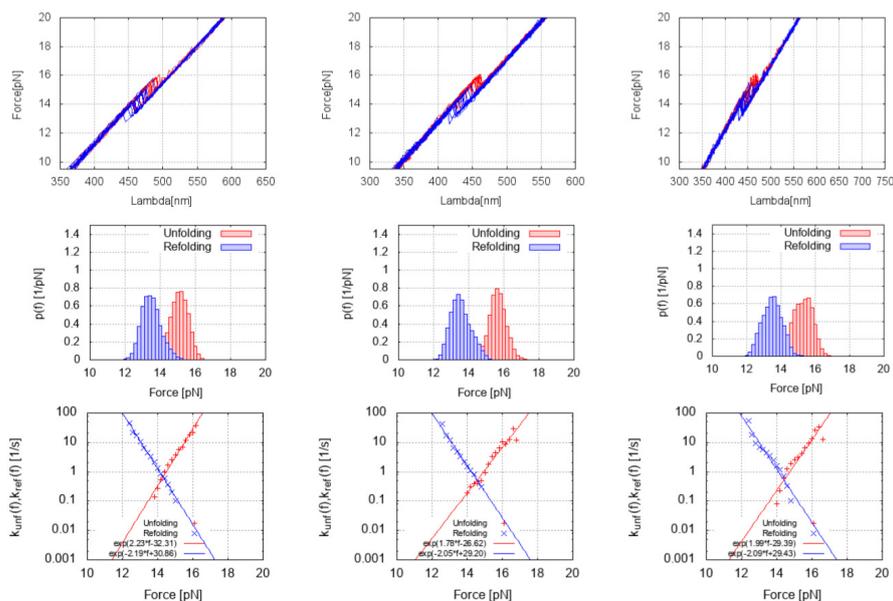
Figure D.8: **Pulling experiments on three molecules of the 20bp hairpin with GT mismatch hairpin sequence, at pulling speed 100 nm/s.** Rows and panels are defined as in Suppl. Fig. D.6
.
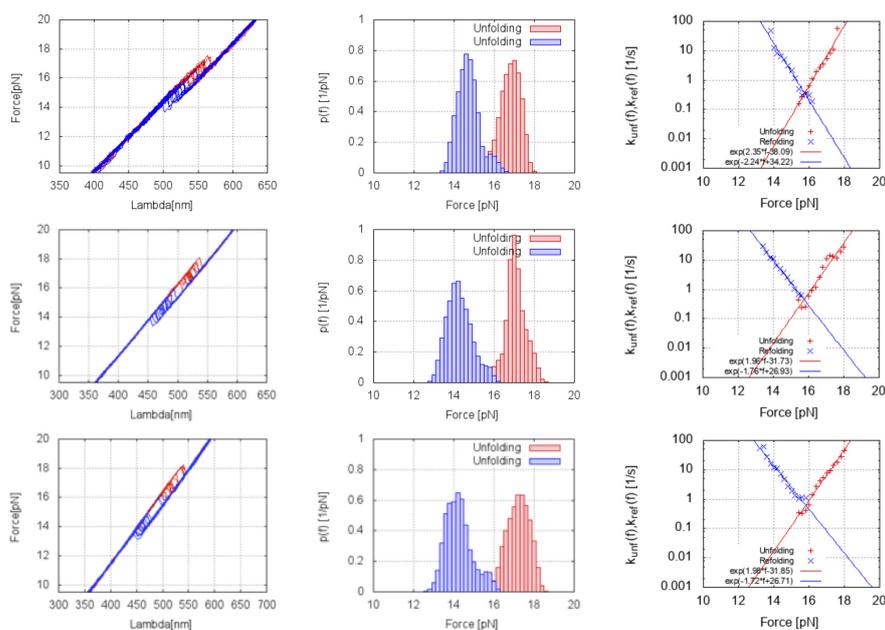
Figure D.9: **Summary of force-pulling experiments for the native 20bp hairpin, at different pulling speeds.** From top to bottom, $v_{pull} =$100, 200, 300 nm/s. Left column: unfolding(red) and refolding (blue) trajectories; each experiment corresponds to about 100 cycles; the rupture force corresponds to the nearly vertical jumps between each pair of trajectories. Central column: histograms of the first-rupture force for unfolding (red) and folding (blue) trajectories; for clarity, the data are convoluted with a Gaussian smearing function with $\sigma =$0.2. Right column: unfolding (red) and folding (blue) kinetic rates, $k^+(f), k^-(f)$; crosses are the experimental data, straight lines are exponential fits according to the Bell-Evans model.
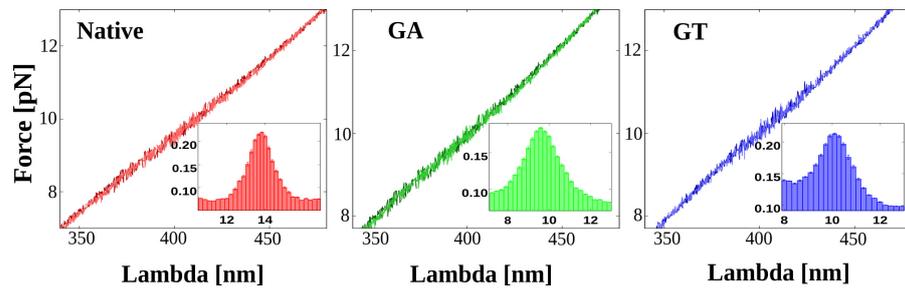
Figure D.10: **10bp hairpin pulling experiments.** Example of a pulling trace of the 10bp hairpin at constant pulling speed 50nm/s for the native hairpin (red), G-A mismatch (green) and G-T (blue). The unfolding trajectory is represented in dark colors and folding trajectory in light colors. Compared to previous pulling experiments for the 20bp, we observe a smaller force rip and faster unfolding and folding kinetic rates (with multiple folding-refolding transitions being observed for each single trajectory). This makes rupture force events noisy and difficult to analyze. In the inset it is possible to observe an increase of the mean square fluctuation of the force close to the coexistence value. This shows the existence of rapid transitions between the folded and the unfolded state even if the noise in the FDC hide the sudden jumps observed in the 20bp hairpin. The coexistence force values obtained in hopping experiments for the 10bp hairpin (main text) are in agreement with the force values at which the mean square fluctuation of force is maximum.

E

ELASTICITY OF SSDNA

E.1 OBTAINING ssDNA

E.1.1 *Short hairpins*

The measured position of the optical trap in the folded branch, λ_F, can be written in terms of the extension of the handles, x_h, the displacement of the bead with respect to the optical trap, x_b, and the extension associated to the orientation of the hairpin, x_d, and the correction due to the drift, λ_0, as:

$$\lambda_F(f) = x_h(f) + x_b(f) + x_d(f) + \lambda_{0,F}. \tag{E.1}$$

$x_d(f)$ is described as a FJC with equal Kuhn and contour length of 2nm (App. E.4).

An equivalent expression to equation E.1 can be written for the unfolded branch, λ_U:

$$\lambda_u(f) = x_h(f) + x_{b(f)} + x_{\text{ssDNA}}(f) + \lambda_{0,U}. \tag{E.2}$$

Where $x_{\text{ssDNA}}(f)$ stands for the extension of the unfolded hairpin ($2n$ bases in the stem and a loop of 20 bases). Subtracting equations E.2 and E.1, we can then write

$$x_{\text{ssDNA}}(f) = \lambda_F(f) - \lambda_U(f) + x_d(f) + \lambda_0(t), \tag{E.3}$$

where $\lambda_0(t)$ stands for the drift correction, which has only a temporal dependence after aligning both cycles at the aligning force, $\lambda_{0,U} = \lambda_{0,F}$, as described in App. E.2. For the case of the splint blocking oligo, the position of the optical trap for the folded and unfolded branches (schematically depicted in figure 7.5A) are written as:

$$\lambda_F(f) = x_{h1}^{dsDNA}(f) + x_{h2}^{29,ssDNA}(f) + x_b(f) + x_d(f) + \lambda_{0,F}. \tag{E.4}$$

$$\lambda_U(f) = x_{h1}^{dsDNA}(f) + x_{h2}^{n,dsDNA}(f) + x_{h2}^{29-n,ssDNA}(f) + x_b(f) + x_{\text{ssDNA}}(f) + \lambda_{0,U}. \tag{E.5}$$

Subtracting equations E.5 and E.4 we can then write the extension of ssDNA as:

$$x_{\text{ssDNA}}(f) + x_{h2}^{n,ssDNA}(f) = \lambda_F(f) - \lambda_U(f) + x_d(f) - x_{h2}^{n,dsDNA} + \lambda_0(t), \tag{E.6}$$

E.1.2 *Long hairpins*

In the case for long hairpins (HNL4, 4 bases in the loop), the trap position, $\lambda(f)$, is written as:

$$\lambda(f) = x_{\text{ssDNA}}(f) + x_h(f) + x_b(f) + x_0. \tag{E.7}$$

$x_{\text{ssDNA}}(f)$ is the molecular extension of the stretched ssDNA, $x_h(f)$ is the extension of the two 29bp handles (characterized in [81]), $x_b(f)$ is the displacement of the bead from the center of the optical trap and x_0 is the shift of the trap position relative to the position detector one. This shift is corrected by imposing that the position of the trap has to be zero at zero force ($\lambda(f = 0) = 0$nm). In order to achieve it, we use the unzipping curve to align the rest of the ssDNA curves to it (as shown in figure E.1). The value of x_b can be obtained by assuming an
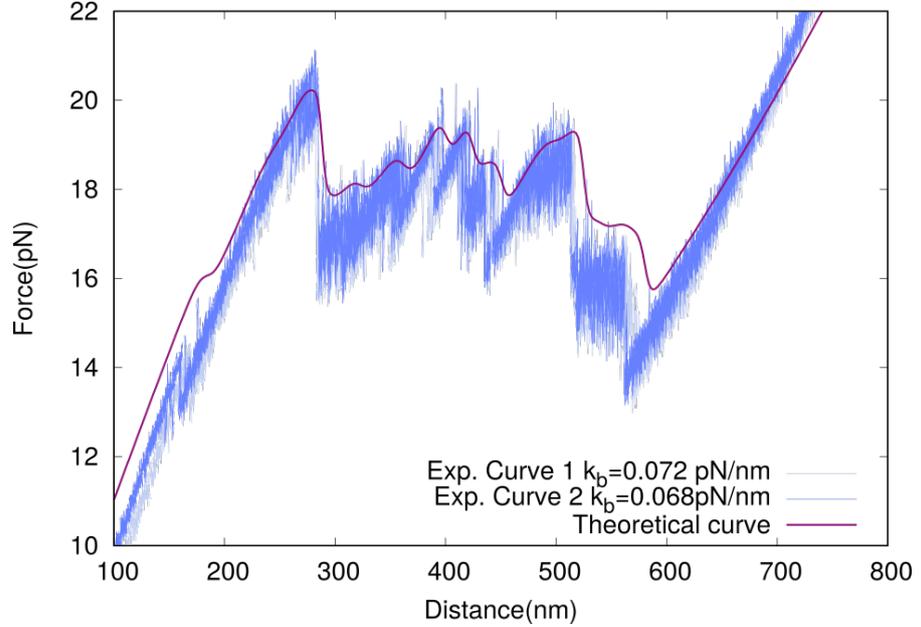
Figure E.1: Two experimental force-distance unzipping curves (each one with a different measured trap stiffness k_b) compared to a theoretical one. In both cases, the agreement between the theoretical rips and the experimental ones guarantee the correct tether of the molecule. A special attention is paid in its beginning, since if there is an incorrect molecular tethering, initial rips may be missing. For obtaining the theoretical curve, we use the elastic parameters from [61] and a standard bead stiffness of $k_b \approx 0.07$pN/nm).

harmonic potential of the optical trap: $f = k_b \cdot x_b$. Hence, by assuming that the trap stiffness, k_b, is constant through the range of forces of study ($0 < f < 40pN$), one can obtain the value of k_b from the power spectrum of an untethered bead at zero force [16] (as shown in figure ??).

Typical values in our experiments for k_b range from 0.06pN/nm to 0.09pN/nm. After the breaking of the molecule, 5s of recording at 40kHz of the force signal is recorded, from which a different trap stiffness for each of the molecules is obtained.
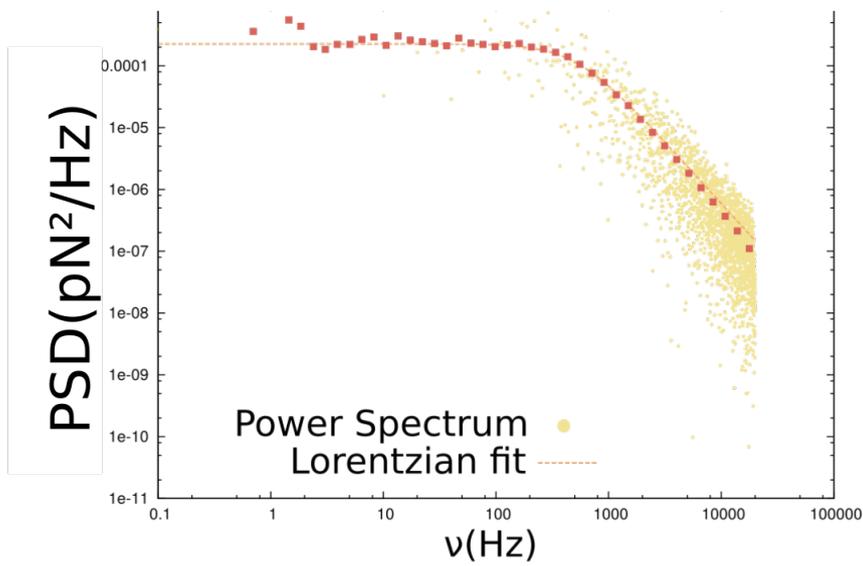
Figure E.2: Log-log representation of the zero force power spectrum of an untethered bead. In this case, the value for the trap stiffness is $k_b = 0.088 \pm 0.004$pN/nm

E.2   DRIFT CORRECTION

The aligning of the folding and unfolding cycles is critical for the determination of the molecular extension of short molecules. The phenomenon of the apparent movement of the optical trap at a certain force is known as drift. This drift of the instrument is a low frequency systematic deviation of the measured trap position due to macroscopic effects, such as temperature changes. The difference of the trap position between the unfolded and folded states $(\lambda_U - \lambda_F)$ is used to determine the molecular extension. The usual aligning method consists on imposing a constant shift λ_0 in each of the force-λ curves to ensure the collapse between the unfolding and refolding curves. In our case, we take a reference curve (an stretching, unfolding one, such as the one shown in magenta in figure E.3A) for each molecule and find the position of the trap, λ_{ref}, around a certain force (black points in the magenta curve) where the hairpin is always unfolded ($\approx$ 20pN, without blocking splint oligo and $\approx$ 50pN with it). Then we calculate the trap position of each of the aligning curves, λ_{align} (curve in green in E.3), around the aforementioned force and calculate the shift that needs to be implemented in the curve in order to match the trap position of the reference curve, $(\lambda_0 = \lambda_{align} - \lambda_{ref})$. In this work we include a temporal correction in
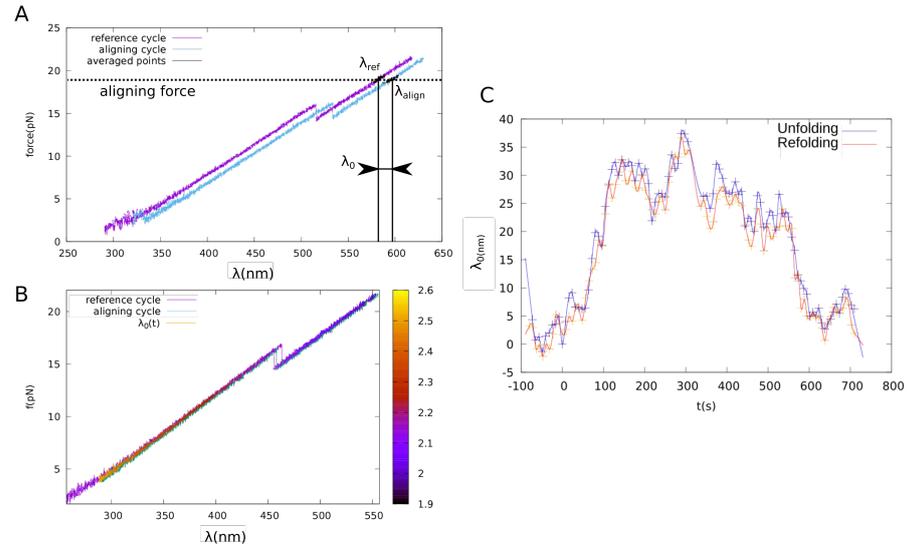
Figure E.3: **Drift correction of short molecules**(A) The trap position, λ_{ref}, around an aligning force of reference cycle(magenta) is determined. The trap position, λ_{align}, of the aligning cycle (green) is determined around the same force. The drift is determined as $\lambda_0 = \lambda_{align} - \lambda_{ref}$. (B) The reference and aligned cycles are plotted, the latter with the colour code taking into account the corrections due to the spline interpolation obtained from (C). (C) Drift determination of the unfolding (blue) and refolding (red) cycles (crosses) with respect to the time (taking as t=0 at the reference cycle). The spline interpolation is the solid line connecting each point.

the drift of a single trajectory. This temporal correction is achieved by interpolating the found correction λ_0 using a spline interpolation

of the different shifts imposed for the unfolding and refolding curves as a function of time, similarly to what is done in [82]. This allows us to achieve sub-nanometric corrections inside a cycle, as shown in figureE.3B. This is critical due to the length of the short molecules we are pulling (with a total length that is around 40nm) and because of the strong dependence of the elastic parameters we are obtaining with the changes of curvature that the drift may cause. A typical plot of the spline interpolation is shown in figure E.3C.

E.3   BLOCKING OLIGONUCLEOTIDE EFFECTS ON THE ELASTICITY 204B SSDNA MOLECULE

The effect on the elastic properties of a dsDNA segment of $\approx$ 25bp – created by the blocking oligo – is expected to be negligible for long enough ssDNA molecules [61]. Nevertheless, this effect should increase when the number of bases of the molecule at which the blockin oligo is attached decreases. In order to check this effect, we obtained an average FEC for the H100L4 with and without the blocking oligo. In the case for the FEC obtained without the blocking oligo method, the first part of the trajectory is not shown, since it corresponds to the unzipping of the hairpin. Despite the unzipping curve that is obtained for the case of the FEC without blocking oligo, the final rip of the unzipping leaves the molecule fully unzipped at $\approx$ 14pN, allowing us to fit the elastic parameters for the same range of forces than in the rest of the molecules studied with the blocking oligo technique. As it is shown in figure E.4 both FEC and the fits for the persistence and contour length are compatible and within error bars. Since H100L4 is the shorter of the molecules studied with the blocking oligo technique, and no effect on the elastic properties is observed, it is possible to conclude that the effect of the segment of dsDNA for hairpins HNL4, with $N > 100$ should also be negligible.
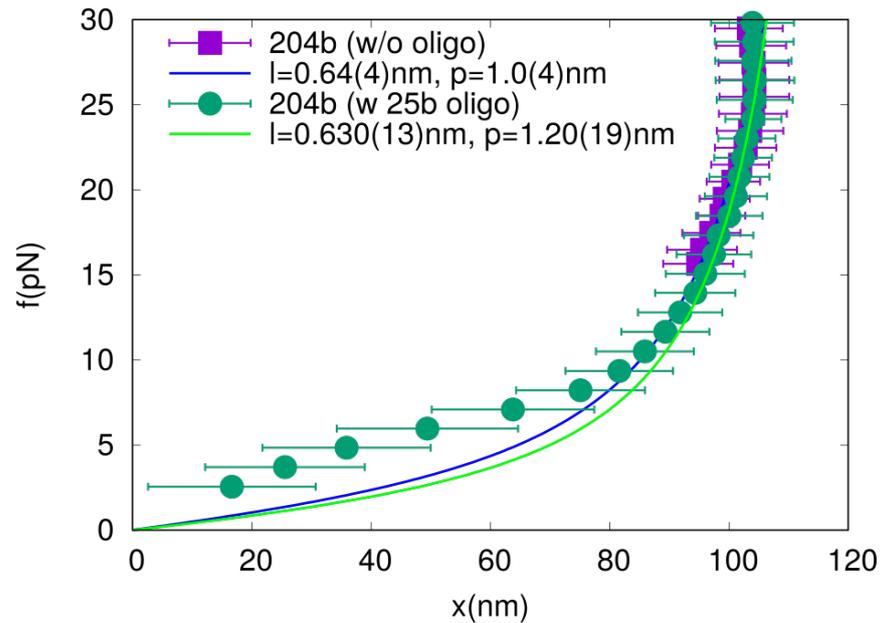
Figure E.4: Average FECs for H100L4, with (green) and without (magenta) the blocking oligo (25 bases). The values for the contour length per base, l, and the persistence length, p of equation 4.13 (inextensible WLC) that best fit the FEC with (blue) and without (orange) blocking oligo are shown in the key.

E.4 ELASTICITY OF THE HAIRPIN ORIENTATION

The contribution of the hairpin orientation can be modeled as a dipole in a magnetic field ([63]):

$$x_d(f) = d \left[coth\left(\frac{f\,d}{k_B\,T} \right) - \frac{k_b\,T}{f\,d} \right], \tag{E.8}$$

where d stands for the diameter of the dsDNA helix ($d = 2.0$nm [34]).

E.5   EFFECTS OF THE ELASTIC MODEL USED FOR SSDNA ON THE BASE-PAIR FREE ENERGY OF FORMATION OF DSDNA

The differences in the energetics given by different elastic parameters or by using different model modify by $\sim 10\%$ the base-pair free energy of formation reported in [82], as shown in figure E.5.
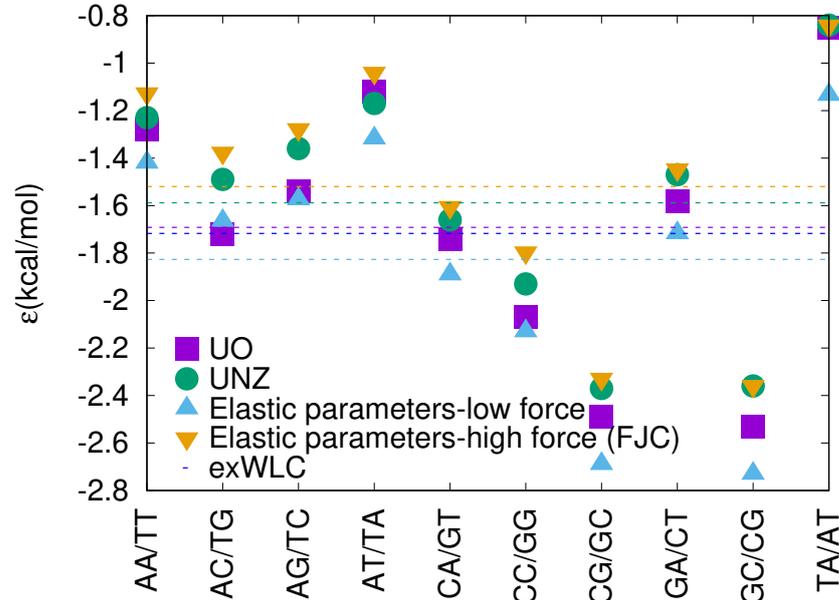
Figure E.5: Base-pair free energy from melting oligonucleotides [108] (square, magenta) and unzipping experiments[82] (green circles). The rhombus points are the modification in energy of the green points when using the low-force ($f < 15$pN) elastic parameters (Figure 7.4B) when compared to the Freely-Jointed Chain (cyan) or the WLC (orange) at high forces ($f > 15pN$). The dashed blue line corresponds to add the constant free energy difference between using the extensible WLC parameters obtained fitting h20L20 ($5f < 40$pN) and the inex.WLC ones at low force (green). The difference in free energies between models and ranges of forces is evaluated at 15pN.

SECONDARY STRUCTURE

F.1 EFFECT OF THE DIAMETER IN THE SECONDARY STRUCTURE PARAMETERS

The projected extension of a single hairpin along the stretching direction o orientation is described in App. E.4, and it requires of a diameter, d, which is usually taken as 2nm. However, in the hamiltonian of the model(Eq. 7.2), at high forces, this diameter results in a minimized energy with a configuration of alternating bases in secondary structure and ideal ssDNA ($\sigma = +1, -1, +1, -1, +1...$). This is due to the fact that each domain in secondary structure (i.e. hairpin) is modelled as a hairpin, which is longer than a single base in ideal ssDNA form, and the configurations that maximizes the extension at high forces (i.e. minimizes the energy). To prevent that from happening without having to add more constraints in the model (which would complicate its expression), we have considered a hairpin diameter $d = 0$nm, which is equivalent to consider the secondary structure domains do not contribute to the extension of the molecule. In order to check the impact of this approximation, we have studied the effects on the fitted parameters of the model (ϵ and γ) for varying d. To avoid the above-mentioned problem at high forces, we have restricted this check within the range of d for which $x_h \leq x_1$, that is $0 \leq d \leq 0.6$nm. As shown in Fig. F.1, the values obtained for ϵ do not vary substantially for any of the hairpins studied. Actually, the fitted values for ϵ and γ increase around $\sim 3\%$, between the fitted values for $d = 0$nm, and $d = 0.6$nm. Extrapolating the differences we obtain between considering ($d = 0.6$nm and $d = 0.0$nm) and assuming it follows a linear behaviour, we expect to be understimating the values obtained for ϵ and $\gamma \sim 10\%$, which fall inside the uncertainties obtained.
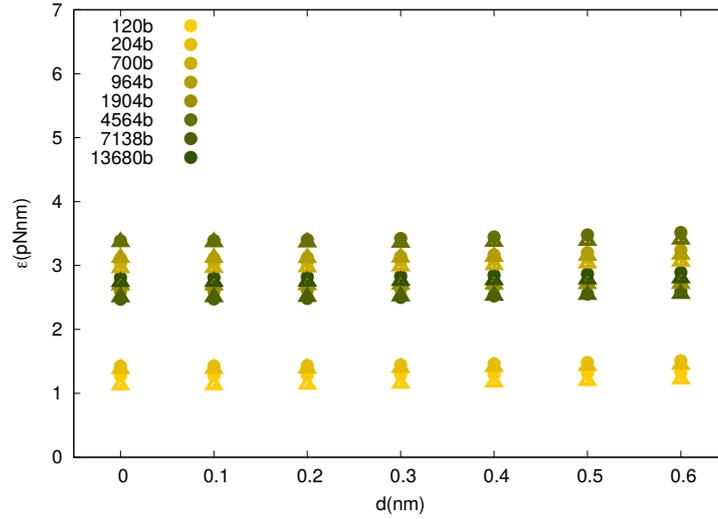
Figure F.1

F.2   UNCERTAINTY CALCULATION

In order to compute the uncertainties of the fitted parameters, we have considered the parametric region for which the χ^2 value varies within 10% of variation with respect to the minimum. The uncertainties obtained are comparable with those obtained from other methods. For the NaCl case, the logarithmic dependence of ϵ with respect to the salt concentration (fitted at [NaCl]$\geq$ 100mM) is within the error bars of the values of ϵ obtained for fitting at lower salt concentrations (10,25,50mM). The constant value of γ is compatible with the fits of the same low salt concentrations.
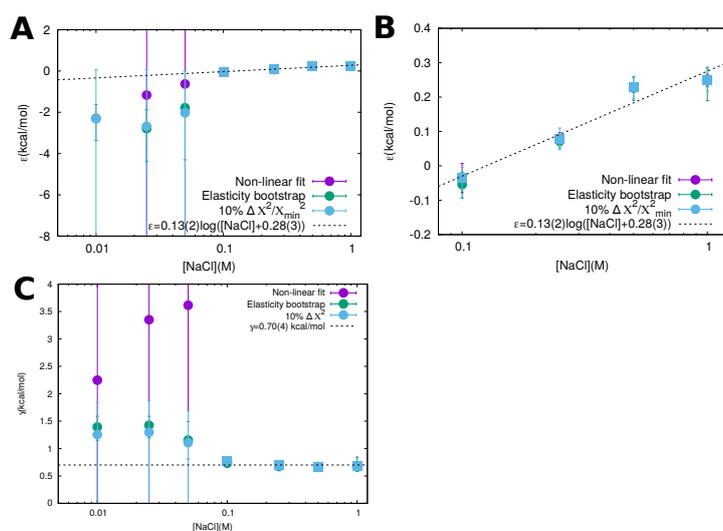
Figure F.2: Obtained fitted values for NaCl and uncertainties using 3 methods: non-linear fit (performed by MATLAB), bootstrap method using a homogeneously distributed values of p and l within the error given in [61] and a 10% variation around the minumum number of the χ^2. (A) The obtained value for ϵ for varying [NaCl]. (B) A zoom for the salt concentrations which we have considered show a strong enough deviation to consider them for fitting a logarithmic dependence with respect to the salt concentration. (C) Values of γ obtained using the same three methods described above for different salt concentrations.

F.3   SEQUENCE EFFECTS

The sequences studied in Chapter 7 have different GC content, which varies from $\sim 44\%$ to $\sim 54\%$. In order to discard that the observed dependence on the length is due to the GC content of the sequences studied, we show the dependence of the basepairing energy, ϵ, obtained from fitting the FECs for each molecule as a function of their GC content.
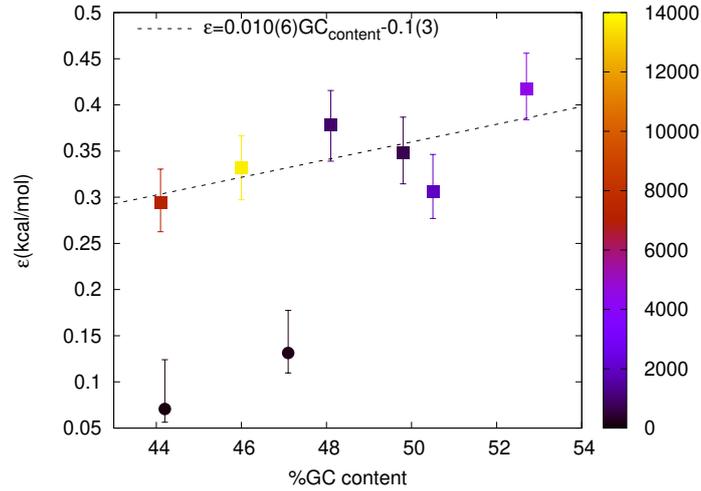
Figure F.3: Obtained values of the average basepairing energy of the secondary structure, ϵ, for the 8 sequences studied. The color of the points indicate the length of the molecule, from short (darker-violet) to long yellow-orange (right panel). The linear fit (dashed lines) has been performed considering only the longer molecules ($700b \leq N \leq 14kb$, squares) and it shows a slight increase ϵ for GC-rich molecules, as expected from the higher stability provided from GC bp compared to that of AT bp (about to times higher free-energy of base-pairing).

As it is shown in Fig. F.3, the correlation between the GC content of the shorter molecules (120b and 204b) and their fitted smaller ϵ does not correlate with their GC content. Nevertheless, a slight trend of increasing ϵ as a function of the GC content of the larger sequences is observed, as shown in the dashed linear fit ($r = 0.6533$).

ACRONYMS

AD	Anti-Digoxigenin
AFM	Atomic Force Microscopy
BSOM	Blockin-Split Oligo Method
CCD	Charge-Coupled Device
dsDNA	double-stranded dsDNA
EDTA	Ethylenediaminetetraacetic Acid
FDC	Force Distance Curve
FEC	Force Extension Curve
FJC	Freely-Jointed Chain
FFT	Fast Fourier Transform
IUPAC	International Union of Pure and Applied Chemistry
NA	Nucleic Acids
NESS	Non-Equilibrium Steady-State
NNBP	Nearest-Neigbour Base-Pairing
PBS	Polarizing Beam Splitter
PCR	Polymerase Chain Reaction
PSD	Position-Sensing Detector
SA	Streptavidin
ssDNA	single-stranded dsDNA
WLC	Worm-Like Chain

BIBLIOGRAPHY

[1] Thomas S Kuhn. *The structure of scientific revolutions*. University of Chicago press, 2012.

[2] Stanford University. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/thomas-kuhn/. Accessed: 2020-09-25.

[3] Roger Schrodinger, Erwin Schrödinger, and Erwin Schrödinger Dinger. *What is life?: With mind and matter and autobiographical sketches*. Cambridge University Press, 1992.

[4] Biophysical Society. *What is Biophysics?* https://www.biophysics.org/what-is-biophysics. Accessed: 2020-09-22.

[5] Josep Maria Huguet. "Statistical and thermodynamic properties of DNA unzipping experiments with optical tweezers." PhD thesis. Universitat de Barcelona, Oct. 2010.

[6] Daniel Wrapp, Nianshuang Wang, Kizzmekia S. Corbett, Jory A. Goldsmith, Ching-Lin Hsieh, Olubukola Abiona, Barney S. Graham, and Jason S. McLellan. "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation." In: *Science* 367.6483 (2020), pp. 1260–1263. ISSN: 0036-8075. DOI: 10.1126/science.abb2507. eprint: https://science.sciencemag.org/content/367/6483/1260.full.pdf. URL: https://science.sciencemag.org/content/367/6483/1260.

[7] Renhong Yan, Yuanyuan Zhang, Yaning Li, Lu Xia, Yingying Guo, and Qiang Zhou. "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2." In: *Science* 367.6485 (2020), pp. 1444–1448. ISSN: 0036-8075. DOI: 10.1126/science.abb2762. eprint: https://science.sciencemag.org/content/367/6485/1444.full.pdf. URL: https://science.sciencemag.org/content/367/6485/1444.

[8] D.S. Goodsell, S. Dutta, C. Zardecki, M. Voigt, H.M. Berman, and S.K. Burley. "The RCSB PDB Molecule of the Month: Inspiring a Molecular View of Biology." In: *Plos Biology* 13.5 (2015). DOI: 10.1371/journal.pbio.1002140.

[9] D. Bohan et al. "Networking Agroecology: Integrating the Diversity of Agroecosystem Interactions." In: *Advances in Ecological Research* 49 (2013), pp. 1–67.

[10] SB Smith, L Finzi, and C Bustamante. "Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads." In: *Science* 258.5085 (1992), pp. 1122–1126. ISSN: 0036-8075.

[11] C Bustamante, JF Marko, ED Siggia, and S Smith. "Entropic elasticity of lambda-phage DNA." In: *Science* 265.5178 (1994), pp. 1599–1600. ISSN: 0036-8075. DOI: 10.1126/science.8079175. eprint: https://science.sciencemag.org/content/265/5178/1599.full.pdf. URL: https://science.sciencemag.org/content/265/5178/1599.

[12] J. D. Watson and F. H. C. Crick. "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid." In: *Nature* 171.4356 (1953), pp. 737–738. ISSN: 1476-4687. DOI: 10.1038/171737a0. URL: https://doi.org/10.1038/171737a0.

[13] Álvaro Martínez Monge. "Free energy and information-content measurements in thermodynamic and molecular ensembles." PhD thesis. Universitat de Barcelona, Mar. 2019.

[14] F. Ritort. "Single-molecule experiments in biological physics: methods and applications." In: *J Phys Condens Matter* 18 (2006), R531–R583.

[15] J. Lipfert, X. Hao, and N. H. Dekker. "Quantitative modeling and optimization of magnetic tweezers." In: *Biophys. J.* 96.12 (2009), pp. 5040–5049.

[16] Marco Ribezzi-Crivellari, Josep M Huguet, and Felix Ritort. "Counter-propagating dual-trap optical tweezers based on linear momentum conservation." In: *Review of Scientific Instruments* 84.4 (2013), p. 043104.

[17] Brian Vohnsen. "A Short History of Optics." In: *Physica Scripta* 2004 (July 2006), p. 75. DOI: 10.1238/Physica.Topical.109a00075.

[18] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and Steven Chu. "Observation of a single-beam gradient force optical trap for dielectric particles." In: *Opt. Lett.* 11.5 (1986), pp. 288–290. DOI: 10.1364/OL.11.000288. URL: http://ol.osa.org/abstract.cfm?URI=ol-11-5-288.

[19] The Nobel Prize. *The Nobel Prize in Physics 2018: Arthur Ashkin - Facts.* https://www.nobelprize.org/prizes/physics/2018/ashkin/facts/. Accessed: 2020-08-27.

[20] Gustav Mie. "Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen." In: *Annalen der Physik* 330.3 (1908), pp. 377–445. DOI: 10.1002/andp.19083300302. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19083300302. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19083300302.

[21] J. P. Barton, D. R. Alexander, and S. A. Schaub. "Theoretical determination of net radiation force and torque for a spherical particle illuminated by a focused laser beam." In: *Journal of Applied Physics* 66.10 (1989), pp. 4594–4602. DOI: 10.1063/1.343813. eprint: https://doi.org/10.1063/1.343813. URL: https://doi.org/10.1063/1.343813.

[22] Steven B. Smith, Yujia Cui, and Carlos Bustamante. "Optical-trap force transducer that operates by direct measurement of light momentum." In: *Biophotonics, Part B*. Vol. 361. Methods in Enzymology. Academic Press, 2003, pp. 134 –162. DOI: https://doi.org/10.1016/S0076-6879(03)61009-8. URL: http://www.sciencedirect.com/science/article/pii/S0076687903610098.

[23] Università di Parma. *Assemblies: Pipette Puller*. http://tweezerslab.unipr.it/cgi-bin/mt/assemblies.pl/Show?_id=7c1b. Accessed: 2020-08-21.

[24] Steven B. Smith, Yujia Cui, and Carlos Bustamante. "Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules." In: *Science* 271.5250 (1996), pp. 795–799. ISSN: 0036-8075. DOI: 10.1126/science.271.5250.795. eprint: http://science.sciencemag.org/content/271/5250/795.full.pdf. URL: http://science.sciencemag.org/content/271/5250/795.

[25] Mark C. Williams, Jay R. Wenner, Ioulia Rouzina, and Victor A. Bloomfield. "Entropy and Heat Capacity of DNA Melting from Temperature Dependence of Single Molecule Stretching." In: *Biophysical Journal* 80.4 (2001), pp. 1932–1939. ISSN: 0006-3495. DOI: 10.1016/S0006-3495(01)76163-2. URL: https://doi.org/10.1016/S0006-3495(01)76163-2.

[26] Jay R. Wenner, Mark C. Williams, Ioulia Rouzina, and Victor A. Bloomfield. "Salt Dependence of the Elasticity and Overstretching Transition of Single DNA Molecules." In: *Biophysical Journal* 82.6 (2002), pp. 3160 –3169. ISSN: 0006-3495. DOI: https://doi.org/10.1016/S0006-3495(02)75658-0. URL: http://www.sciencedirect.com/science/article/pii/S0006349502756580.

[27] Anna Alemany. "Dynamic force spectroscopy and folding kinetics in molecular systems." PhD thesis. Universitat de Barcelona, Sept. 2014.

[28] James D Watson. *Molecular biology of the gene*. Pearson Education India, 2004.

[29] Howard M. Temin and Satoshi Mizutani. "Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus." In: *Nature* 226.5252 (1970), pp. 1211–1213. ISSN: 1476-4687. DOI: 10.1038/2261211a0. URL: https://doi.org/10.1038/2261211a0.

[30] S. Altman, M.F. Baer, M. Bartkiewicz, H. Gold, C. Guerrier-Takada, L.A. Kirsebom, N. Lumelsky, and K. Peck. "Catalysis by the RNA subunit of RNase P — a minireview." In: *Gene* 82.1 (1989), pp. 63 –64. ISSN: 0378-1119. DOI: https://doi.org/10.1016/0378-1119(89)90030-9. URL: http://www.sciencedirect.com/science/article/pii/0378111989900309.

[31] C. David Allis and Thomas Jenuwein. "The molecular hallmarks of epigenetic control." In: *Nature Reviews Genetics* 17.8 (2016), pp. 487–500. ISSN: 1471-0064. DOI: 10.1038/nrg.2016.59. URL: https://doi.org/10.1038/nrg.2016.59.

[32] Francis Crick. "Central Dogma of Molecular Biology." In: *Nature* 227.5258 (1970), pp. 561–563. ISSN: 1476-4687. DOI: 10.1038/227561a0. URL: https://doi.org/10.1038/227561a0.

[33] S Neidle. *Principles of Nucleic Acid Structure 2008*.

[34] R. R. Sinden. *DNA Structure and Function*. Elsevier Science, 1994.

[35] A. R. Ferré-D'Amar? K. Zhou, and J. A. Doudna. "Crystal structure of a hepatitis delta virus ribozyme." In: *Nature* 395.6702 (1998), pp. 567–574.

[36] L. A. Marky and K. J. Breslauer. "Calorimetric determination of base-stacking enthalpies in double-helical DNA molecules." In: *Biopolymers* 21.11 (1982), pp. 2185–2194.

[37] Peter Yakovchuk, Ekaterina Protozanova, and Maxim D. Frank-Kamenetskii. "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix." In: *Nucleic Acids Research* 34.2 (Jan. 2006), pp. 564–574. ISSN: 0305-1048. DOI: 10.1093/nar/gkj454. eprint: https://academic.oup.com/nar/article-pdf/34/2/564/25956946/gkj454.pdf. URL: https://doi.org/10.1093/nar/gkj454.

[38] Christoph Arenz and Oliver Seitz. "DNA Made of Purines Only." In: *Chemistry & Biology* 14.5 (2007), pp. 467 –469. ISSN: 1074-5521. DOI: https://doi.org/10.1016/j.chembiol.2007.05.001. URL: http://www.sciencedirect.com/science/article/pii/S1074552107001470.

[39] Navin Singh and Yashwant Singh. "Effect of defects on thermal denaturation of DNA oligomers." In: *Physical Review E* 64.4 (2001), p. 042901.

[40] M. J. McCauley, L. Furman, C. A. Dietrich, I. Rouzina, M. E. Nunez, and M. C. Williams. "Quantifying the stability of oxidatively damaged DNA by single-molecule DNA stretching." In: *Nucl Acids Res* 46.8 (2018), pp. 4033–4043.

[41] F. Landuzzi, X. Viader-Godoy, F. Cleri, I. Pastor, and F. Ritort. "Detection of single DNA mismatches by force spectroscopy in short DNA hairpins." In: *The Journal of Chemical Physics* 152.7 (2020), p. 074204. DOI: 10.1063/1.5139284.

[42] Walter Gilbert. "Origin of life: The RNA world." In: *nature* 319.6055 (1986), pp. 618–618.

[43] B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. "Global quantification of mammalian gene expression control." In: *Nature* 473.7347 (2011), pp. 337–342.

[44] Douglas A. Bernstein, Morgan C. Zittel, and James L. Keck. "High-resolution structure of the E.coli RecQ helicase catalytic core." In: *The EMBO Journal* 22.19 (2003), pp. 4910–4921. DOI: 10.1093/emboj/cdg500. eprint: https://www.embopress.org/doi/pdf/10.1093/emboj/cdg500. URL: https://www.embopress.org/doi/abs/10.1093/emboj/cdg500.

[45] International Union of Pure and Applied Chemistry. *IUPAC Compendium of Chemical Terminology – The Gold Book*. 2009. URL: http://goldbook.iupac.org/.

[46] H. Staudinger. "Über Polymerisation." In: *Berichte der deutschen chemischen Gesellschaft (A and B Series)* 53.6 (1920), pp. 1073–1085. DOI: 10.1002/cber.19200530627. eprint: https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/cber.19200530627. URL: https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cber.19200530627.

[47] The Nobel Prize. *The Nobel Prize in Chemistry 1974*. https://www.nobelprize.org/prizes/chemistry/1974/summary/. Accessed: 2020-09-20.

[48] Michael Rubinstein, Ralph H Colby, et al. *Polymer physics*. Vol. 23. Oxford university press New York, 2003.

[49] Masao Doi. *Introduction to polymer physics*. Oxford university press, 1996.

[50] John F Marko and Eric D Siggia. "Stretching DNA." In: *Macromolecules* 28.26 (1995), pp. 8759–8770.

[51] C Bouchiat, MD Wang, J-F Allemand, T Strick, SM Block, and V Croquette. "Estimating the persistence length of a worm-like chain molecule from force-extension measurements." In: *Biophysical journal* 76.1 (1999), pp. 409–413.

[52] AM Monge, M Manosas, and F Ritort. "Experimental test of ensemble inequivalence and the fluctuation theorem in the force ensemble in DNA pulling experiments." In: *Physical Review E* 98.3 (2018), p. 032146.

[53] O. Kratky and G. Porod. "Röntgenuntersuchung gelöster Fadenmoleküle." In: *Recueil des Travaux Chimiques des Pays-Bas* 68.12 (1949), pp. 1106–1122. DOI: 10.1002/recl.19490681203. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/recl.19490681203. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/recl.19490681203.

[54] Philip Nelson. *Biological physics*. New York: WH Freeman, 2004.

[55] J. David Moroz and Philip Nelson. "Torsional directed walks, entropic elasticity, and DNA twist stiffness." In: *Proceedings of the National Academy of Sciences* 94.26 (1997), pp. 14418–14422. ISSN: 0027-8424. DOI: 10.1073/pnas.94.26.14418. eprint: https://www.pnas.org/content/94/26/14418.full.pdf. URL: https://www.pnas.org/content/94/26/14418.

[56] Jan Lipfert, Mina Lee, Orkide Ordu, Jacob W. J. Kerssemakers, and Nynke H. Dekker. "Magnetic tweezers for the measurement of twist and torque." eng. In: *Journal of visualized experiments : JoVE* 87 (2014), p. 51503. ISSN: 1940-087X. DOI: 10.3791/51503. URL: https://pubmed.ncbi.nlm.nih.gov/24894412.

[57] Theo Odijk. "Stiff Chains and Filaments under Tension." In: *Macromolecules* 28.20 (1995), pp. 7016–7018. ISSN: 0024-9297. DOI: 10.1021/ma00124a044. URL: https://doi.org/10.1021/ma00124a044.

[58] Rafayel Petrosyan. "Improved approximations for some polymer extension models." In: *Rheologica Acta* 56.1 (2017), pp. 21–26. ISSN: 1435-1528. DOI: 10.1007/s00397-016-0977-9. URL: https://doi.org/10.1007/s00397-016-0977-9.

[59] Michelle D Wang, Hong Yin, Robert Landick, Jeff Gelles, and Steven M Block. "Stretching DNA with optical tweezers." In: *Biophysical journal* 72.3 (1997), p. 1335.

[60] Gert R.. Strobl. *The Physics of Polymers: Concepts for Understanding Their Structures and Behavior; with 295 Figures and 2 Tables.* Springer, 2007.

[61] Alessandro Bosco, Joan Camunas-Soler, and Felix Ritort. "Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions." In: *Nucleic acids research* 42.3 (2014), pp. 2064–2074.

[62] Anna Alemany, Blanca Rey-Serra, Silvia Frutos, Ciro Cecconi, and Felix Ritort. "Mechanical folding and unfolding of protein barnase at the single-molecule level." In: *Biophysical Journal* 110.1 (2016), pp. 63–74.

[63] Anna Alemany and Felix Ritort. "Determination of the elastic properties of short ssDNA molecules by mechanically folding and unfolding DNA hairpins." In: *Biopolymers* 101.12 (2014), pp. 1193–1199.

[64] Jozef Adamcik, Dmitry V Klinov, Guillaume Witz, Sergey K Sekatskii, and Giovanni Dietler. "Observation of single-stranded DNA on mica and highly oriented pyrolytic graphite by atomic force microscopy." In: *FEBS letters* 580.24 (2006), pp. 5671–5675.

[65] Joan Camunas-Soler, Marco Ribezzi-Crivellari, and Felix Ritort. "Elastic Properties of Nucleic Acids by Single-Molecule Force Spectroscopy." In: *Annual review of biophysics* 45.1 (2016), pp. 65–84.

[66] H. Clausen-Schaumann, M. Rief, C. Tolksdorf, and H. E. Gaub. "Mechanical stability of single DNA molecules." In: *Biophys J* 78.4 (2000), pp. 1997–2007. ISSN: 0006-3495. DOI: 10.1016/S0006-3495(00)76747-6. URL: https://www.ncbi.nlm.nih.gov/pubmed/10733978.

[67] Sören Doose, Hannes Barsch, and Markus Sauer. "Polymer properties of polythymine as revealed by translational diffusion." In: *Biophys J* 93.4 (2007), pp. 1224–1234. ISSN: 0006-3495. DOI: 10.1529/biophysj.107.107342. URL: https://www.ncbi.nlm.nih.gov/pubmed/17513377.

[68] Suresh Manohar, Amber R. Mantz, Kevin E. Bancroft, Chung-Yuen Hui, Anand Jagota, and Dmitri V. Vezenov. "Peeling single-stranded DNA from graphite surface to determine oligonucleotide binding energy by force spectroscopy." In: *Nano Lett* 8.12 (2008), pp. 4365–4372. ISSN: 1530-6984. URL: https://www.ncbi.nlm.nih.gov/pubmed/19368004.

[69] MC Murphy, Ivan Rasnik, Wei Cheng, Timothy M Lohman, and Taekjip Ha. "Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy." In: *Biophysical journal* 86.4 (2004), pp. 2530–2537.

[70] A. Buhot and A. Halperin. "Effects of stacking on the configurations and elasticity of single-stranded nucleic acids." In: *Phys. Rev. E* 70 (2 2004), p. 020902. DOI: 10.1103/PhysRevE.70.020902. URL: https://link.aps.org/doi/10.1103/PhysRevE.70.020902.

[71] Changhong Ke, Michael Humeniuk, Hanna S-Gracz, and Piotr E. Marszalek. "Direct Measurements of Base Stacking Interactions in DNA by Single-Molecule Atomic-Force Spectroscopy." In: *Phys. Rev. Lett.* 99 (1 2007), p. 018302. DOI: 10.1103/PhysRevLett.99.018302. URL: https://link.aps.org/doi/10.1103/PhysRevLett.99.018302.

[72] DustinB. McIntosh, Gina Duggan, Quentin Gouil, and OmarA. Saleh. "Sequence-Dependent Elasticity and Electrostatics of Single-Stranded DNA: Signatures of Base-Stacking." In: *Biophysical Journal* 106.3 (2014), pp. 659 –666. ISSN: 0006-3495. DOI: https://doi.org/10.1016/j.bpj.2013.12.018. URL: http://www.sciencedirect.com/science/article/pii/S0006349513058104.

[73] Michael T. Woodside, William M. Behnke-Parks, Kevan Larizadeh, Kevin Travers, Daniel Herschlag, and Steven M. Block. "Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins." In: *Proceedings of the National Academy of Sciences* 103.16 (2006), pp. 6190–6195. DOI: 10.1073/pnas.0511048103. eprint: http://www.pnas.org/content/103/16/6190.full.pdf. URL: http://www.pnas.org/content/103/16/6190.

[74] Huimin Chen, Steve P. Meisburger, Suzette A. Pabit, Julie L. Sutton, Watt W. Webb, and Lois Pollack. "Ionic strength-dependent persistence lengths of single-stranded RNA and DNA." In: *Proceedings of the National Academy of Sciences* 109.3 (2012), pp. 799–804. ISSN: 0027-8424. DOI: 10.1073/pnas.1119057109. eprint: http://www.pnas.org/content/109/3/799.full.pdf. URL: http://www.pnas.org/content/109/3/799.

[75] M. C. Murphy, Ivan Rasnik, Wei Cheng, Timothy M. Lohman, and Taekjip Ha. "Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy." In: *Biophys J* 86.4 (2004), pp. 2530–2537. ISSN: 0006-3495. DOI: 10.1016/S0006-3495(04)74308-8. URL: https://www.ncbi.nlm.nih.gov/pubmed/15041689.

[76] M.-N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette. "Stretching Single Stranded DNA, a Model Polyelectrolyte." In: *Phys. Rev. Lett.* 89 (24 2002), p. 248102. DOI: 10.1103/PhysRevLett.89.248102. URL: https://link.aps.org/doi/10.1103/PhysRevLett.89.248102.

[77] D. B. McIntosh and O. A. Saleh. "Salt Species-Dependent Electrostatic Effects on ssDNA Elasticity." In: *Macromolecules* 44.7 (2011), pp. 2328–2333. ISSN: 0024-9297. DOI: 10.1021/ma1028196. URL: https://doi.org/10.1021/ma1028196.

[78] Sara de Lorenzo, Marco Ribezzi-Crivellari, J. Ricardo Arias-Gonzalez, Steven B. Smith, and Felix Ritort. "A Temperature-Jump Optical Trap for Single-Molecule Manipulation." In: *Biophys J* 108.12 (2015), pp. 2854–2864. ISSN: 1542-0086. DOI: 10.1016/j.bpj.2015.05.017. URL: https://www.ncbi.nlm.nih.gov/pubmed/26083925.

[79] O. A. Saleh, D. B. McIntosh, P. Pincus, and N. Ribeck. "Nonlinear Low-Force Elasticity of Single-Stranded DNA Molecules." In: *Phys. Rev. Lett.* 102 (6 2009), p. 068301. DOI: 10.1103/PhysRevLett.102.068301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.102.068301.

[80] Yeonee Seol, Jinyu Li, Philip C Nelson, Thomas T Perkins, and MD Betterton. "Elasticity of short DNA molecules: theory and experiment for contour lengths of $0.6–7\mu$m." In: *Biophysical journal* 93.12 (2007), pp. 4360–4373.

[81] 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." In: *Biophys. J.* 100 (2011), pp. 1765–1774.

[82] Josep M. Huguet, Cristiano V. Bizarro, Núria Forns, Steven B. Smith, Carlos Bustamante, and Felix Ritort. "Single-molecule derivation of salt dependent base-pair free energies in DNA." In: *Proceedings of the National Academy of Sciences* 107.35 (2010), pp. 15431–15436. ISSN: 0027-8424. DOI: 10.1073/pnas.1001454107. eprint: http://www.pnas.org/content/107/35/15431.full.pdf. URL: http://www.pnas.org/content/107/35/15431.

[83] Josep Maria Huguet, Marco Ribezzi-Crivellari, Cristiano Valim Bizarro, and Felix Ritort. "Derivation of nearest-neighbor DNA parameters in magnesium from single molecule experiments." In: *Nucleic Acids Research* 45.22 (2017), pp. 12921–12931. eprint:

/oup/backfile/content_public/journal/nar/45/22/10.
1093_nar_gkx1161/2/gkx1161.pdf.

[84] Berenike Maier, David Bensimon, and Vincent Croquette. "Replication by a single DNA polymerase of a stretched single-stranded DNA." In: *Proceedings of the National Academy of Sciences* 97.22 (2000), pp. 12002–12007. ISSN: 0027-8424. DOI: 10.1073/pnas.97.22.12002.

[85] Michael T. Woodside, William M. Behnke-Parks, Kevan Larizadeh, Kevin Travers, Daniel Herschlag, and Steven M. Block. "Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins." In: *Proceedings of the National Academy of Sciences* 103.16 (2006), pp. 6190–6195.

[86] Noel L. Goddard, Grégoire Bonnet, Oleg Krichevsky, and Albert Libchaber. "Sequence Dependent Rigidity of Single Stranded DNA." In: *Phys. Rev. Lett.* 85 (11 2000), pp. 2400–2403. DOI: 10.1103/PhysRevLett.85.2400. URL: https://link.aps.org/doi/10.1103/PhysRevLett.85.2400.

[87] M. Egli. "Sugar Pucker and Nucleic Acid Structure." In: (2018).

[88] Karunesh Arora and Tamar Schlick. "Deoxyadenosine sugar puckering pathway simulated by the stochastic difference equation algorithm." In: *Chemical Physics Letters* 378.1 (2003), pp. 1 – 8. ISSN: 0009-2614. DOI: https://doi.org/10.1016/S0009-2614(03)01195-3. URL: http://www.sciencedirect.com/science/article/pii/S0009261403011953.

[89] Junmei Wang, Piotr Cieplak, and Peter A. Kollman. "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?" In: *Journal of Computational Chemistry* 21.12 (2000), pp. 1049–1074. DOI: 10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1096-987X%28200009%2921%3A12%3C1049%3A%3AAID-JCC3%3E3.0.CO%3B2-F. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/1096-987X%28200009%2921%3A12%3C1049%3A%3AAID-JCC3%3E3.0.CO%3B2-F.

[90] Huimin Chen, Steve P. Meisburger, Suzette A. Pabit, Julie L. Sutton, Watt W. Webb, and Lois Pollack. "Ionic strength-dependent persistence lengths of single-stranded RNA and DNA." In: *Proceedings of the National Academy of Sciences* 109.3 (2012), pp. 799–804. ISSN: 0027-8424. eprint: http://www.pnas.org/content/109/3/799.full.pdf.

[91] John SantaLucia. "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics." In: *Proceedings of the National Academy of Sciences* 95.4 (1998), pp. 1460–1465. ISSN: 0027-8424. DOI: 10.1073/pnas.95.4.1460. eprint: http://www.pnas.org/content/95/4/1460.full.pdf. URL: http://www.pnas.org/content/95/4/1460.

[92] P. Vaitiekunas, C. Crane-Robinson, and P. L. Privalov. "The energetic basis of the DNA double helix: a combined microcalorimetric approach." In: *Nucleic Acids Res.* 43.17 (2015), pp. 8577–8589.

[93] Fabian Kilchherr, Christian Wachauf, Benjamin Pelz, Matthias Rief, Martin Zacharias, and Hendrik Dietz. "Single-molecule dissection of stacking forces in DNA." In: *Science* 353.6304 (2016). ISSN: 0036-8075. DOI: `10.1126/science.aaf5508`. eprint: `https://science.sciencemag.org/content/353/6304/aaf5508.full.pdf`. URL: `https://science.sciencemag.org/content/353/6304/aaf5508`.

[94] Jayanthi Ramprakash, Brian Lang, and Frederick P. Schwarz. "Thermodynamics of single strand DNA base stacking." In: *Biopolymers* 89.11 (), pp. 969–979. DOI: `10.1002/bip.21044`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/bip.21044`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.21044`.

[95] Alex Plumridge, Steve P. Meisburger, Kurt Andresen, and Lois Pollack. "The impact of base stacking on the conformations and electrostatics of single-stranded DNA." In: *Nucleic Acids Res* 45.7 (2017), pp. 3932–3943. ISSN: 1362-4962. DOI: `10.1093/nar/gkx140`. URL: `https://www.ncbi.nlm.nih.gov/pubmed/28334825`.

[96] Daniel P. Aalberts, John M. Parman, and Noel L. Goddard. "Single-Strand Stacking Free Energy from DNA Beacon Kinetics." In: *Biophysical Journal* 84.5 (2003), pp. 3212 –3217. ISSN: 0006-3495. DOI: `https://doi.org/10.1016/S0006-3495(03)70045-9`. URL: `http://www.sciencedirect.com/science/article/pii/S0006349503700459`.

[97] Yeonee Seol, Gary M. Skinner, Koen Visscher, Arnaud Buhot, and Avraham Halperin. "Stretching of Homopolymeric RNA Reveals Single-Stranded Helices and Base-Stacking." In: *Phys. Rev. Lett.* 98 (15 2007), p. 158103. DOI: `10.1103/PhysRevLett.98.158103`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.98.158103`.

[98] Susan L Heilman-Miller, D Thirumalai, and Sarah A Woodson. "Role of counterion condensation in folding of the Tetrahymena ribozyme. I. Equilibrium stabilization by cations1 1Edited by J. Doudna." In: *Journal of Molecular Biology* 306.5 (2001), pp. 1157 –1166. ISSN: 0022-2836. DOI: `https://doi.org/10.1006/jmbi.2001.4437`. URL: `http://www.sciencedirect.com/science/article/pii/S0022283601944377`.

[99] Susan L. Heilman-Miller, Jie Pan, D. Thirumalai, and Sarah A. Woodson. "Role of counterion condensation in folding of the Tetrahymena ribozyme II. Counterion-dependence of folding kinetics11Edited by J. Doudna." In: *Journal of Molecular Biology* 309.1 (2001), pp. 57 –68. ISSN: 0022-2836. DOI: `https://doi.org/`

10.1006/jmbi.2001.4660. URL: http://www.sciencedirect.com/science/article/pii/S0022283601946601.

[100] Cristiano V Bizarro, Anna Alemany, and Felix Ritort. "Non-specific binding of Na+ and Mg2+ to RNA determined by force spectroscopy methods." In: *Nucleic acids research* 40.14 (2012), pp. 6922–6935.

[101] Peter Yakovchuk, Ekaterina Protozanova, and Maxim D. Frank-Kamenetskii. "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix." In: *Nucleic Acids Research* 34.2 (Jan. 2006), pp. 564–574. ISSN: 0305-1048. DOI: 10.1093/nar/gkj454. eprint: https://academic.oup.com/nar/article-pdf/34/2/564/25956946/gkj454.pdf. URL: https://doi.org/10.1093/nar/gkj454.

[102] Michael Zuker. "Mfold web server for nucleic acid folding and hybridization prediction." eng. In: *Nucleic acids research* 31.13 (2003), pp. 3406–3415. ISSN: 1362-4962. DOI: 10.1093/nar/gkg595. URL: https://pubmed.ncbi.nlm.nih.gov/12824337.

[103] David R. Jacobson, Dustin B. McIntosh, Mark J. Stevens, Michael Rubinstein, and Omar A. Saleh. "Single-stranded nucleic acid elasticity arises from internal electrostatic tension." In: *Proceedings of the National Academy of Sciences* 114.20 (2017), pp. 5095–5100. ISSN: 0027-8424. DOI: 10.1073/pnas.1701132114. eprint: http://www.pnas.org/content/114/20/5095.full.pdf. URL: http://www.pnas.org/content/114/20/5095.

[104] Andrea Montanari and Marc Mézard. "Hairpin Formation and Elongation of Biomolecules." In: *Phys. Rev. Lett.* 86 (10 2001), pp. 2178–2181. DOI: 10.1103/PhysRevLett.86.2178. URL: https://link.aps.org/doi/10.1103/PhysRevLett.86.2178.

[105] Davide Marenduzzo and Cristian Micheletti. "Thermodynamics of DNA Packaging Inside a Viral Capsid: The Role of DNA Intrinsic Thickness." In: *Journal of Molecular Biology* 330.3 (2003), pp. 485–492. ISSN: 0022-2836. DOI: https://doi.org/10.1016/S0022-2836(03)00584-9. URL: http://www.sciencedirect.com/science/article/pii/S0022283603005849.

[106] Ngo Minh Toan, Davide Marenduzzo, and Cristian Micheletti. "Inferring the Diameter of a Biopolymer from Its Stretching Response." In: *Biophysical Journal* 89.1 (2005), pp. 80 –86. ISSN: 0006-3495. DOI: https://doi.org/10.1529/biophysj.104.058081. URL: http://www.sciencedirect.com/science/article/pii/S0006349505726606.

[107] Ngo Toan and Cristian Micheletti. "Inferring the effective thickness of polyelectrolytes from stretching measurements at various ionic strengths: Applications to DNA and RNA." In: *J. Phys. Condens. Matter* 18 (Jan. 2006). DOI: 10.1088/0953-8984/18/14/S11.

[108] John SantaLucia. "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics." In: *Proceedings of the National Academy of Sciences* 95.4 (1998), pp. 1460–1465. ISSN: 0027-8424. DOI: 10.1073/pnas.95.4.1460. eprint: http://www.pnas.org/content/95/4/1460.full.pdf. URL: http://www.pnas.org/content/95/4/1460.

[109] Michael T. Woodside, Peter C. Anthony, William M. Behnke-Parks, Kevan Larizadeh, Daniel Herschlag, and Steven M. Block. "Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid." In: *Science* 314.5801 (2006), pp. 1001–1004. ISSN: 0036-8075. DOI: 10.1126/science.1133601. eprint: http://science.sciencemag.org/content/314/5801/1001.full.pdf. URL: http://science.sciencemag.org/content/314/5801/1001.

[110] Nicolas Peyret. *Prediction of Nucleic Acid Hybridization: Parameters and Algorithms PhD dissertation*. Wayne State University, Department of Chemistry, Detroit(MI), 2000.

[111] GCG Help. *Mfold*. http://bioinfo.nhri.org.tw/gcg/doc/11.0/mfold.html. Accessed: 2020-09-23.

[112] J. M. Caruthers and D. B. McKay. "Helicase structure and mechanism." In: *Curr. Opin. Struct. Biol.* 12.1 (2002), pp. 123–133.

[113] S. S. Patel and I. Donmez. "Mechanisms of helicases." In: *J. Biol. Chem.* 281.27 (2006), pp. 18265–18268.

[114] Daniel Klaue, Daniela Kobbe, Felix Kemmerich, Alicja Kozikowska, Holger Puchta, and Ralf Seidel. "Fork sensing and strand switching control antagonistic activities of RecQ helicases." In: *Nature Communications* 4.1 (2013), p. 2024. ISSN: 2041-1723. DOI: 10.1038/ncomms3024. URL: https://doi.org/10.1038/ncomms3024.

[115] Vegard Boerve Sordal. "Entropy production and information processing in stochastic thermodynamics: Optimization, measurement, and erasure." PhD thesis. University of Oslo, Sept. 2019.

[116] Debjani Bagchi, Maria Manosas, Weiting Zhang, Kelly A Manthei, Samar Hodeib, Bertrand Ducos, James L Keck, and Vincent Croquette. "Single molecule kinetics uncover roles for E. coli RecQ DNA helicase domains and interaction with SSB." In: *Nucleic Acids Research* 46.16 (July 2018), pp. 8500–8515. ISSN: 0305-1048. DOI: 10.1093/nar/gky647. eprint: https://academic.oup.com/nar/article-pdf/46/16/8500/25802553/gky647.pdf. URL: https://doi.org/10.1093/nar/gky647.

[117] C. J. Fischer and T. M. Lohman. "ATP-dependent translocation of proteins along single-stranded DNA: models and methods of analysis of pre-steady state kinetics." In: *J. Mol. Biol.* 344.5 (2004), pp. 1265–1286.

[118] Gábor M. Harami, Nikolett T. Nagy, Máté Martina, Keir C. Neuman, and Mihály Kovács. "The HRDC domain of E. coli RecQ helicase controls single-stranded DNA translocation and double-stranded DNA unwinding rates without affecting mechanoenzymatic coupling." In: *Scientific Reports* 5.1 (2015), p. 11091. ISSN: 2045-2322. DOI: 10.1038/srep11091. URL: https://doi.org/10.1038/srep11091.

[119] J. Rosing and E. C. Slater. "The value of G degrees for the hydrolysis of ATP." In: *Biochim. Biophys. Acta* 267.2 (1972), pp. 275–290.

[120] Behzad Rad and Stephen C. Kowalczykowski. "Efficient coupling of ATP hydrolysis to translocation by RecQ helicase." In: *Proceedings of the National Academy of Sciences* 109.5 (2012), pp. 1443–1448. ISSN: 0027-8424. DOI: 10.1073/pnas.1119952109. eprint: https://www.pnas.org/content/109/5/1443.full.pdf. URL: https://www.pnas.org/content/109/5/1443.

[121] Kata Sarlós, Máté Gyimesi, and Mihály Kovács. "RecQ helicase translocates along single-stranded DNA with a moderate processivity and tight mechanochemical coupling." In: *Proceedings of the National Academy of Sciences* 109.25 (2012), pp. 9804–9809. ISSN: 0027-8424. DOI: 10.1073/pnas.1114468109. eprint: https://www.pnas.org/content/109/25/9804.full.pdf. URL: https://www.pnas.org/content/109/25/9804.

[122] Timothy M Lohman, Eric J Tomko, and Colin G Wu. "Non-hexameric DNA helicases and translocases: mechanisms and regulation." In: *Nature Reviews Molecular Cell Biology* 9.5 (2008), pp. 391–401.

[123] Gavin E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences." In: *Phys. Rev. E* 60 (3 1999), pp. 2721–2726. DOI: 10.1103/PhysRevE.60.2721. URL: https://link.aps.org/doi/10.1103/PhysRevE.60.2721.

[124] Felix Ritort. "Nonequilibrium fluctuations in small systems: From physics to biology." In: *Adv. Chem. Phys.* 137 (June 2007).

[125] R. Ait-Haddou and W. Herzog. "Brownian ratchet models of molecular motors." In: *Cell Biochem. Biophys.* 38.2 (2003), pp. 191–214.

[126] David Cubero and Ferruccio Renzoni. *Brownian Ratchets: From Statistical Physics to Bio and Nano-motors*. Cambridge University Press, 2016.

[127] M. D. Betterton and F. Jülicher. "Opening of nucleic-acid double strands by helicases: active versus passive opening." In: *Phys Rev E Stat Nonlin Soft Matter Phys* 71.1 Pt 1 (2005), p. 011904.

[128] Ashok Garai, Debashish Chowdhury, and M Betterton. "Two-state model for helicase translocation and unwinding of nucleic acids." In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 77 (June 2008), p. 061910. DOI: 10.1103/PhysRevE.77.061910.

[129] Maria Manosas, Xu Guang Xi, David Bensimon, and Vincent Croquette. "Active and passive mechanisms of helicases." In: *Nucleic Acids Research* 38.16 (Apr. 2010), pp. 5518–5526. ISSN: 0305-1048. DOI: 10.1093/nar/gkq273. eprint: https://academic.oup.com/nar/article-pdf/38/16/5518/16767403/gkq273.pdf. URL: https://doi.org/10.1093/nar/gkq273.

[130] P. Modrich. "DNA mismatch correction." In: *Annu Rev Biochem* 56 (1987), pp. 435–466.

[131] A. Tikhomirova, I. V. Beletskaya, and T. V. Chalikian. "Stability of DNA duplexes containing GG, CC, AA, and TT mismatches." In: *Biochemistry* 45 (2006), pp. 10563–10571.

[132] N. Peyret, P. A. Seneviratne, H.T. Allawi, and J. SantaLucia. "Nearest-neighbor thermodynamics and NMR of DNA sequences with internal AA, CC, G, and TT mismatches." In: *Biochemistry* 38 (1999), pp. 3468–3477.

[133] G. Rossetti, P. D. Dans, I. Gomez-Pinto, I. Ivani, C. Gonzalez, and M. Orozco. "The structural impact of DNA mismatches." In: *Nucl Acids Res* 43 (2015), pp. 4309–4321.

[134] M. F. Goodman, S. Creighton, L. B. Bloom, and J. Petruska. "Biochemical basis of DNA replication fidelity." In: *J Crit Rev Biochem Mol Biol* 28 (1993), pp. 83–126.

[135] J. Wildenberg and M. Meselson. "Mismatch repair in heteroduplex DNA." In: *Proc Natl Acad Sci USA* 72 (1975), pp. 2202–2206.

[136] C. Kunz, Y. Saito, and P. Schär. "DNA Repair in mammalian cells. Mismatched repair: variations on a theme." In: *Cell Mol Life Sci* 66 (2009), pp. 1021–1038.

[137] P. Modrich. "Mechanisms in eukaryotic mismatch repair." In: *J Biol Chem* 281 (2006), pp. 30305–30309.

[138] R. D. Kolodner. "Mismatch repair: mechanisms and relationship to cancer susceptibility." In: *Trends Biochem Sci* 20 (1995), pp. 397–401.

[139] Sheila S. David, Valerie L. O'Shea, and Sucharita Kundu. "Base-excision repair of oxidative DNA damage." In: *Nature* 447.7147 (2007), pp. 941–950. ISSN: 1476-4687. DOI: 10.1038/nature05978. URL: https://doi.org/10.1038/nature05978.

[140] A. Meyerhans and J.-P. Vartanian. "The fidelity of cellular and viral polymerases and its manipulation." In: *Origin and evolution of viruses*. Ed. by E. Domingo, R. G. Webster, and J. F. Holland. New York: Academic Press, 1999. Chap. 5.

[141] P. M. Lizardi, X. Huang, Z. Zhu, P. Bray-Ward, D. C. Thomas, and D. C. Ward. "Mutation detection and single-molecule counting using isothermal rolling-circle amplification." In: *Nature Genet* 19 (1998), pp. 225–232.

[142] P. Hardenbol, J. Banér, M. Jain, M. Nilsson, E. A. Namsaraev, G. A. Karlin-Neumann, H. Fakhrai-Rad, M. Ronaghi, T. D. Willis, and Landegren U.et al. "Multiplexed genotyping with sequence-tagged molecular inversion probes." In: *Nature Biotech* 21 (2003), pp. 673–678.

[143] Q. Xu, S. Q. Huang, F. Ma, B. Tang, and C. Y. Zhang. "Controllable mismatched ligation for bioluminescence screening of known and unknown mutations." In: *Anal Chem* 88 (2016), pp. 2431–2439.

[144] F. M De La Vega, K. D. Lazaruk, M. D. Rhodes, and M. H. Wenz. "Assessment of two flexible and compatible NP genotyping platforms: TaqMan SNP genotyping assays and the SNPlex genotyping system." In: *Mut Res Fundam Mol Mech Mutagen* 573 (2005), pp. 111–135.

[145] S. Tyagi and F. R. Kramer. "Molecular beacons: probes that fluoresce upon hybridization." In: *Nature Biotech* 14 (1996), pp. 303–308.

[146] R. Wagner, P. Debble, and M. Radman. "Mutation detection using immobilized mismatch binding protein (MutS)." In: *Nucl Acids Res* 23 (1995), pp. 3944–3948.

[147] B. J. Till, C. Burtner, L. Comai, and S. Henikoff. "Mismatch cleavage by single-strand specific nucleases." In: *Nucl Acids Res* 32 (2004), pp. 2632–2641.

[148] D. J. Patel, S. A. Kozlowski, L. A. Marky, J. A. Rice, C. Broka, J. Dallas, K. Itakura, and K. Breslauer. "Structure, dynamics, and energetics of deoxyguanosine·thymidine wobble base pair formation in the self-complementary d(CGTGAATTCGCG) duplex in solution." In: *Biochemistry* 21.3 (1982), pp. 437–444.

[149] D. J. Patel, A. Pardi, and K. Itakura. "DNA conformation, dynamics, and interactions in solution." In: *Science* 216.4546 (1982), pp. 581–590.

[150] D. Rabinovich, T. Haran, M. Eisenstein, and Z. Shakked. "Structures of the mismatched duplex d(GGGTGCCC) and one of its Watson-Crick analogues d(GGGCGCCC)." In: *J Mol Biol* 200.1 (1988), pp. 151–161.

[151] T. Brown, O. Kennard, G. Kneale, and D. Rabinovich. "High-resolution structure of a DNA helix containing mismatched base pairs." In: *Nature* 315.6020 (1985), pp. 604–606.

[152] J. I. Friedman and J. T. Stivers. "Detection of damaged DNA bases by DNA glycosylase enzymes." In: *Biochemistry* 49.24 (2010), pp. 4957–4967.

[153] Y. Qi, M. C. Spong, K. Nam, A. Banerjee, S. Jiralerspong, M. Karplus, and G. L. Verdine. "Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme." In: *Nature* 462.7274 (2009), pp. 762–766.

[154] Y. Suzuki and O. K. Dudko. "Single-molecule rupture dynamics on multidimensional landscapes." In: *Phys Rev Lett* 104 (2010), p. 048101.

[155] A. Alemany and F. Ritort. "Force-dependent folding and unfolding kinetics in DNA hairpins reveals transition-state displacements along a single pathway." In: *J Phys Chem Lett* 8 (2017), pp. 895–900.

[156] Bernie D Sattin, Andrew E Pelling, and M Cynthia Goh. "DNA base pair resolution by single molecule force spectroscopy." In: *Nucl Acids Res* 32.16 (2004), pp. 4876–4883.

[157] Wenjing Liu, Yourong Guo, Kaizhe Wang, Xingfei Zhou, Ying Wang, Junhong Lü, Zhifeng Shao, Jun Hu, Daniel M Czajkowsky, and Bin Li. "Atomic force microscopy-based single-molecule force spectroscopy detects DNA base mismatches." In: *Nanoscale* 11.37 (2019), pp. 17206–17210.

[158] M. Manosas, D. Collin, and F. Ritort. "Force-dependent fragility in RNA hairpins." In: *Phys Rev Lett* 96 (2006), p. 218301.

[159] Evan Evans and Ken Ritchie. "Dynamic strength of molecular adhesion bonds." In: *Biophys. J.* 72 (1997), pp. 1541–1555.

[160] Evan Evans. "Probing the relation between force-lifetime-and chemistry in single molecular bonds." In: *Annu. Rev. Biophys. Biomol. Struct.* 30 (2001), pp. 105–128.

[161] A. Mossa, J. M. Huguet, and F. Ritort. "Investigating the thermodynamics of small biosystems with optical tweezers." In: *Physica E* 42 (2010), pp. 666–671.

[162] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. "GROMACS: A message-passing parallel molecular dynamics implementation." In: *Comp. Phys. Comm* 91.1–3 (1995), pp. 43–56.

[163] Erik Lindahl, Berk Hess, and David van der Spoel. "GROMACS 3.0: a package for molecular simulation and trajectory analysis." In: *Molecular modeling annual* 7.8 (2001), pp. 306–317.

[164] B. E. K. Snodin, Ferdinando Randisi, Majid Mosayebi, and et al. "Introducing improved structural properties and salt dependence into a coarse-grained model of DNA." In: *J. Chem. Phys.* 142 (2015).

[165] Changbong Hyeon and D. Thirumalai. "Capturing the essence of folding and functions of biomolecules using coarse-grained models." In: *Nature Communications* 2.487 (2011).

[166] O. Henrich, Y.A. Gutiérrez Fosado, Curk, and et al. "Coarse-Grained Simulation of DNA using LAMMPS an implementation of the oxNA model and its application." In: *Eur. Phys. J. E* (2018).

[167] Thomas E. Ouldridge, A. A. Louis, and Jonathan P. K. Doye. "Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model." In: *J. Chem. Phys.* 134 (2011).

[168] George I. Bell. "Models for the specific adhesion of cells to cells." In: *Science* 200 (1978), pp. 618–627.

[169] A. Garai, S. Mogurampelly, S. Bag, and P. K. Maiti. "Overstretching of B-DNA with various pulling protocols: Appearance of structural polymorphism and S-DNA." In: *J. Chem. Phys.* 147 (2017), p. 225102.

[170] A. K. Sahoo, B. Bagchi, and P. K. Maiti. "Unfolding Dynamics of Ubiquitin from Constant Force MD Simulation: Entropy-Enthalpy Interplay Shapes the Free-Energy Landscape." In: *J. Phys. Chem. B* 123 (2019), pp. 1228–1236.

[171] B. J. Berne, G. Ciccotti, and D. F. Coker. *Classical and Quantum Dynamics in Condensed Phase Simulations.* World Scientific Publishing Co. Pte. Ltd., 1998.

[172] S. Cocco, J. F. Marko, and R. Monasson. "Slow nucleic acid unzipping kinetics from sequence-defined barriers." In: *Eur PHys J E: Soft Matter Biol Phys* 10 (2003), pp. 153–161.

[173] Borkovec Michal Hanggi Peter Talkner Peter. "Reaction-rate theory: fifty years after Kramers." In: *Rev. Mod. Phys.* 62 (1990), p. 251.

[174] Josep Maria Huguet, Nuria Forns, and Felix Ritort. "Statistical properties of metastable intermediates in DNA unzipping." In: *Physical review letters* 103.24 (2009), p. 248106.

[175] F Manca, S Giordano, PL Palla, F Cleri, and L Colombo. "Monte Carlo simulations of single polymer force-extension relations." In: *Journal of Physics: Conference Series.* 383.1 (2012), pp. 012016–.

[176] F Manca, S Giordano, PL Palla, F Cleri, and L Colombo. "Two-state theory of single-molecule stretching experiments." In: *Physical Review E* 87.3 (2013), p. 032705.

[177] Jeff Hooyberghs, Paul Van Hummelen, and Enrico Carlon. "The effects of mismatches on hybridization in DNA microarrays: determination of nearest neighbor parameters." In: *Nucl. Acids Res.* 37 (2009), e37.

[178] Jan Gieseler et al. *Optical Tweezers: A Comprehensive Tutorial from Calibration to Applications.* 2020. arXiv: 2004.05246 [physics.optics].

[179] I. Sela-Culang, V. Kunik, and Y. Ofran. "The structural basis of antibody-antigen recognition." In: *Frontiers Immunol.* 4 (2013), p. 302.

[180] C. E. Chivers, A. L. Koner, E. D. Lowe, and M. Howarth. "How the biotin–streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer." In: *Biochem. J.* 435 (2011), pp. 55–63.

[181] F. Pincet and J. Husson. "The solution to the streptavidin-biotin paradox: the influence of history on the strength of single molecular bonds." In: *Biophys. J.* 89.6 (2005), pp. 4374–4381.

[182] U. Seifert. "Stochastic thermodynamics, fluctuation theorems and molecular machines." In: *Rep. Prog. Phys.* 75 (2012), p. 126001.

[183] Joan Camunas-Soler, Silvia Frutos, Cristiano V Bizarro, Sara de Lorenzo, Maria Eugenia Fuentes-Perez, Roland Ramsch, Susana Vilchez, Conxita Solans, Fernando Moreno-Herrero, Fernando Albericio, et al. "Electrostatic binding and hydrophobic collapse of peptide–nucleic acid aggregates quantified using force spectroscopy." In: *ACS nano* 7.6 (2013), pp. 5102–5113.

[184] Alan B Poritz. "Hidden Markov models: A guided tour." In: (1988), pp. 7–13.