UNIVERSITAT DE BARCELONA

Calorimetric force spectroscopy experiments in DNA and protein folding

Marc Rico Pastó

Tesi doctoral

Calorimetric force spectroscopy experiments in DNA and protein folding

Marc Rico Pastó

Universitat de Barcelona

UNIVERSITAT DE BARCELONA

Universitat de Barcelona

FACULTAT DE FÍSICA

CALORIMETRIC FORCE SPECTROSCOPY EXPERIMENTS IN DNA AND PROTEIN FOLDING

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

Director: Dr. Felix Ritort Farran

Tutor: Dr. Giancarlo Franzese

Setembre de 2021

AGRAÏMENTS

Després de cinc anys plens de bons moments i alguns una mica costa amunt, ha arribat el moment de concloure aquesta tesi i d'agrair les vivències i les persones que m'han acompanyat en aquest viatge.

Voldria començar pel director d'aquesta tesi, el Professor Felix Ritort. Gràcies per obrir-me el teu laboratori ja fa vuit anys. Qui m'havia de dir llavors, fent el projecte final de grau, que gaudiria tant de la recerca. Gràcies per la confiança i per encomanar-me la teva passió per la ciència, per empènyem a ser curòs i precís, i en definitiva, per guiar-me en aquests primers passos com a investigador i ensenyar-me a fer les preguntes adequades per anar un pas més enllà.

També voldria agrair tota l'ajuda que em va donar el Marco Ribezzi i l'Anna Alemany durant els meus primers anys al laboratori. Que hauria estat de mi sense vosaltres!
Marco, sempre recordaré quan em vas dir: ¿Quieres mis programas o discutimos la teoría? Des de llavors discutir amb tu ha estat, essencialment, una de les experiències més enriquidores d'aquest viatge. Gràcies per dedicar-me el teu temps i per ajudar-me quan anava ben perdut. Tenim pendent un vermutet, eh?!
I tu Anna, que vas passar de ser la profe canyera de problemes de física estadística a ser la meva companya de despatx i, amb el temps, algú que considero una bona amiga. Probablement, mai sabràs quant n'he aprés de tu! Professional i personalment. Els meus anàlisis de dades sempre agraïran que compartisim despatx. I jo, les nostres xarrades que netegen l'ànima. Gràcies de tot cor.

Els següents que em venen al cap són els veterans del laboratori, de qui vaig aprendre molt durant el temps que vam compartir: en Joan Camuñas, la Maria Martí i la Maria Mañosas. Del Joan vaig aprendre

que la tranquil·litat i la perseverança són clau en els experiments, i de la Maria Martí a ser curós amb l'instrument (quantes hores vam passar allà dins tancats!). En els últims temps, gràcies al projecte PROSEQO, he gaudit molt de treballar amb la Maria Mañosas.

Una de les coses per la que estic més agraït és per haver compartit aquesta aventura amb tres persones molt especials: l'Álvaro, el Xavi i la Marta. Res hauria estat el mateix sense vosaltres. Álvaro, estuvimos compartiendo clase durante años aunque apenas nos conocíamos y a raíz de compartir despacho, te convertiste en uno de mis mejores amigos. Siempre que piense en esta etapa recordaré nuestros viernes por la tarde con música de Queen, las cervezas en cualquier terraza arreglando el mundo y nuestros debates futboleros - Visca el Barça!. Gracias por todo. I amb el Xavi, quins records del viatge a Madrid els tres, per mi va ser un punt d'inflexió al laboratori. Xavi, hem fet una tesi junts, hem pujat a pics nevats als Alps i hem anat explicant per Madrid que la Voll-Damm és catalana. Gràcies per empenye'ns quan més ho necessitàvem. Has estat el motor d'aquest equip. I a tu, Marta, que vas arribar per revolucionar el V334, gràcies per encomanar-nos la teva força. Vas escollir un camí diferent dins del laboratori, amb totes les complicacions que això implica, i aquí estem, dipositant junts les nostres tesis. Finalment, ho tenim!

A la Isabel, la persona amb la que més hores he passat al laboratori: gràcies. Por compartir tu experiencia y aconsejarme, por hacer más llevaderas las jornadas maratonianas en el laboratorio con las pinzas, por acogerme en tu despacho y por nuestras charlas de camino a casa. Has sido una parte muy importante de este viaje.

També volia agrair l'ajuda d'una persona que encara que va estar poc temps al laboratori em va marcar molt, l'Annamaria. He aprendido y disfrutado muchísimo trabajando contigo y con nuestra querida Barnasa. Gracias a ti existen dos capítulos de esta tesis. Siempre has encontrado un momento para hacer un Skype y animarme en este último tramo. Grazie mille!

Als nous doctorants, Paolo, Jaime i Alejandro; encara que hem pogut compartit poc temps pel laboratori, sempre m'heu fet passar una bona estona. Us queda un camí dur però gratificant, aprofiteu molt el temps que passa volant! Ah, i cuideu de les pinces!
A la colla del moment més divertit d'aquest laboratori: Laura, Carmina, Teo, Sebastian, Regina, Patrick, Fabio, Vegard i Aurelien; que sapigueu que tinc l'espasa làser que em veu regalar per controlar les pinces al menjador de casa. *May the force be with you*.
I us recordo amb especial estima a vosaltres, els meus *padawans*: Aleix, Carla, David i Laia. Vaig aprendre molt ajudant-vos a fer el treball de fi de grau. Gràcies pels bons moments.

També penso en els meus *cool physicists*, amb qui fa temps que no comparteixo el dia a dia però seguim donant-nos suport. Sempre em feu riure amb els vostres 200 missatges de Whatsapp en menys 15 minuts. Aleix, Ignasi, Irini, Laura, Lluís, Maria, Néstor i Núria; sembla impossible però s'ha acabat!

A partir d'aquí, voldria agrair a tots aquells que sempre esteu fent-me costat incondicionalment, la família i els amics.

Els que m'acompanyeu d'ençà que érem uns nens, Guille, Ignasi, Néstor i Sebas. Encara que estiguem esparcits pel món, el poc temps que compartim em carrega les piles i em fa tornar amb energia renovada. Gràcies per formar part del meu viatge. Us estimo.

Lis y Cande, que llegasteis hace poco a mi vida pero desde entonces me habéis acompañado con muchas risas y un montón de vermuts. Gracias por vuestro apoyo e interés durante todo este tiempo. Parece imposible pero se acabó el día de la marmota! No hace falta que os diga donde vamos a celebrarlo...

Aquesta tesi no hauria estat possible sense el suport dels tiets (Josep Maria i Rosa, Edu, Conchi i Josep), cosins (Albert, Tis i Adri) i de les tres maries (Merche, Amparo i Toñi). Gràcies per acompanyar-me i per la vostra energia. Hi ha dues persones fonamentals a la meva vida que

sempre m'acompanyen, l'Àvia i l'Oscar. Us trobo a faltar, aquesta tesi també va per vosaltres.

Àngel i Susanna, gràcies per voler formar part d'aquesta tesi. Quina paciència teniu amb mi! Gràcies per el vostre suport incondicional tots aquests anys i per les sobretaules discutint d'experiments amb nanopors. Només cal veure on s'ha escrit gran part d'aquesta tesi. Us estimo molt.

Eric, gràcies per fer-me riure i enfadar quan més ho he necessitat, ets el millor germà que es pot tenir. I a vosaltres, Mama i Papa, no sé ni que escriure, us estimo. M'heu donat el suport i les oportunitats que he necessitat sempre, empenyent-me a anar més enllà. Sou els pares que un desitja i jo en gaudeixo cada dia.

Les últimes paraules les vull dedicar a la coautora d'aquesta tesi, l'Anna. Ens vam conèixer ja fa gairebé setze anys i no ens hem separat de llavors ençà, i que bé ens ho hem muntat! Gràcies per donar-me les empentes que necessito per seguir endavant. Aquesta tesi no existiria sense tu, i aquest Marc que escriu, tampoc. Tot el que vingui després d'aquesta aventura sé que serà fantàstic perquè ho farem junts. T'estimo, tenim feta la tesi!

Palamós i Barcelona, Setembre de 2021

CONTENTS

ACRONYMS

AFM	Atomic Force Microscopy
CEBA	Continuous Effective Barrier Approach
CTF	Continuous Time Feedback
dsDNA	double-stranded DNA
DTF	Discrete Time Feedback
eCEBA	extended Continuous Effective Barrier Approach
FDCs	Force Distance Curves
FECs	Force Extension Curves
FJC	Freely Jointed Chain
FT	Fluctuation Theorem
MT	Magnetic Tweezers
OT	Optical Tweezers
SMEs	Single-Molecule Experiments
ssDNA	single-stranded DNA
WLC	Worm-Like Chain

RESUM DE LA TESIS EN CATALÀ

Durant el segle XX, amb el desenvolupament de la física quàntica, el camp de la física va fer un gran salt convertint-se en un motor de desenvolupament tecnològic. Un dels avenços més significatius del moment va ser la invenció de làser (acrònim de *light amplification by stimulated emission of radiation*). Aquest dispositiu basat en fonaments de la física quàntica, com és l'emissió estimulada, va ser l'ingredient clau per la invenció de les pinces òptiques. L'any 1970 en els laboratoris d'IBM, Arthur Ashkin va demostrar que la pressió de radiació de la llum pot ser utilitzada per controlar objectes microscòpics. Per l'invent de les pinces òptiques, A. Ashkin va ser guardonat amb el Premi Novel de Física l'any 2018.

Les pinces òptiques, juntament amb altres tecnologies com les pinces magnètiques i el microscopi de força atòmica, va donar lloc al sorgiment dels experiments de molècula individual. Aquests experiments van significar un gran avenç en el camp de la biofísica i la física estadística. Des del punt de vista de la biologia, aquests experiments van permetre mesurar per primera vegada el plegament molecular amb una resolució espacial i temporal sense precedents, permetent veure clarament estats cinètics que d'altra forma eren emmascarats per les condicions experimentals. A més a més, des d'un punt de vista físic, aquests experiments han permès estudiar sistemes fortament influenciats per fluctuacions tèrmiques. Aquest fet ha permès realitzar experiments que només es podien estudiar des d'un punt de vista teòric en el passat. La present tesi s'ha realitzat sota aquests tres paraigües: experiments de molècula individual, física estadística i biofísica.

En la primera part de la tesi, Part I, es presenta una introducció general al camp de la biofísica, així com els detalls específics de l'aparell

de pinces òptiques utilitzat al llarg de la tesi. En la segona part, Part II, es presenta un estudi en el context de la teoria de la informació termodinàmica. Aquest camp de la física va sorgir arran de l'experiment teòric proposat per James Clerk Maxwell, actualment conegut com a Dimoni de Maxwell, on un ens imaginari extreu treball net a partir de les fluctuacions tèrmiques de les partícules d'un gas dintre d'un dipòsit. L'acció de l'ens és el que es coneix com una retroalimentació o *feedback* en anglès. Actualment, la majoria d'experiments amb protocols de retroalimentació utilitzen sistemes on en absència de feedback el sistema està en equilibri, és a dir no hi ha una producció efectiva d'entropia. En aquesta tesi, s'ha estudiat el cas on el sistema per si sol està fora d'equilibri. Dintre d'aquest marc, s'ha definit un protocol general de retroalimentació anomenat retroalimentació de primera observació (*First-time feedback* en anglès). Aquest protocol presenta dos casos particulars que han estat estudiats en sistemes en equilibri, com són el protocol discret i continu. Per aquests dos casos particulars, s'han realitzat experiments de molècula individual per determinar la conversió d'informació en energia útil i la reducció del treball dissipat del sistema. Els resultats obtinguts han demostrat que els protocols de retroalimentació poden ser útils per reduir la dissipació en sistemes petits fora de l'equilibri. Tenint present aquest resultat, s'ha anat un pas més enllà i s'ha introduït per primera vegada el concepte de conversió d'informació en mesura. Aquesta conversió quantifica una millora en la determinació d'una mesura a partir d'un protocol de feedback. En els experiments realitzats, el que s'ha estudiat és com el protocol millora la mesura d'energia lliure de les molècules. Els resultats obtinguts han demostrat que la reducció en dissipació no és suficient per a millorar la determinació de l'energia lliure. Per finalitzar aquesta part, s'ha introduït un nou concepte anomenat retroalimentació estratègica. Aquesta es basa a concatenar de forma efectiva protocols continus i discrets per reduir la dissipació de forma més eficient. Aquests resultats obren noves preguntes a respondre dintre del camp de la informació termodinàmica.

Als anys setanta, l'estudi del plegament de les proteïnes va prendre una gran notorietat quan Cyrus Levinthal va notar que a causa del gran

nombre de graus de llibertat una cadena polipeptídica desplegada necessitarà un temps astronòmic per plegar-se en l'estructura nativa de la proteïna. Així i tot, experimentalment s'havia observat el plegament molecular amb tècniques de granel. Aquesta aparent paradoxa segueix sent un tema de plena actualitat amb constants descobriments teòrics i experimentals. En aquesta tesi (Part III) s'estudia el plegament de la proteïna Barnasa utilitzant experiments de molècula individual en un ampli rang de temperatures, $7 - 37^{o}$C. En aquesta secció s'ha caracteritzat el paisatge d'energia lliure que governa el plegament i desplegament de la proteïna utilitzant experiments fora d'equilibri i eines teòriques com el teorema de fluctuació i el model cinètic de Bell i Evans. Els resultats obtinguts posen en relleu que el plegament de les proteïnes venen governats per tres desigualtats: $\Delta S^{\dagger} \gg \Delta S^{*}$, $\Delta H^{\dagger} \gg \Delta H^{*}$ i $\Delta C_p^{\dagger} \ll \Delta C_p^{*}$. Aquestes desigualtats ens diuen que l'estat de transició (ET) entre l'estat natiu (N) i el desplegat (D) té una diferència d'entropia i entalpia més gran en relació amb l'estat N que respecte l'estat D. Altrament, el canvi de graus de llibertat és més gran entre D i ET que entre N i ET. Aquest fet denota que l'estat ET té l'estructura bàsica de l'estat N sense tenir tots els enllaços que mantenen l'estructura formada. A més a més, utilitzat el model cinètic de Bell-Evans s'ha determinat la posició de l'estat ET en funció de la temperatura i la força aplicada. Els resultats obtinguts posen de manifest que la força funciona millor que la temperatura per desnaturalitzar una proteïna. Els resultats obtinguts en aquests capítols suggereixen un mecanisme general pel plegament de les proteïnes.

En la quarta part de la tesi (Part IV), s'han determinat les propietats elàstiques, cinètiques i termodinàmiques en funció de la temperatura de diferents tipus de molècules d'ADN. L'ADN és la molècula encarregada de transmetre la informació genètica entre generacions. És per aquest motiu, que l'estudi de l'ADN és a dir d'avui un tema de gran interès en la biologia, la medicina i la física. En el capítol 6, s'ha determinat les propietats elàstiques en funció de la temperatura utilitzant el model elàstic per biopolímers conegut com a *Worm-Like Chain* (cadena de tipus cuc, en anglès). Utilitzant aquest model s'ha vist que la rigidesa de les molècules de cadena simple d'ADN es tornen més rígides quan

s'augmenta la temperatura. Mentre que les cadenes dobles d'ADN es tornen més toves. En futures investigacions es vol modelitzar aquests comportaments per obtenir un coneixement més profund de la naturalesa mecànica de l'ADN. En els capítols 7 i 8, s'ha estudiat l'energia de formació de l'ADN de cada motiu d'ADN en funció de la temperatura. Una caracterització acurada del comportament en temperatura és essencial per exemple per optimitzar les PCR (reacció en cadena de la polimerasa) on es varia la temperatura per multiplicar fragments d'ADN. En el capítol 7 s'ha mesurat l'entropia i entalpia en funció de la temperatura per determinar el canvi de capacitat calorífica d'una parella de bases G-C (guanina-citosina) i A-T (adenina-timina). Per altra banda, en el capítol 8 s'ha determinat l'energia de formació dels setze motius diferents que presenta l'ADN tenint en compte el primer veí d'un parell de bases. En un futur seria interessant ampliar l'anàlisi elaborada en el capítol 8 tenint en compte els mètodes emprats per determinar l'entropia i entalpia en funció de la temperatura en el capítol 7.

Finalment, en la secció V de la tesi (Part V) es presenten els resultats obtinguts de dues col·laboracions fetes durant el transcurs de la realització d'aquesta tesi. Primerament, s'ha dut un pas més enllà un mètode desenvolupat en el passat en el mateix laboratori que permet caracteritzar la barrera cinètica que intervé les transicions entre estats moleculars. En aquest cas, s'ha analitzat unes molècules que presenten estats intermedis al llarg del plegament i desplegament, així com s'han estudiat dues molècules petites d'ADN a diferents temperatures. En segon lloc s'ha estudiat per mitjà d'experiments de molècula individual l'enllaç del fàrmac antitumoral conegut com a Netropsina. Els resultats presentats demostren que la Netropsina s'enllaça a l'ADN de dues formes. De forma no específica a regions riques en adenines i timines i de forma específica al motiu AATT.

Part I

PRELIMINARIES

1

GENERAL INTRODUCTION

The rise of the industrial revolution brought out the interest in thermal machines and motors, resulting in a significant boost in the science of thermodynamics. Concretely, thermodynamics settled the possibility to characterize the temperature, pressure or heat capacity of macroscopic systems by only considering their relation with their surroundings independently of their internal structure. One of the greatest exponents of the development in thermal machines is the Carnot cycle published by Sadi Carnot (1796 – 1832) in 1824 in his book *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance* [1].

The Carnot cycle is an ideal theoretical thermodynamic cycle that converts heat from a thermal bath to work, which provides a theoretical upper limit on the efficiency of classical macroscopic engines. Aiming to reach this ideal limit, in 1893, Rudolf Diesel (1858 – 1913) proposed a motor based on the Carnot cycle with an expected efficiency of 70% [2]. Nowadays is worldwide known as Diesel's motor with car's motor efficiencies around 50%. The existence of an upper-efficiency limit opened the way for the formulation of the second law of thermodynamics. Briefly, the energy required to do any thermodynamic transformation is always equal to or larger than the system's energy. In other words, the second law says that the work necessary to bring a macroscopic system from an initial state to a final state is always equal (reversible processes) or larger (irreversible processes) than the energy difference between the states. Nevertheless, in microscopic systems, the

energy difference is on the same order of magnitude as the energy exchanged between the environment and the system. This fact is relevant in the context of thermodynamic cycles because identical repetitions of the same process require different work, and rare trajectories where the work is less than the energy difference violate the thermodynamic second law. This fact highlights the relevance of fluctuations to describe thermodynamic transformations in such systems and stresses the need to define a new framework to study these systems.

Statistical thermodynamics is the framework that uses statistics and probability theory to explain the connection between the macroscopic properties of materials and the microscopic behaviors of their constituent particles. Indeed, the statistical thermodynamics proves that in microscopic systems, the average work done under a thermodynamic cycle considering a large number of repetitions always satisfies the thermodynamic second law. While classical thermodynamics explains the response of macroscopic systems under thermodynamic equilibrium, i.e., no net flows either within a system or between systems, statistical thermodynamics has been applied in non-equilibrium or irreversible processes that are driven by drivings. The Austrian physicist Ludwig Boltzmann (1844 - 1906) developed the fundamental interpretation of entropy for any system in terms of collecting accessible *micro-states*. Moreover, from the energy levels of the different micro-states, a partition function can be derived to calculate all the thermodynamic properties of the system, e.g., the energy difference between the initial and final states.

The statistical framework brings us the possibility to derive the thermodynamic properties, e.g., entropy or heat capacity, of microscopic systems such as cells, biomolecules, or viruses, from equilibrium or non-equilibrium conditions. From a physical perspective, the mechanisms inside cells, e.g., the molecular motors, are of particular interest. Interestingly, although the strong dependence on fluctuations, recent experiments have found that molecular machines, like F1–ATPase have efficiencies near the 100% [3].

This thesis focuses on the study of characterizing the kinetic properties at different temperatures of molecular systems under applied mechanical forces to map the molecular free energy landscape as a function of temperature. Moreover, this thesis focus on the study of the information-energy conversion in feedback protocols applied on nucleic acids molecules. In this chapter, the context to study such molecular systems is presented. In section 1.1 the main ideas of thermodynamics of small systems are sketched. In section 1.2, a brief introduction to single-molecule force spectroscopy techniques is provided. Finally, a summary of the work presented in this thesis is given in section 1.3

1.1 THERMODYNAMICS OF SMALL SYSTEMS

As previously mentioned, statistical mechanics defines the perfect framework to study thermodynamics from a probabilistic perspective. In a system with N elemental units (e.g., atoms, particles, molecules, etc.), the measurable thermodynamic properties are the result over all the different energy levels occupied by the N units. In macroscopic systems, the number of particles tends to infinity, so the fluctuations ($\sim 1/\sqrt{N}$) are negligible. On the contrary, for small systems where $N \to 1$, the fluctuations are of great relevance [4]. Concretely, for small systems, the typical energy exchanged between the system and the surrounding are of the order of Brownian fluctuations, i.e., $\sim k_B T$, where k_B is the Boltzmann constant and T the temperature of the bath. The study of small systems has become a trending topic since the development of single-molecule techniques and the progress of stochastic thermodynamics.

In biophysics, which is a bridge science to connect physics and biology, methodologies and theories of physics are used to understand biological systems. Such biological systems range from single biomolecules, like nucleic acids and proteins, to living cells or populations. Traditionally, nucleic acids and proteins are studied from bulk assays, such as calorimetric experiments or NMR spectroscopy, where the studied sample contains a large number of molecules (in the order of Avogadro

number, $N_A = 6.02 \times 10^{23}$). In contrary, since the development of single-molecule experiments (SME), one molecule is studied individually by exerting mechanical forces to derive its thermodynamic properties.

1.2 FORCE SPECTROSCOPY TECHNIQUES

There are different single-molecule techniques used to characterize the kinetic and energies of nucleic acids and proteins, such as super-resolution microscopy based on single-molecule fluorescence (e.g., FRET or STORM), single-channel recording, and single-molecule forces spectroscopy techniques, to say some examples. Among all existing single-molecule techniques, in this thesis we will focus in force spectroscopy techniques.

The most common single-molecule force spectroscopy technique are Magnetic Tweezers (MT), Optical Tweezers (OT), and Atomic Force Microscopy (AFM). The history of these techniques is the following. First, in 1949, Francis Crick (1916 – 2004) and Arthur Hughes demonstrated that tiny magnetic beads, phagocytoced by whole cells grown in culture, could be manipulated by an external magnetic field [5]. This experiments were the embryo of the actual magnetic tweezers. Second, in 1970, Arthur Ashkin (1922 - 2020) reported for the first time the detection of optical forces pushing dielectric micron-sized particles [6]. One of the results of this research was the invention of the optical tweezers in 1986. For this invention, Ashkin earned the 2018 Novel Prize of Physics. Third, in 1986, Gerd Binnig (1947 -) and Heinrich Rohrer (1933 – 2013) developed the famous Atomic Force Microscopy, which is a high-resolution type of scanning probe microscopy with resolutions $\sim$1000 times better than the optical diffraction-limit [7]. For this invention Binnig and Rohrer earned the 1986 Nobel Prize for Physics.

The three presented techniques, MT, OT, and AFM, exert forces in the pico Newton (pN = 10^{-12} N) range and measure molecular extensions with nanometric resolution (nm = 10^{-9} m) [8]. This range of force and distance allows for work values in the order of pN$\times$nm $\sim k_B T$ at room

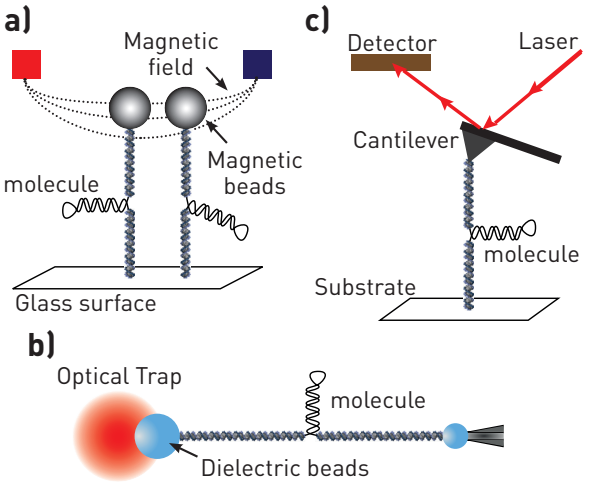

Figure 1.1.: **Single-molecule force spectroscopy techniques.** (a) Magnetic Tweezers. (b) Optical Tweezers. (c) Atomic Force Microscopy

temperature, 25°C, allowing to measure thermal fluctuations and molecular energies. The main features of the three techniques are detailed below, accompanied by a schematic cartoon of the three experimental instruments in Fig.1.1.

- MT (Figure 1.1a) are based on the principle that magnetic objects subjected to an external magnetic field $\vec{B}$ experience a force $\vec{f}$ proportional to the gradient of $\vec{B}$, $\vec{f} \propto \nabla \vec{B}$. The instruments based on this principle operate as follows: the nucleic acids or protein molecules are tethered between two molecular handles. One handle is attached at a glass surface, while the other is attached at the surface of microscopic magnetized beads. By moving the magnets, which generate the magnetic field, relative to the glass surface we change the force exerted to the molecular construction. The molecular extension, x, at a given force (magnet position) is determined by the equipartition law: $f = k_B T \cdot x / \langle \Delta x^2 \rangle$. The stiffness of these instruments are proportional to the size of the beads, the magnetic

field, and the molecule under study. However, the typical stiffness of the magnetic traps are $\sim 10^{-4}$pN/nm covering a force range between 10^{-2} and 10^{2} pN.

Magnetic tweezers present two important advantages over the other two instruments. On the one hand, MT allow exerting torque on the molecules, allowing to study the elastic and torsional properties of the molecules. On the other hand, MT allow parallel force measurements of different beads simultaneously, facilitating data collection.

- OT (Figure 1.1b) are based on the optical gradient forces generated by a highly focused laser beam interacting with a dielectric microscopic bead with a refraction index higher than that of the surrounding medium (more details will be provided in chapter 2). A typical experimental setup is as follows: the molecule under study is tethered via molecular handles between two dielectric beads (typically on the order of 1-3 μm). One bead is kept fixed by air suction (or using a fixed optical trap), while the optical trap controls the other bead. Moving the optical trap relatively to the fixed bead allows us to exert forces on the molecule. The connections between the handles and the beads are specifically chosen to be different between beads, but both strong enough to do not break during the experiments (see chapter 2 for further details).

  The operating force range in OT instruments is $10^{-1} - 10^{2}$ pN with a force resolution of 10^{-1} pN and trap stiffness close to $10^{-1} - 10^{-2}$ pN/nm. All these parameters can be changed by varying the bead size, the difference between the refraction index of the beads and the surrounding medium, and the laser power. One of the main advantages of the OT is the possibility to directly measure the force exerted on the bead without post-processing techniques. Moreover, the OT can be easily combined with other single-molecule techniques, such as fluorescence or single nano-channels sensors, expanding the instrument's range of applications.

- AFM (Figure 1.1c) is widespread used in topographic imaging, however, AFM is also used as a force spectroscopy tool. The principle to use the AFM as force spectroscopy tool is the following: One end of the molecule under study is connected at the tip of a cantilever, while the other is fixed in a substrate. In the backplane of the tip, we focus a laser to measure the depletion of the light when the substrate is moved up and down. The depletion of the light is proportional to the force exerted on the molecule. The typical force range of AFM instruments is $10^1 - 10^4$ pN, with stiffness ~ 10 pN/nm. This stiffness gives a spatial resolution of ~ 1 Å.

All these techniques has been used to characterize the thermodynamic properties and to measure the molecular free energy landscape that governs the folding process different biomolecules in *in vitro* conditions [9–11].

1.3 SUMMARY OF THE THESIS

In this thesis, equilibrium and non-equilibrium single-molecule experiments are carried out to derive the thermodynamic, kinetic, and elastic properties of different molecular constructions. Furthermore, the essential protein folding problem has been studied from the obtained results. In addition, the scale and resolution of such systems provide an excellent platform to test the new developments in the field of thermodynamic information. The results of this thesis are divided into six parts.

The first part of this thesis, Part I, is divided into two chapters to provide the basic concepts to understand the results of the thesis. In chapter 1, a brief introduction to the single-molecule biophysics field and a summary of the force spectroscopy techniques are provided. In chapter 2, the molecular constructions used in this thesis are presented (section 2.1). Moreover, a detailed description of the physics behind the optical trapping phenomenon is provided in section 2.2. In addition, in

section 2.2, the main features of the instrument used in this thesis are presented.

In Part II, different kinds of feedback protocols are employed to investigate nucleic acids hairpins' unfolding and folding process under different experimental conditions. The obtained results shown in chapter 3.

In Part III, the protein folding problem is studied from pulling experiments using the paradigmatic two-state protein barnase. The obtained results are presented in two chapters

- In chapter 4, the *temperature-dependent* enthalpy, entropy and energy of barnase are derived using calorimetric force spectroscopy experiments.
- In chapter 5, the molecular *free-energy landscape* of protein barnase is derived from kinetic measurements using the Bell-Evans model.

In Part IV, the temperature dependence of thermodynamic, elastic, and kinetic properties of nucleic acid hairpins are derived using a temperature-jump optical trap. The results from this investigation are divided into three chapters.

- In chapter 6, the elastic properties of single-stranded and double-stranded DNA molecules are derived from pulling experiments in a temperature range between $5 - 50^{o}$C.
- In chapter 7, DNA hairpins with specific sequences have been used to derive the temperature-dependent enthalpy, entropy, and free-energy difference for GC and AT base pairs. Furthermore, from the temperature dependence of these thermodynamic potentials, the heat capacity change of each base pair has been derived.
- In chapter 8, unzipping experiments at different temperatures have been carried out to derive the temperature-dependent free-energy difference for the 16 different DNA motifs of the nearest-neighbor model.

In Part V, the results of two collaborations based on single-molecule experiments are presented.

- In chapter 9, the molecular free-energy landscape of different DNA hairpins exhibiting different *intermediate state* along folding and unfolding are studied to provide new insights into the molecular folding problem.
- In chapter 10, kinetic measurements derived from non-equilibrium pulling experiments are used to characterize the binding of *netropsin* to DNA hairpins.

Finally, in Part VI the main conclusions derived along this work are summarized and future lines of research are presented.

2

EXPERIMENTAL SETUP

This chapter presents an introductory description of the biomolecules (i.e., nucleic acids and proteins) used in this thesis and a general description of the optical tweezers instrument used throughout the thesis.

2.1 BIOMOLECULES

Biomolecule is the term used to refer to biological molecules present in organisms. These molecules play several roles in different biological processes, such as cell replication, ion transport, or storing genetic information. Such molecules are often endogenous, i.e., produced within the organism, but cells usually need exogenous biomolecules, i.e., produced outside the organism, to carry out biological processes. Biochemistry and biophysics study the formation and reactions of biomolecules, e.g., drug delivery, protein formation, and nucleic acid thermodynamics.

See below a brief biochemical description of the biomolecules studied in this thesis.

2.1.1 *Nucleic acids*

Nucleic acid is the umbrella term for deoxyribonucleic acid (DNA) and ribonucleic acids (RNA) molecules, which are essential for life as they

store and carry the genetic information for the development, growth, and reproduction of all known organisms [12].

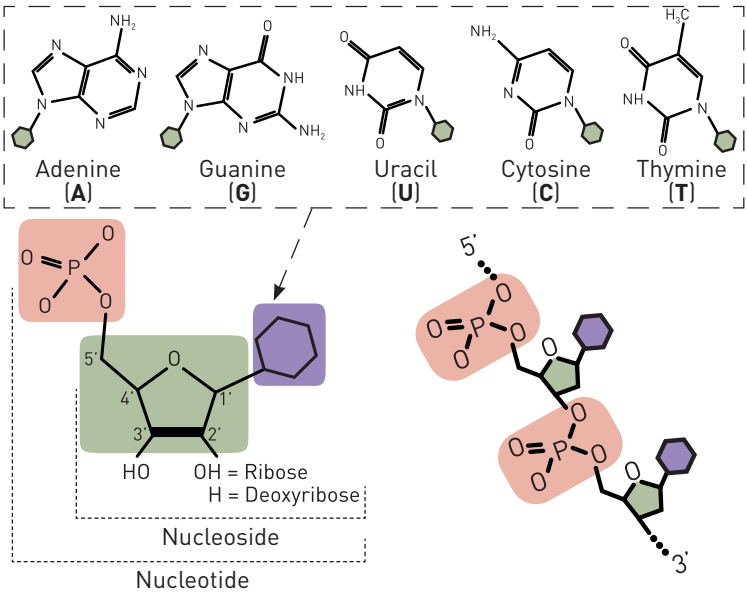

Figure 2.1.: **Chemical structure of nucleic acids.** The chemical structure of a nucleotide is formed by a phosphate group (red), plus a pentose sugar (green) and a nucleobase (purple). The nucleotides are connected by phosphodiester bonds between the phosphate group and the 5-carbon sugar.

Nucleic acids are biopolymers composed of *nucleotides*; the latter are compounds formed by a phosphate group and a *nucleoside*. The nucleoside is composed of a nitrogenous base or nucleobase and a pentose (i.e., five-carbon sugar) (Fig. 2.1). The pentose can be either a ribose in RNA or deoxyribose in DNA. Five different nucleobases exist which are: Adenine (A), Guanine (G), Uracil (U), Cytosine (C), and Thymine (T). While A, G, and C are common for RNA and DNA, U appears only in RNA and T in DNA. The bases are classified into two different groups: purines (A and G) and pyrimidines (C, U, and T). A phosphodiester bond links individual nucleotides from the 3′ sugar carbon to the 5′ the

following sugar carbon forming a nucleotide chain. The prime notation denotes the directionality of the molecule (Fig. 2.1). Finally, the structure of nucleic acids are ranked into four levels [13]:

- **Primary structure**: It corresponds to the nucleotide sequence.
- **Secondary structure**: It refers to interactions between bases.
- **Tertiary structure**: It indicates the three-dimensional structure of the molecule.
- **Quaternary structure**: It refers to ensembles between different tertiary structures.

In particular, this thesis is focused on the study of thermodynamic and kinetic properties of DNA molecules.

Deoxyribonucleic acid (DNA)

DNA molecules are found inside the cell nucleus, and mitochondria of Eukaryotic cells combined with packaging proteins called histones [14]. The complex DNA-histones is named *chromosomes* and its primary function is to store, preserve and transmit the genetic information from generation to generation.

In 1953, James Watson and Francis Crick discovered the tertiary structure of DNA molecules, i.e., the 3D structure of DNA was resolved [15]. This discovery was possible thanks to the crucial experimental works done by Maurice Wilkins, and Rosalind Franklin [16, 17]. Those results show that DNA is formed by two anti-parallel single-stranded DNA (ssDNA) molecules (see Fig. 2.1), one oriented from the direction $5'\rightarrow 3'$ whereas the other runs from $3'\rightarrow 5'$. The resulting double-stranded DNA (dsDNA) molecule is a right-handed helix keeping the negatively charged phosphates in the major groove of the helix (see Fig. 2.2). The nitrogenous bases are located in the inner part of the double helix and therefore are physically protected from the surrounding medium. The two nucleic acid strands are joined by base-pairing, mainly following the canonical Watson-Crick complementary rule (Fig. 2.2):

- **A** interacts with **T** via two hydrogen bonds: $A = T$
- **G** interacts with **C** via three hydrogen bonds: $G \equiv C$

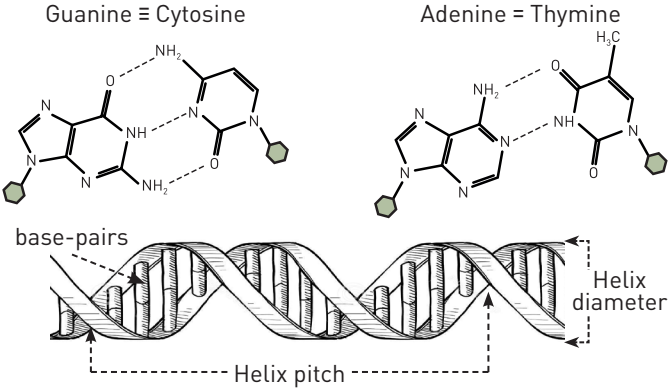

Figure 2.2.: **Structure of dsDNA.** Canonical Watson-Crick base-pairs for DNA (top). Illustration of the double-helix pointing out its structural parameters (bottom).

The most common double-helices structure of canonical dsDNA molecules is named B-form. This structure consists of a right-handed helix with a diameter of 2.1nm with 10 base-pairs (bp) per turn. The distance between consecutive bp is 0.34nm [18]. Nevertheless, in solutions with alcohol or high ionic strength other structures have been observed, such as the A-form (right-handed double helix with 11bp per turn) or the Z-form (left-handed double helix with 12bp per turn).

2.1.2 *Proteins*

Proteins are macromolecules carrying out several functions within organisms, e.g., DNA replication, contributing to structure cells and organisms, and exerting mechanical forces inside and outside cells. The shape and function of the proteins are defined by their amino acid sequence, which is codified in the DNA. These biomolecules are formed by one or more long chains of amino acids residues named polypeptide chains.

Amino acids are the building compounds of the polypeptide chains, analogous to the nucleotides in nucleic acids. Amino acids are organic compounds that contain amino ($–NH_2$) and carboxyl (–COOH) functional groups, along with a side chain specific to each amino acid. The genetic code specifies 20 standard amino acids. Amino acids are linked together through covalent bonds between the amino group of one amino acid and the carboxyl group of the following amino acid. Like amino acids, the structure of proteins are ranked into four levels:

- **Primary structure**: It corresponds to the protein sequence, i.e., the amino acid sequence. The sequences are read from the amino-termini (N-termini) to the carboxyl-termini (C-termini).

- **Secondary structure**: It refers to interactions between bases amino acids. The most abundant secondary structures are:

  - *α-helices*: right-handed spirals.

  - *β-sheets*: β-strands (two polypeptide chains joined in parallel) connected laterally by at least two or three backbone hydrogen bonds.

  - *random coils*: stable structures with amino acid chains collapsed randomly.

- **Tertiary structure**: It indicates the three-dimensional structure of the protein (combination of α-helices, β-sheets, and random coils.

- **Quaternary structure**: It refers to ensembles between proteins.

This thesis is focused on the thermodynamic study of protein Barnase. Then, some specific details of this protein will be presented below.

Barnase

Barnase is a 110 amino acids bacterial ribonuclease globular protein secreted by *Bacillus amyloliquefaciens*, which in physiological conditions degrades RNA in the absence of its protein inhibitor barstar [19]. The

high solubility and stability of barnase makes a paradigmatic protein to investigate the folding kinetics of globular proteins, by combining phenomenological approaches (e.g., the phi-value analysis) with protein engineering and site-directed mutagenesis methods [20].

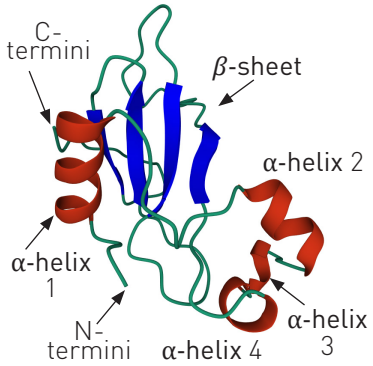

Figure 2.3.: **Structure of Barnase.** 3D structure of Barnase obtained with 1.5Å resolution. In red are highlighted the α-helices, in blue the β-strands (a single β-sheet), and in green the random coiled polypeptide chain.

The tertiary structure of native barnase has been obtained with X-ray diffraction with 1.50 Å resolution [21]. The structure is formed with four α-helices (helix 1: Phe7-Tyr17; helix 2: Lys27-Leu33; helix 3: Ala37-Lys39; helix 4: Leu42-Val45) containing a total of 25 amino acids surrounding a single β-sheet (with four β-strands) located in the protein core. The rest of the amino acids are in a random coiled conformation (Fig. 2.3).

2.2 OPTICAL TWEEZERS

The understanding of light has been and continues to be one of science's motors. In the 10th century, Hasan Ibn al-Haytham (965 - 1040) did revolutionary works and experiments with lenses and mirrors to explain the reflection and refraction and to understand the vision mechanisms. For his works, Ibn al-Haytham is worldwide recognized as the *father of*

modern optics. A Latin translation of his most widely recognized work, *Kitab al-Manazir*, was made in the late 12th or early 13th century and influenced the intellectuals of the moment in Europe to investigate the nature of light.

In the 13th century, Roger Bacon (1215 - 1294), inspired by the works done by Ibn al-Haytham 300 years before, realized that lenses could be used to correct human vision. In the 16th century, the invention of the first optical microscope was a tipping point in science, especially in biology. In 1665, Robert Hooke (1635-1703) detailed the first microscopic observations and defined, for the first time, the concept of the cell. Afterward, Anton van Leeuwenhoek (1632 - 1723), considered one of the *fathers of microbiology*, wrote approximately 560 letters to scientific institutions concerning his observations and discoveries and popularised the use of optical microscopes. In parallel, there was a great debate related to the nature of light. On the one hand, Robert Hooke, among others, considered light as a wave to explain reflection and refraction. On the other hand, Isaac Newton (1642 - 1727) explained white light as a mixture of particles of different colors. For a century, Newton's theory was the accepted one, thanks to its prestige.

At the dawn of the 20th century, Max Planck (1858 – 1947) demonstrated that the energy stored within an object should be composed of an integer number of discrete, equal-sized parts (denoted as *quanta*). Later, Albert Einstein (1879 - 1955) proposed the existence of light quanta to explain the photoelectric effect. These light quanta are the so-called *photons*. Photons exhibit wave-particle duality, featuring properties of both waves and mass-less particles. The photon theory is the most consolidated theory nowadays to explain the nature of light. In 1960, the first laser (an acronym for Light Amplification by Stimulated Emission of Radiation), based on the stimulated emission stated by Einstein, was built to study the interaction between matter and light. Moreover, the use of lasers emerged as an influential tool for science and designing new instrumental devices.

2.2.1 *Optical trapping principles*

Geometrical optics is the most commonly used method to illustrate light propagation, reflection, and refraction. This methodology describes the propagation of light by rays, each one representing a certain number N of photons with a momentum, $\vec{P}$, equal to:

$$\vec{P} = N\frac{h}{\lambda}\hat{u} \quad . \tag{2.1}$$

In Eq.(2.1), $\hat{u}$ is a unitary vector indicating the direction propagation of the ray, h is the Planck constant, and λ is the wavelength of light. The conservation of momentum is the physical principle of the optical trapping. To explain optical trapping, let us consider a laser beam with a Gaussian intensity profile passing through a micro-sized (bigger than the light wavelength) sphere that reflects a small amount of light and refracts most of it.

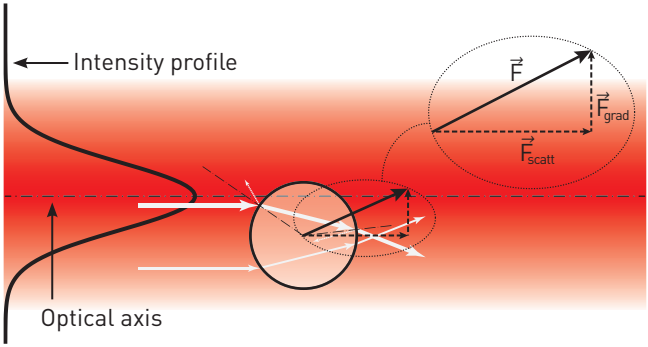

Figure 2.4.: **Optical trapping principles.** Geometrical optical scheme of the interaction between a Gaussian laser beam Gaussian and a microscopic sphere, highlighting the resulting forces.

In a Gaussian laser beam, the rays in the center are more intense than those at the edges (Fig. 2.4), i.e., the number of photons propagating along the center is larger than at the edges. The propagation direction of the laser beam is known as the optical axis. Let us consider now that

one of these rays, e.g., the most intense one, interacts with a micro-sized dielectric sphere (hereafter referred to as a bead). First, a small amount of the incoming ray (thick white arrow in Fig. 2.4) is reflected at the external surface of the bead (dashed thin white arrow in Fig. 2.4). The rest of the light is refracted inside the bead changing its propagation direction. The new direction is calculated using Snell's law,

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \quad . \tag{2.2}$$

In Eq.(2.2), n_1 and n_2 are the refraction indices of the medium and bead, respectively, and θ_1 and θ_2 are the incidence angle respect the normal of the interface in the surrounding medium and bead, respectively. The same procedure is required to calculate the leaving angle when the ray leaves the bead. Optical trapping is achieved when in conditions where $n_1 < n_2$, i.e., water and glass beads.

Finally, considering Newton's second law, a change in the linear momentum is translated into a force exerted on the photons. Moreover, as the force exerted to the photons is a result of the interaction between light and the bead, it means that according to Newton's third law, a force in the opposite direction is exerted to the bead (solid black arrow in Fig. 2.4). This resulting force is commonly named *radiation pressure* or force of light. The total net force exerted to the bead is the resulting radiation pressure of all rays passing through the bead. The net force $\vec{F}$ is decomposed into two different contributions, the *scattering* and the *gradient* forces (dashed black arrows in Fig. 2.4). On the one hand, the scattering force pushes the bead along the optical axis. On the other hand, the gradient force pushes the bead towards the optical axis.

In 1970, Arthur Ashkin (1922 – 2020) experimentally demonstrated that the radiation pressure from visible laser light (514.5 nm) accelerates microscopic glass spheres along the propagation direction [6]. Moreover, in the same work, Ashkin proposed to use two counter-propagating lasers focused at the same point to suppress the scattering force to trap dielectric beads at the focal point. The latter is the main principle of the optical tweezers instrument.

2.2.2 *Optical Tweezers instrument*

This section presents the main features and characteristics of the optical tweezers instrument used throughout this thesis. The instrument used in this thesis is an updated version [22] of the original instrument designed by Steve B. Smith and Carlos Bustamente in 2003 [23]. The instrument used in this thesis is based on the original idea proposed by Ashkin in 1970 with an extra collimated laser to change the temperature near the optical trap.

Optical path

The optical trapping instrument is based on two counter-propagating lasers generating a single optical trap. The optical path of each trapping laser is symmetric concerning the other one (see Fig. 2.5). One of the main characteristics of this instrument is its short optical paths. For this reason, this instrument is named miniaturized tweezers or MiniTweezers. The optical scheme of this instrument is presented in Fig.2.5.

Two identical 845nm wavelength laser diodes (Laser A and B in Fig.2.5) are directed through a single-mode optical fiber to generate and guide the trapping laser beams. The optical fiber is glued inside a brass tube, and two perpendicular piezoelectric actuators tilt the brass tube to change the position of the trap. The junction of the piezoelectric actuators and brass tube is known as a wiggler. At the output of the wiggler, the light exits the optical fiber and is emitted into the air. Both lasers are emitted with linear polarization. Each polarization is perpendicular to the other to avoid *trash-talk* between the two lasers. To explain the function of the elements inside the instrument, we will focus on the optical path of laser B (remember that the optical system is symmetrical for laser A). After leaving the wiggler, the light passes through a polarized pellicle reflecting about 8% of the light and letting the remaining 92% pass through it. Therefore, most of the light is used to form the optical trap. The reflected light passes through a converging lens, which focuses the light

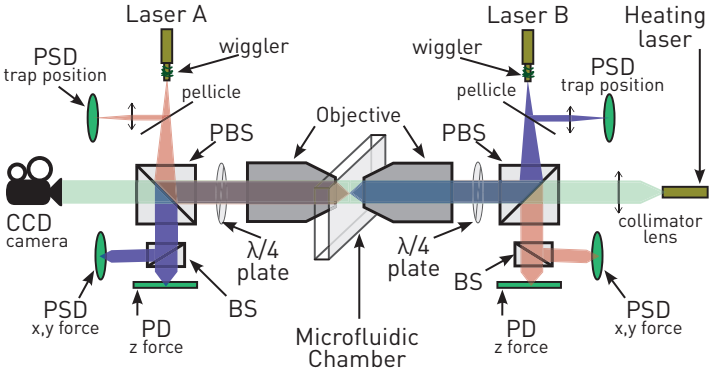

Figure 2.5.: **Optical paths of the MiniTweezers.**

on a photo-sensitive detector (PSD) to measure the trap's position. After passing through the pellicle, the light is collimated and directed to a polarizing beam splitter (PBS), reflecting the light to the microscope objective. However, before focusing the light, it passes through a $\lambda/4$ plate or retarder, which converts the linearly polarized light into circularly polarized light.

The laser beam is focused and subsequently collected via two Olympus x60 1.2NA microscope objectives. The microscope objective that focuses the light is working in under-filling conditions. Then, even the laser beam is tilted, the second objective collects all the rays of the beam. This fact is relevant because it defines one of the most distinctive features of this instrument. The under-filling condition allows us to measure the exerted force by directly observing the deflection of light passing through the trapped object. To directly measure the force, the light is sent to two detectors using a beam splitter (BS). Half of the light is used to measure the scattering force via a photo-detector (PD), while the other half is used to measure the gradient force via a PSD.

Force detection

As already mentioned, one of the most representative features of the MiniTweezers is that force is measured directly from the light deflection. To study how the deflection of light is proportional to the exerted force it is necessary to study the rays passing through the trapped bead (Fig. 2.6). First, let us consider the case where a dielectric bead is placed in

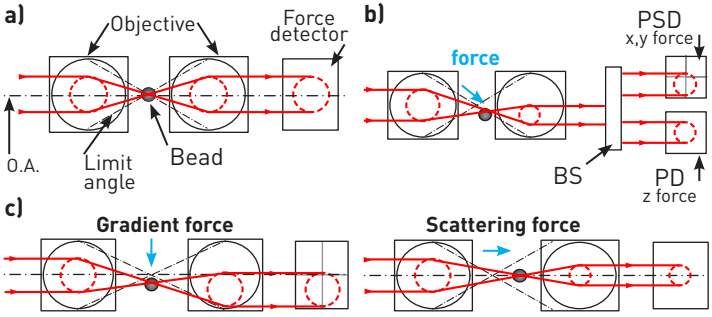

Figure 2.6.: **Force detection.** (a) Light deflection when the bead is in the center of the optical trap, i.e., zero force. (b) Force deflection when the bead is displaced concerning the focal point in the perpendicular (X,Y direction), and propagation (Z direction) direction. (c) Left: Case where only gradient force is applied to the bead. Right: Case where only scattering force is exerted to the bead.

the focal point of the laser, i.e., no force is exerted to the bead (Fig. 2.6a). In that particular case, the rays would not be affected by the presence of the bead. In figure 2.6a, the red rays correspond to the edges of the Gaussian beam, while the dashed black lines correspond to the limit angle of the microscope objectives. Let us consider the case where an external force displaces the bead in the three (X, Y, and Z) directions, i.e., the bead is entirely out of focus (Fig. 2.6b). In this case, the light passing through the bead is deflected and unfocus, i.e., the beam changes its diameter. The light is split into two to measure the gradient force and scattering force. On the one hand, a PSD detector measures the deflection relative to the zero-force case to measure the gradient force

(Fig. 2.6c-left). On the other hand, a PD detector measures the change in diameter to determine the scattering force (Fig. 2.6c-right).

The gradient and scattering forces are calibrated using Stoke's law that relates the drag force f that undergoes the bead moving through a fluid at a constant velocity v:

$$f = 6\pi\mu R v \tag{2.3}$$

where R is the radius of the bead, and μ is the viscosity of the medium. Using beads with known radius and distilled water, the force is calibrated by measuring the deflection and defocus of light as a function of the velocity:

$$f_{x(y)} = C_{x(y)} PSD_{x(y)}(v_{x(y)}) \text{ and } f_z = C_z PD_z(v_z)\,. \tag{2.4}$$

In Eq.(2.4) $C_{x,y,z}$ is the calibration factor, $PSD_{x,y}$ is the measurement from the PSD detector and PD is the PD detector measurement.

Temperature controller

Finally, the last distinguishing feature of the instrument is the capacity to raise the temperature near the optical trap [22]. The temperature increment near the optical trap is achieved by heating the medium using a third 1435nm wavelength laser. This last laser is focused on the back-focal plane of the microscopy objectives. Consequently, the heating laser is collimated at the optical trap position (analogous to Köhler illumination). The heating laser enters the optical from the right arm of the symmetric optical path (see Fig.2.5).

The 1435nm wavelength is chosen due to the light absorption peak of water. Water is the solvent used to prepare the buffer solutions throughout the thesis. This peak is about three orders of magnitude higher than the absorption corresponding to the wavelength of the trapping lasers (845nm).

A Stokes test is performed to determine the temperature inside the microfluidic chamber where the single-molecule experiments are carried out. The temperature inside the chamber at discrete laser powers is determined from the slope of the force versus velocity curves by varying the velocity and the laser power (Fig. 2.7). The instrument is placed in a 25°C temperature-controlled laboratory. From this initial temperature, the heating laser raises the temperature of the buffer surrounding the optical trap from the initial 25°C to around 50°C. Contrary to the force calibration, when the heating laser is turned on, the force, velocity, and radius are known, but the temperature-dependent viscosity is not. Using the Vogel-Fulcher equation equation, the temperature at each laser power is derived from the viscosity.

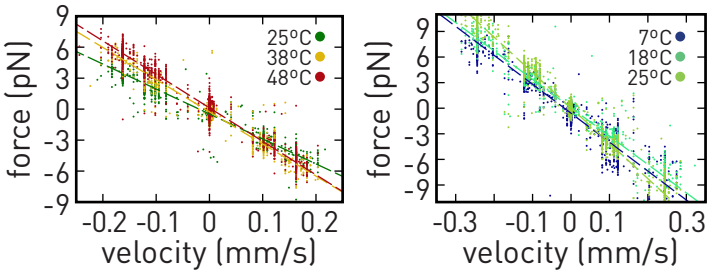

Figure 2.7.: **Stokes' test.** Stokes' test measured inside (right) and outside (left) the icebox.

Moreover, as this instrument is a miniaturized optical path instrument, the instrument can be placed inside an icebox with an internal temperature of around 5°C. This fact allows us to obtain a temperature range ranging from 5°C to 50°C.

2.3 SINGLE-MOLECULE EXPERIMENTS

Commonly, to carry out single-molecule experiments using an optical tweezers instrument is necessary to tether the molecule under study between two micro-sized dielectric beads, where optical traps control one

or both beads. In particular, in this thesis, the molecules are tethered between two dielectric beads with diameters 3.15μm (Kisker Biotechnologies) and 2.17μm (Spherotech, SVP-20-5). The optical trap controls the 3.15μm bead (bead type I), while the 2.17μm bead (bead type II) is kept fix at the tip of a micro-pipette by air suction.

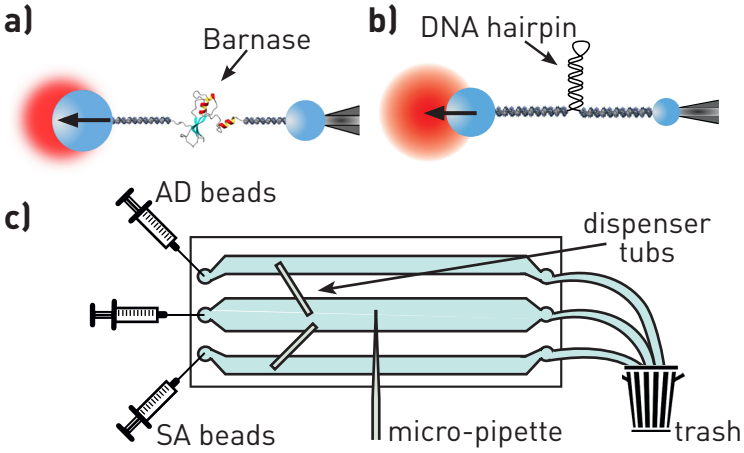

Figure 2.8.: **Microfluidic chambers.** (a) Barnase and (b) DNA hairpins tethered between two dielectric beads. In both cases, the AD bead is controlled with the optical trap, while the SA bead is fixed by air suction at the tip of a glass micro-pipette. (c) Illustration of the microfluidic chambers used throughout this theses. The upper (bottom) channel provides the AD (SA) beads to the central one via a dispenser tube.

This thesis is focused on the study of the thermodynamic properties of DNA and protein molecules. In particular, it is studied the protein barnase (Fig. 2.8a) and different sorts of DNA hairpins (Fig. 2.8b). The synthesis details are provided in App. A. For both kinds of biomolecules, the molecule under study is attached to the beads via dsDNA handles. In addition, these handles act as spacers and prevent non-specific interaction between the beads and the molecule. Protein barnase is attached to two identical 500bp dsDNA in its N- and C-termini via cysteine-thiol

chemical reduction. The DNA hairpins are designed as follows: an oligonucleotide chain with the sequence of the studied hairpin flanked by two 29 bases ssDNA handles are ordered to SigmaAldrich. A complementary 29 bases ssDNA oligo is attached to the ssDNA handle to produce the desired dsDNA handles.

The 5'-end of one handle is labeled with one biotin, while the 3'-end of the other handle is labeled with a digoxigenin. Biotin- and digoxigenin-labeled ends specifically bind to streptavidin and anti-digoxigenin. To have reproducible experiments, the type I beads are coated with anti-digoxigenin (AD bead), whereas the type II beads are coated with streptavidin (SA bead). This procedure is the same for barnase and DNA hairpins.

A sketch of the microfluidic chamber where the experiments are carried out can be seen in Fig. 2.8c. The chamber consists of three microfluidic channels. The experiments are carried out in the central channel, while the top and bottom channels are used as dispenser channels for the AD and SA beads, respectively. With this configuration, the central channel is kept clean of beads during the experiments. The microfluidic chamber is handmade, and it is made by a *NescoFilm* layer, which has the channels cut off, sandwiched between two Cover Glass, No. 1.5, 24 x 60mm. One of the glass coverslips has six holes to introduce and remove the buffer solution from the channels. Two glass dispenser tubes (Garner Glass CO.), with an outer diameter of 100μm and an inner diameter of 40μm, connect the dispenser channels with the central one.

In this thesis have been carried out two different single-molecule experiments: *hopping* and *pulling* experiments.

Hopping experiments

Equilibrium hopping experiments are carried out under to possible modes: passive or active. In the passive mode, the optical trap is kept fixed at different positions to record the thermally activated unfolding and folding transitions from the force signal. For a given trap position, i.e., fixed dis-

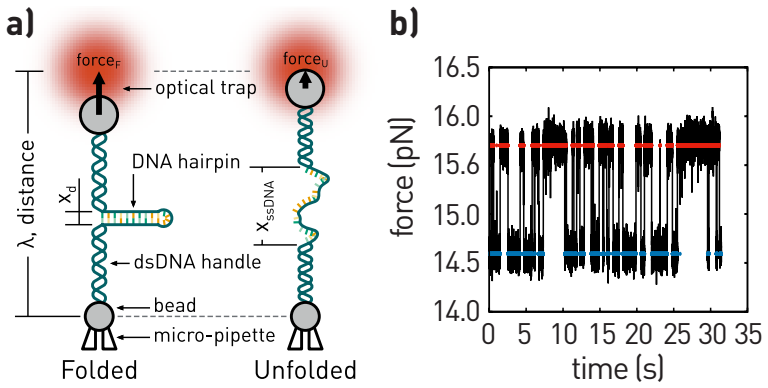

Figure 2.9.: **Passive equilibrium hopping experiments.** (a) Scheme of the molecular construct tethered at a given trap position when the hairpin is folded (left) and unfolded (right). (b) Force-time trace recorded at fixed trap position.

tance between the center of the optical trap and the fixed end, molecules with only two states (folded and unfolded) present two force levels. Each force level is proportional to the elasticity of the elements of the system formed by the trapped bead and the molecular construction. The bead is displaced concerning the center of the optical trap exerting an external force to the molecular system. Relative to the unfolded case, when the molecule is folded, i.e., short molecular extension, the AD bead is more displaced, resulting in a higher force (Fig. 2.9a). Recorded time traces typically span a few minutes at each trap position, which allows us to characterize the force dependence of such transitions (Fig. 2.9b).

In contrast, in active hopping experiments, a force feedback protocol is activated to keep controlled the position of the bead relative to the center of the optical trap (Fig. 2.10). In these experiments, the trap position signal is recorded as a function of time to measure the thermally activated transition between the different states involved in the folding/unfolding pathway. A force feedback protocol implemented in these sort of experiments has been studied in references [24, 25].

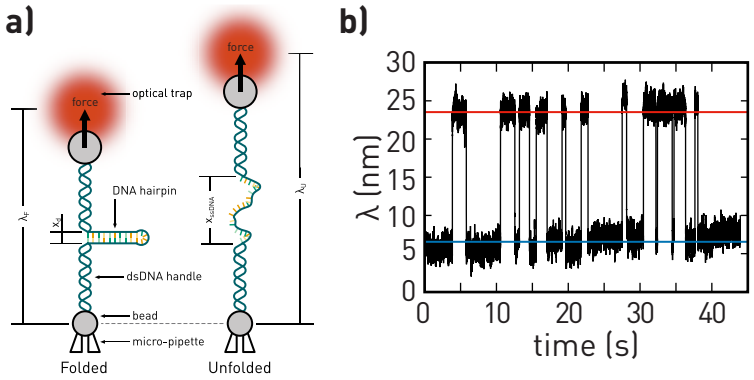

Figure 2.10.: **Active equilibrium hopping experiments.** (a) Scheme of the molecular construct tethered at a given force when the hairpin is folded (left) and unfolded (right). (b) Trap position-time trace recorded at fixed force.

Pulling experiments

In pulling experiments, the optical trap is repeatedly moved up and down, increasing and reducing the distance between the molecular construct's ends. The movement is done at a given constant speed. Force is ramped between an initial force, where the molecule is folded, and a maximum force value, where the molecule is unfolded (Fig. 2.11a). For small DNA hairpins or the studied protein (barnase), during stretching (red curves in Fig. 2.11b), the unfolding event is observed as a sudden force rip in the force versus trap position λ (distance) curve (FDC). Moreover, thanks to the folding reversibility, the studied molecules (barnase and small DNA hairpins) are mechanically refolded when distance is reduced, i.e., when the force is released (Fig. 2.11b). Upon force release (blue curves in Fig. 2.11b), a folding transition is detected as a force jump at forces below the unfolding ones. Notice that the force-distance curves shown in Fig. 2.11b exhibit two different force branches, highlighted with dashed lines, associated with the elastic response of the

system when the hairpin is in its folded and unfolded conformation (see Fig. 2.10a).

In contrast, in pulling experiments using long DNA hairpins, the FDCs exhibit a saw-tooth force pattern. In Figure 2.11c, are shown one illustrative unfolding and folding trajectory. Each force rip in the FDCs corresponds to the release of several base pairs, typically between 20 and 100bp. For this reason, these experiments are commonly known as unzipping experiments.

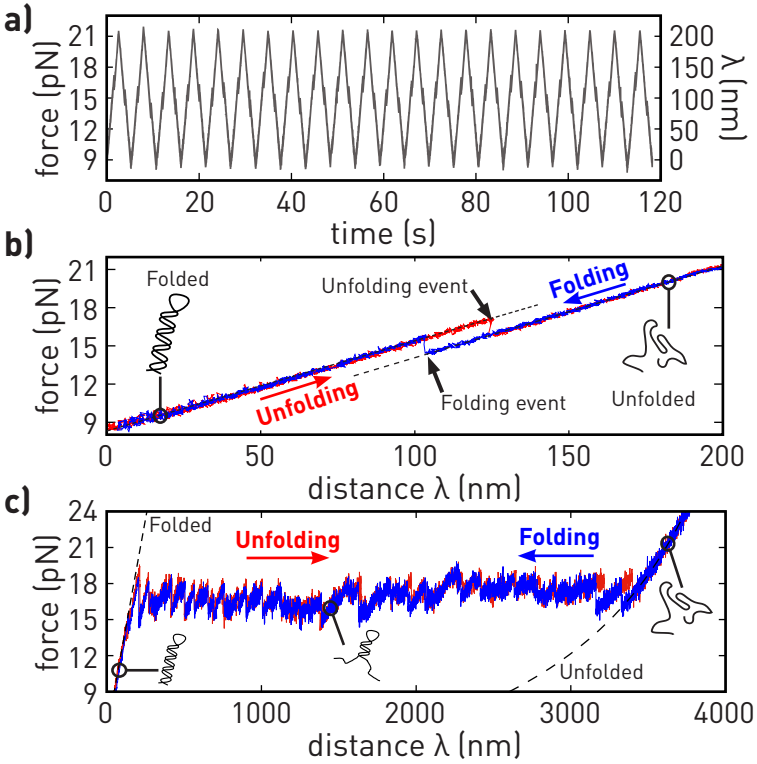

Figure 2.11.: **Non-equilibrium pulling experiments.** (a) Force and trap position signal as a function of time. (b) Unfolding (red) and folding (blue) force-distance curves recorded at room temperature using an small DNA hairpin. (c) Force-distance curve measured during the unzipping process of a long DNA hairpin.

Part II

FEEDBACK

3

INFORMATION FEEDBACK

MOTIVATION

In 1867, James Clerk Maxwell proposed a thought experiment to violate the second principle of thermodynamics [26, 27]. In such an experiment, an entity known as the Maxwell demon separates the fast and slow gas particles inside a box into two different reservoirs. This process is made by just measuring the speed of the molecules and opening and closing a gate to separate them. Both measurement and gate manipulation are done without energy cost. Therefore, a temperature gradient between the reservoirs is produced. In 1929, Leó Szilárd rethought the experiment by Maxwell to demonstrate that the knowledge of some experimental information has thermodynamic implications. The experiment proposed by Szilárd is known as the Szilard engine and has boosted a new field of research, namely the thermodynamics of small systems under feedback control or feedback thermodynamics. Szilárd engine has been experimentally realized in colloidal systems [28–31], electronic [32, 33] and optical devices [34–36], single molecules [37] and quantum systems [38–40].

In this chapter we address a new application of feedback thermodynamics. The goal is to apply feedback to reduce dissipation in small non-equilibrium systems. Most macroscopic and microscopic systems in daily life applications dissipate and require feedback to reduce dissi-

pation. Related to dissipation reduction is free energy determination, a relevant question for molecular thermodynamics. The understanding of dissipation reduction is vital in the nanoscale where dissipative molecular processes are remarkably efficient (for instance, the typical $> 90\%$ efficiency of the ATPase synthase). To address this interesting problem, in this chapter we derive and test a fluctuation theorem describing dissipation reduction in such kinds of feedback protocol. Moreover, we introduce the concept of feedback strategy to optimize free energy determination.

3.1 INTRODUCTION AND HISTORICAL CONTEXT

Maxwell's demon (MD) thought experiment [26, 27] has led to the insight that information enhances the capacity to extract energy from a system. In 1961 Landauer demonstrated that any irreversible logical operation (such as bit erasure) dissipates at least $k_B T \log 2$ per stored bit of information [41]. Bennett [42] applied this result to the Szilard's engine – a single particle version of the MD that extracts energy from a thermal bath – and showed that bit erasure is an entropy producing step necessary to restore the initial *blank* state of the memory of the demon. The extension of stochastic thermodynamics [43] to include information and feedback has led to novel fluctuation theorems (FTs) for work and information [44, 45]. Generalized Jarzynski equalities have been derived for isothermal feedback processes, where an external agent performs a single measurement on a system and applies a protocol ω_m that depends on the measurement outcome $m = 1, 2, \ldots, M$. A main equality useful for measurements with feedback reads [44],

$$\langle \exp\left[-(W - \Delta G)/k_B T\right] \rangle = \sum_{m=1}^{M} P_{\leftarrow}(m|\omega_m) \equiv \gamma \tag{3.1}$$

with k_B the Boltzmann constant and T the temperature. Here, W is the work performed on the thermodynamic system, ΔG is the free energy difference and $P_{\leftarrow}(m|\omega_m)$ is the probability to measure m along

the time reversal ($\leftarrow$) of the original protocol ω_m. Equation (3.1) defines the efficacy parameter $\gamma(\leq M)$, which quantifies the reversibility of the feedback process [46], reaching its maximum value (M) for reversible feedback processes where $P_{\leftarrow}(m|\omega_m) = 1$ for all m. Without feedback $\omega_m \equiv \omega$ and $\gamma = 1$, with Eq.(3.1) the Jarzynski equality. It is convenient to define the logarithm of the efficacy $\mathrm{Y} = \log \gamma$, which is a bound of the work that can be extracted in an isothermal feedback process. Y might be called thermodynamic information or information utilization.

For discrete-time single measurements Y is bounded from above, $-\infty < \mathrm{Y} \leq \log M$ ($M = 2$ being the one-bit Landauer limit). Jensen's inequality applied to Eq.(3.1) yields,

$$\langle W_d \rangle \equiv \langle W \rangle - \Delta G \geq -k_B T \mathrm{Y} \Rightarrow \langle W_d \rangle + k_B T \mathrm{Y} \geq 0 \tag{3.2}$$

where $\langle W_d \rangle$ is the dissipated work. Without feedback, $\mathrm{Y} = 0$ and $\langle W_d \rangle_0 \geq 0$ where the subscript 0 denotes the non-feedback case. In previous experimental realizations of the MD, feedback measurement is operated in equilibrium systems where $\mathrm{Y} = \log M$ is the maximum extractable work for equally likely outcomes, $\langle W \rangle = -k_B T \mathrm{Y} = -k_B T \log M$.

For non-equilibrium processes, Eq.(3.2) shows that the dissipation is bounded by $-k_B T \mathrm{Y}$. For $\mathrm{Y} > 0$, which correspond to the information-to-work conversion, dissipation is reduced by at most $-k_B T \mathrm{Y}$. Conversely, one could apply feedback protocols where $\mathrm{Y} < 0$ (information-to-heat conversion) and dissipation increased by at least $-k_B T \mathrm{Y}$. The latter case is a non-productive feedback for dissipation reduction. This kind of feedback has similarities with feedback in control theory, where protocols regulate experimental variables, e.g., by keeping them constant. These types of protocols counteract deviations from a system's specific preset conditions rather than rectifying thermal fluctuations, leading to increased dissipation ($\mathrm{Y} < 0$).

For the case $\mathrm{Y} > 0$, the dissipated work is reduced with respect to the non-feedback case, $\langle W_d \rangle \leq \langle W_d \rangle_0$. We define the information-to-

energy efficiency, η_I, as the amount of dissipation reduction, $\Delta\langle W_d\rangle = \langle W_d\rangle_0 - \langle W_d\rangle$, relative to the second law bound, $\langle W_d\rangle_0 + k_B T\Upsilon$:

$$\eta_I = \frac{\Delta\langle W_d\rangle}{\langle W_d\rangle_0 + k_B T\Upsilon} < 1. \tag{3.3}$$

This definition generalises the information-to-work conversion efficacy for non-equilibrium systems. Notice, that for cyclic and reversible MD devices $\langle W_d\rangle_0 = 0$ so $\eta_I = -\langle W_d\rangle / k_B T\Upsilon$, which is the standard MD efficiency.

As previously mentioned in the motivation to this chapter, related to dissipation reduction is free energy determination in molecular systems. It is an open question whether, by reducing dissipation, feedback can improve free energy determination, what we call information-to-measurement conversion. Free energy determination can be improved if $\langle W_d\rangle + k_B T\Upsilon < \langle W_d\rangle_0$, which implies that $\langle W_d\rangle + k_B T\Upsilon$ decreases approaching the lower bound (equal to 0, Eq. (3.2)) and weakening of the second law with feedback. Free energy determination improvement is related to the Jarzynski relation Eq.(3.1) and its bias, B_N, for N number of work (W) measurements. Inserting $\Upsilon = \log\gamma$ in Eq.(3.1) we define,

$$\begin{aligned} B_N &= \langle \Delta G_N\rangle - \Delta G \\ \Delta G_N &= -k_B T \log\left(\frac{1}{N}\sum_{i=1}^{N} e^{-\frac{W_i}{k_B T}}\right) + k_B T\Upsilon \end{aligned} \tag{3.4}$$

with ΔG_N the Jarzynski ΔG-estimator for N experiments and $\langle ...\rangle$ the average over many realizations of the N experiments. The average of the exponential of (with minus sign) work in Eq.(3.4) is biased for finite N, whereas the bounded and finite sum defining $\Upsilon(=\log\gamma)$ in Eq.(3.1) is not. B_N is positive and monotonically decreasing with N [47], vanishing in the limit $N \to \infty$. Therefore, improved free energy determination requires that B_N for $N = 1$ decreases with feedback relative to the non-feedback case. From Eq.(3.4) we have,

$$\begin{aligned} B_1 = \langle \Delta G_1\rangle - \Delta G = \langle W\rangle - \Delta G + k_B T\Upsilon = \\ = \langle W_d\rangle + k_B T\Upsilon \geq 0 \end{aligned} \tag{3.5}$$

which is equivalent to Eq.(3.2). Therefore, weakening of the second law implies reducing B_1, leading to improved free energy determination for finite N.

To quantify information-to-measurement conversion we define the information-to-measurement efficiency η_M as the relative difference between the second law inequality bounds with feedback, $\langle W_d \rangle + k_B T \mathrm{Y} \geq 0$, and without feedback, $\langle W_d \rangle_0 \geq 0$:

$$\eta_M = 1 - \frac{\langle W_d \rangle + k_B T \mathrm{Y}}{\langle W_d \rangle_0} = \eta_I + (\eta_I - 1) \frac{k_B T \mathrm{Y}}{\langle W_d \rangle_0}. \tag{3.6}$$

This is our second definition of efficiencies in non-equilibrium conditions. Equation (3.6) leads to a new inequality, $\eta_M \leq \eta_I$. Note that $\eta_M = 1$ if and only if $\eta_I = 1$, in which case dissipation reduction is maximal, $\langle W_d \rangle = -k_B T \mathrm{Y}$, and $\Delta G = \langle W \rangle + k_B T \mathrm{Y}$ can be determined with certainty. Improved free energy determination requires $\eta_M > 0$, whereas for $\eta_M \leq 0$ no gain in free energy determination is obtained with feedback: $\langle W_d \rangle$ decreases with respect to $\langle W_d \rangle_0$ by exactly or less than $k_B T \mathrm{Y}$. In general, optimal free energy determination is obtained by maximizing η_M.

We address information-to-energy and information-to-measurement conversion by combining theory and experiment. We introduce a new feedback-FT for multiple repeated measurements that is applicable to DNA unzipping experiments with optical tweezers. From the pulling experiments (see Chap. 2, Sec. 2.3) and using a short (20 base-pairs) DNA tetra-loop hairpin named L4 (Fig. 3.1), we measure the work distributions and Y to extract the efficiencies η_I, η_M. We investigate whether reduced dissipation with feedback ($\mathrm{Y} > 0$) improves free energy determination. We apply two different feedback protocols, namely discrete-time feedback (DTF), and continuous-time feedback (CTF). DTF and CTF being particular cases of the new feedback protocol.

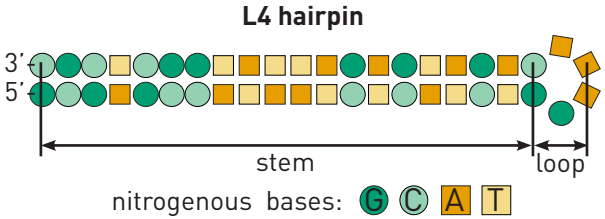

Figure 3.1.: **DNA sequence.** The DNA hairpin is formed a 20 base-pairs stem ended with a GAAA tetra-loop.

3.2 RESULTS

To address DTF and CTF in full generality we introduce the 1stTime-Feedback (1stTF) protocol, a repeated-time feedback protocol suitable for pulling experiments (Fig.3.2) that interpolates between DTF and CTF. In the 1stTF protocol the trap-position range $[\lambda_{min}, \lambda_{max}]$ is split into discrete $M+1$ steps, $\{\lambda_k; 0 \leq k \leq M\}$ with boundaries $\lambda_0 = \lambda_{min}; \lambda_M = \lambda_{max}$. The feedback protocol needs at least one intermediate measure position, so $M \geq 2$.

The 1stTF protocol works as follows; The initially folded (F) molecule is pulled at a loading rate r_F and measurements are taken at every discrete position λ_k along the forward ($\rightarrow$) process (Fig.3.2). The state of the molecule, $\sigma_k (= F, U)$, is monitored at every λ_k until a λ_{k^*} is reached where the molecule is observed to be in U for the first time, i.e., $\{\sigma_{k^*} = U;\ \sigma_k = F\ ,\ 0 \leq k < k^*\}$. At λ_{k^*} the loading rate is changed to r_U (Fig.3.2, top). Like in generic isothermal feedback processes, the conditional reverse ($\leftarrow$) process is the time-reverse of $\rightarrow$ [48]: the unloading rate is equal to r_U from λ_{max} to λ_{k^*}, after which the unloading rate is changed back to r_F between λ_{k^*} and λ_{min} (Fig.3.2, bottom). Therefore, no feedback is implemented on $\leftarrow$.

Let $p^{\sigma}_{\rightarrow,k}(r)$ and $p^{\sigma}_{\leftarrow,k}(r)$ be the probability to observe the molecule in state σ (F or U) at λ_k along $\rightarrow$ and $\leftarrow$, respectively, at the pulling rate r.

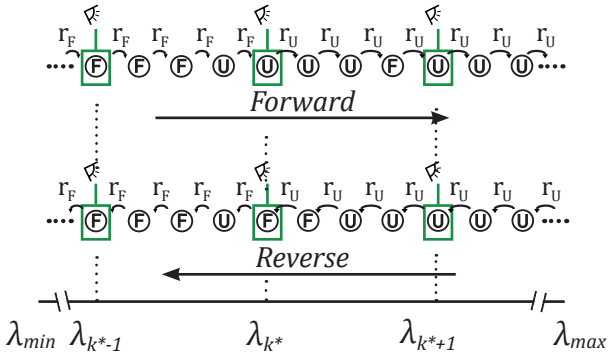

Figure 3.2.: **Schematics of the general 1^{st}time-feedback protocol in a pulling experiment.** Note that the molecule can execute multiple transitions at intermediate values of λ (between λ_{k^*-1} and λ_{k^*}) where no observations are made. A feedback response is only triggered when state U is observed for the first time at the specific positions defined by the predetermined set $\{\lambda_k\}$.

We derived a detailed and full work-FT for 1^{st}TF (Appendix B). The full work-FT reads,

$$\frac{\rho_{\rightarrow}(W)}{\rho_{\leftarrow}(-W)} = \exp\left[\frac{W - \Delta G_{FU} + k_B T \Upsilon_M}{k_B T}\right]$$
$$\text{with } \Upsilon_M = \log\left(\sum_{k=1}^{M} \frac{p^U_{\leftarrow,k}(r_U)}{p^U_{\leftarrow,k}(r_F)} \widetilde{\psi}_k\right). \tag{3.7}$$

W is the work measured as the area under the FDC between λ_{min} and λ_{max}, and ΔG_{FU} is the free energy difference between the state U at λ_{max} and the state F at λ_{min}, i.e., $\Delta G_{FU} = G_U(\lambda_{max}) - G_F(\lambda_{min})$. Work distributions are given by,

$$\rho_{\rightarrow}(W) = \sum_{k=1}^{M} \rho_{\rightarrow}(W|k)\,\psi_k \tag{3.8}$$

$$\rho_{\leftarrow}(-W) = e^{-\Upsilon_M} \sum_{k=1}^{M} \rho_{\leftarrow}(-W|k) \frac{p^U_{\leftarrow,k}(r_U)}{p^U_{\leftarrow,k}(r_F)} \widetilde{\psi}_k. \tag{3.9}$$

$\rho_{\rightarrow}(W|k)$ and $\rho_{\leftarrow}(-W|k)$ are the partial work distributions along $\rightarrow$ and $\leftarrow$ conditioned to those $\rightarrow$ paths where U is observed for the first time at λ_k. ψ_k and $\widetilde{\psi}_k$ $(1 \leq k \leq M)$ are the probabilities along $\rightarrow$ and $\leftarrow$ to observe U at λ_k for the first and last time at the single rate r_F. Note that while ψ_k can be measured from $\rightarrow$ pulls with the feedback on, the *reverse* quantities $\widetilde{\psi}_k$, $p^U_{\leftarrow,k}(r_F)$, and $p^U_{\leftarrow,k}(r_U)$ are measured from reverse pulls at either the unloading rates r_F or r_U without feedback. Υ_M in Eq.(3.7) denotes the thermodynamic information,

$$\Upsilon_M = \log\left(\sum_{k=1}^{M} \psi_k \exp J_k\right)$$
$$\text{with } J_k = \log\left(\frac{p^U_{\leftarrow,k}(r_U)\,\widetilde{\psi}_k}{p^U_{\leftarrow,k}(r_F)\,\psi_k}\right), \tag{3.10}$$

where J_k is the *partial* thermodynamic information along $\rightarrow$, restricted to those paths where U is observed for the first time at λ_k. Equation (3.10) expresses Υ_M as a partition sum or potential of mean force over the partial contributions J_k. In the feedback case where $r_U = r_F$, $J_k = \log(\widetilde{\psi}_k/\psi_k)$ and $\Upsilon_M = \log\left(\sum_{k=1}^{M}\widetilde{\psi}_k\right) = 0$ and Crooks FT [49] is recovered.

The DTF and CTF feedback protocols are particular cases of the 1$^{\text{st}}$TF protocol: DTF corresponds to $M = 2$; whereas $M \to \infty$ yields CTF. Let us assume that the time between consecutive measurements at λ_k is equal to τ, in which case $M \to \infty$ corresponds to $\tau \to 0$. To implement CTF experimentally we take the lowest possible value of τ as dictated by the maximum data acquisition frequency of the instrument ($\tau \approx 1$ms).

3.2.1 *Discrete-time feedback*

In DTF, the pulling rate along $\rightarrow$ is changed from r_F to r_U at a given trap position λ_1 if $\sigma = U$, otherwise it remains unchanged (Fig 3.3). Therefore, if $\sigma = F$ at λ_1 the pulling rate is constant and equal to r_F throughout the pulling cycle. Whereas if $\sigma = U$ at λ_1, the pulling rate

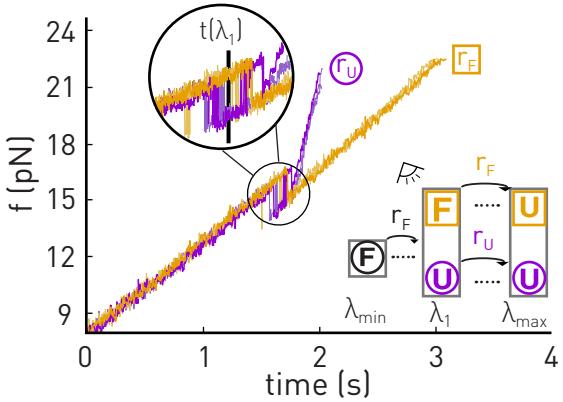

Figure 3.3.: **Discrete-Time Feedback.** Experimental force-time unfolding curves. The molecule is pulled at $r_F = 4$pN/s and the observation is made at λ_1: if the molecule is folded ($\sigma = F$) the pulling rate remains unchanged ($r = r_F$, orange curves); if it is unfolded ($\sigma = U$) the pulling rate is changed to $r_U = 17$pN/s, making refolding events less likely (purple curves). The schematic summarizes the feedback protocol and the reaction to the measurement outcome.

along $\leftarrow$ starts at $r = r_U$ and switches back to r_F at λ_1. A detailed feedback-FT is derived either from the detailed form of Eq.(3.7) for $M = 2$ or from the extended fluctuation relation [50] (mathematical details are provide in Appendix C)

$$\frac{\rho_{\rightarrow}(W|\sigma)}{\rho_{\leftarrow}(-W|\sigma, r_\sigma)} = \exp\left[\frac{W - \Delta G_{FU} + k_B T\ I_\sigma}{k_B T}\right]$$
$$\text{with } I_\sigma \equiv \log\left(\frac{p^\sigma_{\leftarrow}(r_\sigma)}{p^\sigma_{\rightarrow}(r_F)}\right);\ \sigma = \{F, U\} \tag{3.11}$$

where $\sigma\,(= F, U)$ is the measurement outcome at λ_1 along $\rightarrow$. $\rho_{\rightarrow}(W|\sigma)$ and $\rho_{\leftarrow}(W|\sigma, r_\sigma)$ are the normalized work distributions conditioned to those trajectories passing through σ at λ_1 along $\rightarrow$ and $\leftarrow$, respectively. $p^\sigma_{\rightarrow}(r_F)$, $p^\sigma_{\leftarrow}(r_\sigma)$ are the probabilities to measure σ at λ_1 along $\rightarrow$ and $\leftarrow$, at the respective pulling rates. Finally, I_σ is the partial thermodynamic information of measurement outcome σ. Note that I_σ can take any sign depending on the ratio $p^\sigma_{\leftarrow}(r_\sigma)/p^\sigma_{\rightarrow}(r_F)$, which can be larger

or smaller than 1. Moreover, while all trajectories in $\rightarrow$ are classified in one of the two groups $\sigma = (F, U)$, only those that revisit again the same σ contribute to $\rho_{\leftarrow}(-W|\sigma, r_\sigma)$. Therefore, the normalization condition along $\rightarrow$, $\sum_\sigma p^\sigma_{\rightarrow}(r_F) = 1$, is not applicable to $\leftarrow$, i.e., $\sum_\sigma p^\sigma_{\leftarrow}(r_\sigma) \neq 1$.

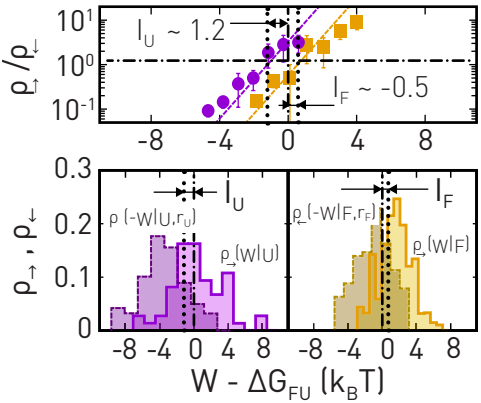

Figure 3.4.: **Verification of the detailed work-FT.** Bottom: $\rho_{\rightarrow}(W|\sigma)$ (solid lines) and $\rho_{\leftarrow}(-W|\sigma, r_\sigma)$ (dashed lines) for $\sigma = F, U$ (orange, purple) trajectories. Top: Test of the detailed feedback-FT Eq.(3.11) for $\sigma = F, U$.

We apply DTF with $r_U > r_F$ to extract partial work distributions $\rho_{\rightarrow}(W|\sigma)$ and $\rho_{\leftarrow}(-W|\sigma, r_\sigma)$ by classifying trajectories depending on the outcome σ at λ_1, and the protocol under which they are operated. Figure 3.4 shows the partial work distributions for $\sigma = F, U$ for $r_F = 4$ pN/s, $r_U = 17$ pN/s. For the F (U) subset we find that the work distributions are shifted rightwards (leftwards) with respect to the non-feedback case with crossing points (W^*_σ) such that $W^*_F > \Delta G_{FU}$ ($W^*_U < \Delta G_{FU}$). These shifts reflect the fact that hairpin unfolding is on average more (less) energy-costly for the F (U) subset than without feedback at the pulling rate r_F. From Eq.(3.11) the measured shift, defined as $W^*_\sigma - \Delta G_{FU}$, equals $-k_B T I_\sigma$. The $\rho_{\rightarrow}(W|\sigma)$ and $\rho_{\leftarrow}(-W|\sigma, r_\sigma)$ fulfil Eq.(3.11) crossing at values $W^*_\sigma = \Delta G_{FU} - k_B T I_\sigma$ with $I_F \approx -0.5$ and $I_U \approx 1.2$ (Fig.3.4, top).

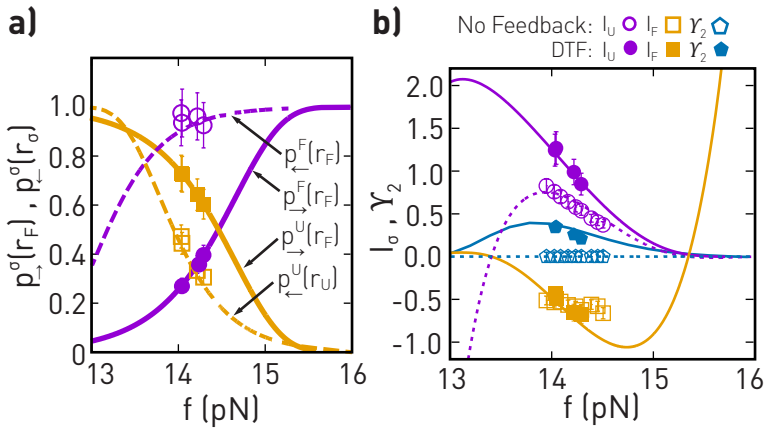

Figure 3.5.: **Thermodynamic information for DTF.** (a) Probabilities $p^{\sigma}_{\rightarrow}(r_F)$, $p^{\sigma}_{\leftarrow}(r_\sigma)$ and (b) information terms I_σ and Υ_2 as a function of the force in U measured at λ_1. Orange (purple) data are for $\sigma = F(U)$. In (b) we also show I_σ without feedback (empty symbols). Theoretical predictions from the Bell-Evans model are shown as solid and dashed lines.

Figure 3.5 shows $p^{\sigma}_{\rightarrow}(r_F)$, $p^{\sigma}_{\leftarrow}(r_\sigma)$, $I_{F,U}$, and Υ_2 versus the force in U at λ_1. We choose force as a reference value to present the results because it is more informative than the trap position λ, the relative distance between the trap position and an arbitrary initial position in the light-lever detector. In addition, the force dependence of the experimental values is adjusted to the Bell-Evans model. The Bell-Evans model is a kinetic model commonly used to characterize the folding-molecular free energy landscape. This model simplifies the landscape with a kinetic barrier between the different states. This model will be exploited in future chapters of this thesis. Here, the Bell-Evans model is used to model the force-dependent $I_{F(U)}$ and thermodynamic information.

Combining the detailed feedback-FT Eq.(3.11) for $\sigma = F, U$ yields the full work-FT Eq.(3.7) for $M = 2$ (mathematical details are provided in Appendix C),

$$\frac{\rho_{\rightarrow}(W)}{\rho_{\leftarrow}(-W)} = \exp\left[\frac{W - \Delta G_{FU} + k_B T \Upsilon_2}{k_B T}\right] \tag{3.12a}$$

$$\text{with } \Upsilon_2 = \log\left(\sum_{\sigma=F,U} p^{\sigma}_{\leftarrow}(r_\sigma)\right) =$$

$$= \log\left(\sum_{\sigma=F,U} p^{\sigma}_{\rightarrow}(r_F)\exp I_\sigma\right), \tag{3.12b}$$

Υ_2 being the thermodynamic information. The forward and reverse work distributions are given by,

$$\rho_{\rightarrow}(W) = \sum_{\sigma} p^{\sigma}_{\rightarrow}(r_F)\,\rho_{\rightarrow}(W|\sigma) \tag{3.13}$$

$$\rho_{\leftarrow}(-W) = e^{-\Upsilon_2}\sum_{\sigma} p^{\sigma}_{\leftarrow}(r_\sigma)\,\rho_{\leftarrow}(-W|\sigma, r_\sigma) \tag{3.14}$$

For $r_F = r_U$ we have $p^{F}_{\leftarrow}(r) + p^{U}_{\leftarrow}(r) = 1$ yielding $\Upsilon_2 = 0$ and Crooks FT [49] as expected. Figure 3.6a tests Eq.(3.12a) and Fig. 3.6b shows Υ_2 and the efficiencies η_I, η_M obtained from the experiments. Results are compared with numerical simulations of the DNA pulling experiments (details in Appendix D) and a prediction by the Bell-Evans model. For $r_U > r_F$, $p^{F}_{\leftarrow}(r_F) + p^{U}_{\leftarrow}(r_U) > 1$ and $\Upsilon_2 > 0$. In general, $-\infty < \Upsilon_2 \leq \log 2$ $(0 \leq p^{\sigma}_{\leftarrow}(r_\sigma) \leq 1)$ showing that the Landauer bound holds for two-state molecules pulled under DTF. Saturating the bound, $\Upsilon_2 = \log 2$, requires full reversibility [46], i.e., $p^{\sigma}_{\leftarrow}(r_\sigma) = 1$, which is obtained for arbitrary r_F in the limit $r_U \to \infty$ ($p^{U}_{\leftarrow}(r_U) = 1$) and $\lambda_1 \to \lambda_{min}$ (i.e., maximally stable F or $p^{F}_{\leftarrow}(r_F) = 1$).

We find $\eta_M \sim -0.05$ and $\eta_I \sim 0.04$ showing that information-to-measurement conversion is much less efficient than information-to-work conversion. Moreover, $\eta_M < 0$ throughout the whole force range shows that DTF does not improve free energy prediction.

To better understand this result we have calculated the efficiencies η_I and η_M for DTF in the two-state Bell-Evans model, where force is the

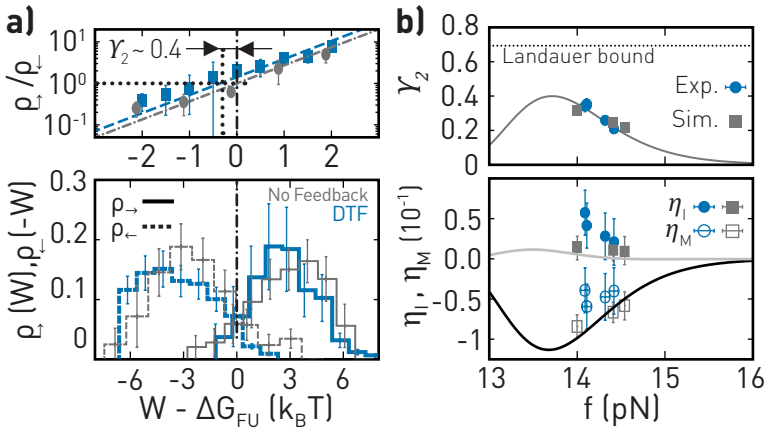

Figure 3.6.: **Efficiencies for DTF.** (a) Top: Test of the full feedback-FT Eq.(3.12a) with shift Υ_2 (blue) with respect to the case without Feedback (gray). Bottom: Work distributions. (b) Thermodynamic information, Υ_2 (top), and efficiencies η_I and η_M (bottom) measured at different λ_1 (grey: simulations; blue: experiments). Solid lines are the Bell-Evans model in the single-hopping approximation (Sec. 3.2.2). In general, Υ_2 is bounded from above by the Landauer limit for binary measurements ($\log 2$, dashed line). Notice that $\eta_M < \eta_I$, as expected.

control parameter, using the single-hopping approximation (mathematical details in Sec. 3.2.2). The single-hopping approximation neglects multiple transitions after the measurement position λ_1. Therefore, the analytical results derived from this approximation are lower bounds to the true efficiencies. Figure 3.6b shows that the analytical results capture the trend of the experimental data but systematically underestimate the measured efficiencies η_I and η_M. The fact that $\eta_M < 0$ throughout the force range shows that although DTF does reduce dissipation it does not improve free energy prediction. This conclusion is supported by the results shown in Fig.3.7. Figure 3.7 shows the experimental free energy bias (Eq.(3.4)) as a function of N number of pulling experiments at the

experimental conditions shown in Fig. 3.6: bias with feedback does not decrease with respect to the non-feedback case.

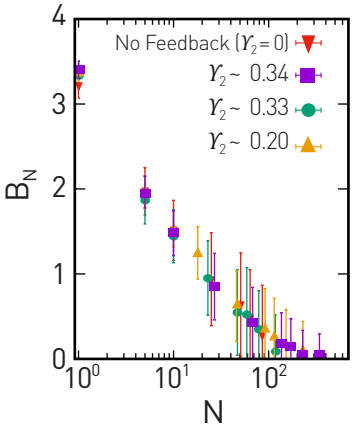

Figure 3.7.: **Bias for three different DTF conditions compared to the non-feedback case.** There is no improvement in free energy prediction: B_N is equal for all the explored cases compared to the case without feedback.

3.2.2 *The Bell-Evans model in the single-hopping approximation*

To better understand under which conditions η_I and η_M are optimal under DTF protocols, we have carried out an analysis based on the two-state Bell-Evans model where force is the control parameter. To analyze dissipation reduction, we define $\Delta\langle W_d\rangle = \langle W_d\rangle_0 - \langle W_d\rangle > 0$, the change in the average dissipated work upon implementing feedback. It has been shown that the Bell-Evans approximation where force is controlled provides qualitatively identical and quantitatively comparable results to the experimental condition where the trap position is controlled [51]. In Figure 3.8, we show a typical trajectory (state versus force) in the DTF protocol in the model where the initially folded molecule is pulled at $\dot{f} = r_F$. The pulling rate is then changed to $\dot{f} = r_U$ at a given force value (f_1) if the molecule is in U. $\Delta\langle W_d\rangle$ is only determined by the contribution

of those trajectories that are in U at f_1: trajectories that are in F at f_1 do not change the pulling rate and therefore do not contribute to $\Delta\langle W_d\rangle$.

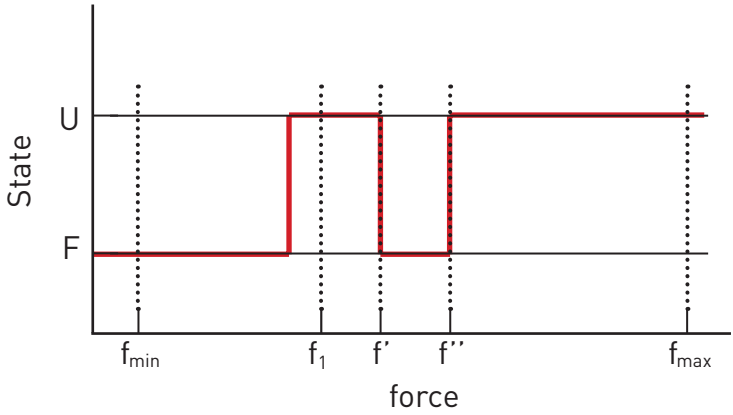

Figure 3.8.: **Single-hopping approximation.** In the single-hopping approximation the molecule refolds and unfolds again before f_{max} after the measurement position (f_1). The forces that determine the trajectories $U \rightarrow F \rightarrow U$ between f_1 and f_{max} are f' and f'', which are the folding and unfolding forces, respectively.

To determine $\Delta\langle W_d\rangle$, we restrict the analysis to single-hopping trajectories of the type $U \rightarrow F \rightarrow U$ after f_1. The average dissipated work in the range (f_1, f_{max}) for the U-type trajectories is given by:

$$\langle W_d\rangle_{U\rightarrow F\rightarrow U} = P_{U\rightarrow F\rightarrow U}(f_1)\langle f'' - f'\rangle x_m \tag{3.15}$$

where x_m stands for the difference in molecular extension between U and F; f' and f'' are the folding and unfolding forces of steps $U \rightarrow F$ and $F \rightarrow U$ for the trajectory $U \rightarrow F \rightarrow U$ (Fig.3.8); $P_{U\rightarrow F\rightarrow U}(f_1)$ is the fraction of trajectories of the type $U \rightarrow F \rightarrow U$, which in the current

single-hopping approximation equals $1 - P_s^U(f_1, f_{max})$, where $P_s^U(f_1, f)$ is the survival probability of U between f_1 and f. $\langle f'' - f' \rangle$ is given by,

$$\langle f'' - f' \rangle = \\ - \int_{f_1}^{f_{max}} df' \frac{\partial P_s^U(f_1, f')}{\partial f'} \int_{f'}^{f_{max}} df'' P_s^F(f', f'') \tag{3.16}$$

where $P_s^F(f', f'')$ is the survival probability of F between f' and f''.

Moreover, $\Delta\langle W_d \rangle$ is proportional to the difference of the average dissipated work between f_1 and f_{max} calculated at the pulling rates r_U and r_F. Equations (3.15), (3.16) must be calculated at the pulling rates r_F and r_U to obtain the dissipation reduction in the single-hopping approximation,

$$\Delta\langle W_d \rangle = \\ p_{\rightarrow}^U(r_F) \left[\langle W_d \rangle_{U \rightarrow F \rightarrow U}(r_U) - \langle W_d \rangle_{U \rightarrow F \rightarrow U}(r_F) \right] \tag{3.17}$$

where $p_{\rightarrow}^U(r_F)$ is the fraction of trajectories observed at U and the dissipated work is restricted to the range (f_1, f_{max}). For practical purposes, we take $f_{max} \to \infty$ as the molecule always ends in U at f_{max}. Equations (3.15), (3.16), (3.17) are numerically calculated for generic Bell-Evans rates where survival probabilities have simple analytical expressions. For the specific case relevant to the experiments, the transition state is located at half distance between F and U, $x^\dagger = x_m/2$, we have

$$k_{F \rightarrow U}(f) = k_0 \cdot \exp\left(\beta f x^\dagger\right) \tag{3.18a}$$

$$k_{F \leftarrow U}(f) = k_0 \cdot \exp\left[\beta(\Delta G - f x^\dagger)\right] \tag{3.18b}$$

where k_0 is the kinetic rate at zero force, and ΔG is the free energy difference between F and U ($\beta = 1/k_B T$). Finally, for sufficiently high forces where $k_{F \leftarrow U}(f_1)/(\beta x^\dagger r) \ll 1$, the average dissipation reduction is obtained to first order in $1/r$:

$$\Delta\langle W_d \rangle = p_{\rightarrow}^U(r_F) \left[\frac{x_m k_{F \leftarrow U}(f_1)^2}{2 k_{F \rightarrow U}(f_1)(\beta x^\dagger)^2} \right] \left(\frac{1}{r_F} - \frac{1}{r_U} \right) \tag{3.19}$$

which is positive for $r_U > r_F$ as expected, and negative otherwise. As the molecule always starts in F at f_{min}, again, for practical purposes we take $f_{min} \to -\infty$. The survival probability of U along $\to$ ($p^U_{\to}(r_F)$) is expressed as:

$$p^U_{\to}(r_F) = 1 - \exp\left(-\frac{k_{F\to U}(f_1)}{\beta x^\dagger r_F}\right) \tag{3.20}$$

To calculate the efficiencies, we also need the values of $\langle W_d\rangle_0$ and k_BTY_2 for DTF, the latter being given by Eq.(3.12b). $\langle W_d\rangle_0$ is estimated from the mean first unfolding force $\langle f_{F\to U}\rangle$ in the Bell-Evans approximation,

$$\begin{aligned}\langle W_d\rangle_0 &= x_m\left(\langle f_{F\to U}\rangle - f_c\right) = \\ &= x_m\left(\frac{1}{\beta x^\dagger}\log\left(\frac{\beta x^\dagger r}{k_0}\right) - f_c\right)\end{aligned} \tag{3.21}$$

where $f_c = \Delta G/x_m$ is the coexistence force (force where U and F have the same probability of being observed). We also have,

$$k_BTY_2 = \log(p^F_{\leftarrow}(r_F) + p^U_{\leftarrow}(r_U)) \tag{3.22a}$$

$$p^U_{\leftarrow}(r) = \exp\left(-\frac{k_{F\leftarrow U}(f)}{\beta x^\dagger r}\right) \tag{3.22b}$$

and $p^F_{\leftarrow}(r) = 1 - p^U_{\leftarrow}(r)$. We have calculated these quantities for the parameters that best fit the experimental pulling curves for L4 without feedback: $k_0 = 2\cdot 10^{-14}\mathrm{s}^{-1}$, $x_m = 18\mathrm{nm}$, $x^\dagger = 9\mathrm{nm}$, $k_BT = 4.114\mathrm{pN}\cdot\mathrm{nm}$, $\beta = 1/k_BT \sim 0.24\mathrm{pN}^{-1}\cdot\mathrm{nm}^{-1}$, $\Delta G = 64.2\mathrm{k_BT}$, and $f_c = 14.66\mathrm{pN}$. From these values we calculate the efficiencies defined in Eqs.(3.3), (3.6),

$$\eta_I = \frac{\Delta\langle W_d\rangle}{\langle W_d\rangle_0 + k_BTY_2} \tag{3.23a}$$

$$\eta_M = \frac{\Delta\langle W_d\rangle - k_BTY_2}{\langle W_d\rangle_0} \tag{3.23b}$$

In Figure 3.9 we show η_I and η_M versus force for $r_F = 4\mathrm{pN/s}$ and $r_U = 17\mathrm{pN/s}$. The red continuous line corresponds to the results obtained from Eqs.(3.23a, 3.23b) by numerically solving Eqs. (3.15-3.17) to determine $\Delta\langle W_d\rangle$ in the single-hopping approximation. The dashed line shows the $1/r$ leading term, Eq.(3.19), which holds for sufficiently high

forces ($f > 14.45$pN where $k_{F\leftarrow U}(f)/(\beta x^{\dagger} r_F) \leq 0.3$). The circles and squares are the experimental and simulated results shown in Figure 3.6.

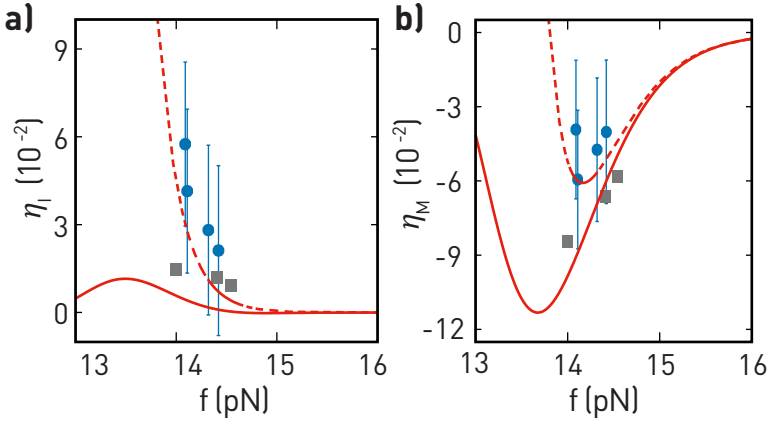

Figure 3.9.: **Efficiencies for the single-hopping approximation.** (a) Dissipation-reduction efficiency η_I (b) Information-to-measurement efficiency η_M. Red solid line corresponds to the exact solution of $\Delta\langle W_d\rangle$ derived from Eqs. (3.15-3.17), and the red dashed line corresponds to the approximation where $k_{F\leftarrow U}(f_1)/(\beta x^{\dagger} r) \ll 1$ (Eq.(3.19)).

3.2.3 *Continuous-time feedback*

CTF is the limit of $\Delta\lambda(=\lambda_{k+1}-\lambda_k)$, $\tau \to 0$ for the 1$^{\text{st}}$TF. The mathematical details are provide in Appendix B. The detailed feedback-FT reads,

$$\frac{\rho_{\rightarrow}(W|\lambda)}{\rho_{\leftarrow}(-W|\lambda)} = \exp\left[\frac{W-\Delta G_{FU}+k_BT\,J(\lambda)}{k_BT}\right]; \tag{3.24a}$$

$$J(\lambda) = \log\left(\frac{p^U_{\leftarrow}(\lambda,r_U)\,\widetilde{\psi}(\lambda)}{p^U_{\leftarrow}(\lambda,r_F)\,\psi(\lambda)}\right) \tag{3.24b}$$

while the full feedback-FT reads,

$$\frac{\rho_{\rightarrow}(W)}{\rho_{\leftarrow}(-W)} = \exp\left[\frac{W - \Delta G_{FU} + k_B T \Upsilon_\infty}{k_B T}\right]; \tag{3.25a}$$

$$\Upsilon_\infty = \log\left(\int_{\lambda_{min}}^{\lambda_{max}} \frac{p_{\leftarrow}^{U}(\lambda, r_U)}{p_{\leftarrow}^{U}(\lambda, r_F)} \widetilde{\psi}(\lambda) d\lambda\right) = \\ = \log\left(\int_{\lambda_{min}}^{\lambda_{max}} \psi(\lambda) \exp J(\lambda) d\lambda\right) \tag{3.25b}$$

with the forward and reverse work distributions given by,

$$\rho_{\rightarrow}(W) = \int_{\lambda_{min}}^{\lambda_{max}} \rho_{\rightarrow}(W|\lambda)\psi(\lambda)d\lambda \tag{3.26}$$

$$\rho_{\leftarrow}(-W) = e^{-\Upsilon_\infty} \int_{\lambda_{min}}^{\lambda_{max}} \rho_{\leftarrow}(-W|\lambda) \cdot \\ \frac{p_{\leftarrow}^{U}(\lambda, r_U)}{p_{\leftarrow}^{U}(\lambda, r_F)} \widetilde{\psi}(\lambda) d\lambda \tag{3.27}$$

where $\psi(\lambda)$ ($\widetilde{\psi}(\lambda)$) is the probability density to observe the first (last) unfolding (folding) event $F \rightarrow U$ ($F \leftarrow U$) along $\rightarrow$ ($\leftarrow$); $p_{\leftarrow}^{U}(\lambda, r)$ is the probability density of the molecule being in U at λ along $\leftarrow$ at the unloading rate r.

Similarly to the partial thermodynamic information I_σ in Eq.(3.11), if we define $I_\sigma^{rr'}(\lambda) = \log\left(p_{\leftarrow}^{\sigma}(\lambda, r')/p_{\rightarrow}^{\sigma}(\lambda, r)\right)$, we have $J(\lambda) = I_U^{r_F r_U}(\lambda) - I_U^{r_F r_F}(\lambda) + I_\psi(\lambda)$, with $I_\psi(\lambda) = \log\left(\widetilde{\psi}(\lambda)/\psi(\lambda)\right)$. Notice that for $r_U = r_F$, i.e., no feedback case, $\rho_{\leftarrow}(-W|\lambda) = \rho_{\leftarrow}(-W)$ and $\Upsilon_\infty = 0$, but $J(\lambda) = I_\psi(\lambda) \neq 0$.

Equations (3.24b) and (3.25b) can be further simplified by neglecting multiple hopping transitions between F and U. In contrast to the previous single-hopping approximation, in this mean field approximation (MFA), we are considering trajectories with only one transition. Neglecting multiple hopping transitions between F and U along $\rightarrow$ ($\leftarrow$) implies that, once the molecule jumps to U (F) at a given λ it remains in that state until reaching λ_{max} (λ_{min}). Therefore we identify $\psi(\lambda) = c \cdot p_{\rightarrow}^{F}(\lambda, r_F)\, p_{\rightarrow}^{U}(\lambda, r_F)$ and $\widetilde{\psi}(\lambda) = c' \cdot p_{\leftarrow}^{F}(\lambda, r_F)\, p_{\leftarrow}^{U}(\lambda, r_F)$ where ψ ($\widetilde{\psi}$) is the normalized probability along $\rightarrow$ ($\leftarrow$) to observe U at

a given value of λ for the first (last) time at the single loading (unloading) rate r_F, and $p^{\sigma}_{\rightarrow(\leftarrow)}(\lambda, r)$ is the fraction of trajectories observed to be in $\sigma(= F, U)$ at λ with loading rate r. The $p^{\sigma}_{\rightarrow(\leftarrow)}(\lambda, r)$ are derived using the Bell-Evans model (Eq. (3.20) and (3.22b)), and c, and c' are normalizing factors.

Terms $J(\lambda)$ and Υ_∞ in Eqs.(3.24b , 3.25b) are simplified as,

$$J_{MFA}(\lambda) = I_F(\lambda) + I_U(\lambda) + \log\left(\frac{c'}{c}\right) \tag{3.28a}$$

$$\Upsilon_{MFA} = \log\left(\int_{\lambda_{min}}^{\lambda_{max}} \psi(\lambda) e^{J_{MFA}(\lambda)} \; d\lambda\right) = \\ = \log\left(\frac{\int_{\lambda_{min}}^{\lambda_{max}} p^{F}_{\leftarrow}(\lambda, r_F)\, p^{U}_{\leftarrow}(\lambda, r_U) d\lambda}{\int_{\lambda_{min}}^{\lambda_{max}} p^{F}_{\leftarrow}(\lambda, r_F) p^{U}_{\leftarrow}(\lambda, r_F) d\lambda}\right) \tag{3.28b}$$

with $I_F(\lambda)$ and $I_U(\lambda)$ in Eq.(3.28a) the partial thermodynamic information as in Eq.(3.11) for DTF at a given λ. As mentioned before, the transition state of molecule L4 is located at half-distance of the molecular extension, resulting into nearly symmetric forward and reverse processes without feedback. Therefore, $c' \cong c$ and $\log(c'/c) \cong 0$. Equation (3.28a) shows $J_{MFA}(\lambda)$ in CTF equals the sum of the $I_\sigma(\lambda)$ corresponding to DTF. Υ_{MFA} in Eq.(3.28b) is a good approximation for Υ_∞ under highly irreversible pulling conditions where the molecule executes a single unfolding (folding) transition during $\rightarrow$ ($\leftarrow$).

We tested CTF in DNA hairpin pulling experiments (Fig.3.10). The molecule initially in F is pulled at $r_F = 5$pN/s and the state monitored by recording the force every $\tau = 1$ms until the first force jump is observed at a given trap position λ^*. The force rip indicates that state U has been visited for the first time at λ^*. Then the pulling rate is increased to $r_U = 23$pN/s until the preset maximum force is reached. For the reverse process the optical trap moves backwards at $r_U = 23$pN/s until λ^* and then the pulling rate switched back to $r_F = 5$pN/s. By repeatedly pulling, we collect enough statistics to test Eqs.(3.24a , 3.25a) and measure $J(\lambda)$ and Υ_∞. In Figure 3.11 (bottom), we plot $\rho_\rightarrow(W|\lambda)$, $\rho_\leftarrow(-W|\lambda)$ for three selected λ^*, while in the top panel of Fig. 3.11, we

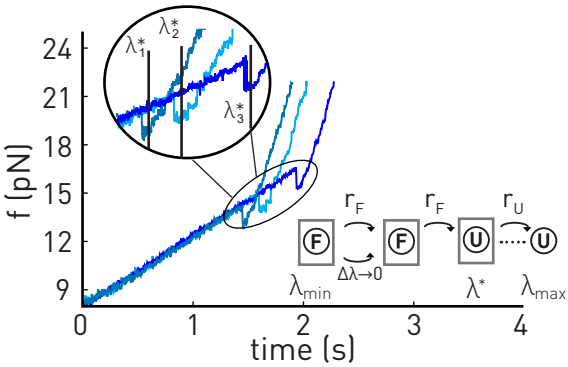

Figure 3.10.: **Continuous-time feedback.** Experimental unfolding curves obtained with $r_F = 5\text{pN/s}$ and $r_U = 23\text{pN/s}$.

test Eq.(3.24a). By determining the crossing work values between $\rho_{\rightarrow}$ and $\rho_{\leftarrow}$, $W^*(\lambda) = \Delta G_{FU} - k_B T J(\lambda)$, we extract $J(\lambda)$.

Figure 3.12a shows the values of $\widetilde{\psi}$ and $p^{UI}_{\leftarrow}$ directly determined from experimental FDCs for the two loading rates, $r_F = 5\text{pN/s}$ and $r_U = 23\text{pN/s}$ (symbols) as a function of force. This has been fitted to the Bell-Evans model (solid lines) to extract the kinetic parameters of hairpin L4, in order to compare them with the simulations and the MFA. Figure 3.12b shows the experimental values of $J(\lambda)$ determined from the detailed feedback-FT Eq.(3.24a) (filled squares) together with the predictions by the fits to the Bell-Evans model using Eq.(3.24b) (dashed line) and the MFA, Eq.(3.28a), assuming that $\log(c'/c) = 0$ (solid line).

In Figure 3.13a we test the full feedback-FT Eq.(3.25a). For comparison we also show the non-feedback case. We emphasize the importance of properly weighing $\rho_{\leftarrow}(-W|\lambda)$ to build $\rho_{\leftarrow}(-W)$. An *unweighted* reverse work distribution ($\int \rho_{\leftarrow}(-W|\lambda)\, d\lambda$, blue) does not fulfil the FT (inset, blue points), and the slope of the fitting line ($\sim$0.08) is far below 1. Figure 3.13b (top) shows Υ_∞ for different experimental conditions (black circles) and results obtained in simulations (gray squares) of a hairpin model compared to the theoretical values determined from Eq.(3.25b) using the Bell-Evans model, and the MFA (dashed line).

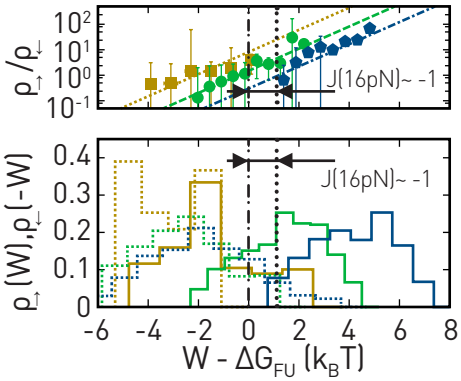

Figure 3.11.: **Experimental test of the detailed feedback-FT for CTF.** FDCs have been classified according to the value of λ at which the earliest unfolding event is detected. FDCs have been grouped in bins of equal width $\Delta\lambda \sim$ 6nm and Eq.(3.24a) applied to each interval. We show work histograms for three different λ values. The corresponding forces at the presented λ values are: 15pN (brown), 15.5pN (green) and 16pN (blue). The value of $J(\lambda)$ for $f = 16$pN is highlighted.

In Figure 3.13b (bottom), we show the efficiencies η_I and η_M versus r_U/r_F. As shown in Figure 3.13b dissipation reduction is larger for CTF as compared to DTF. Indeed, $\Delta\langle W_d\rangle$ is not bounded by the Landauer limit $k_B T \log 2$ in the CTF. However, $\eta_M \sim -0.1$ is slightly negative as in DTF, showing that dissipation reduction does not necessarily improve free energy determination. In Figure 3.14, we plot the experimental free energy bias Eq.(3.4) as a function of N experiments at the conditions shown in Fig. 3.13: Again, as for DTF, we observe that the bias with feedback does not decrease relative to the non-feedback case.

3.2.4 *Efficient information-to-measurement conversion: feedback strategies*

Here, we ask under which conditions feedback does improve free energy determination by increasing η_M. As previously shown, the dissipation

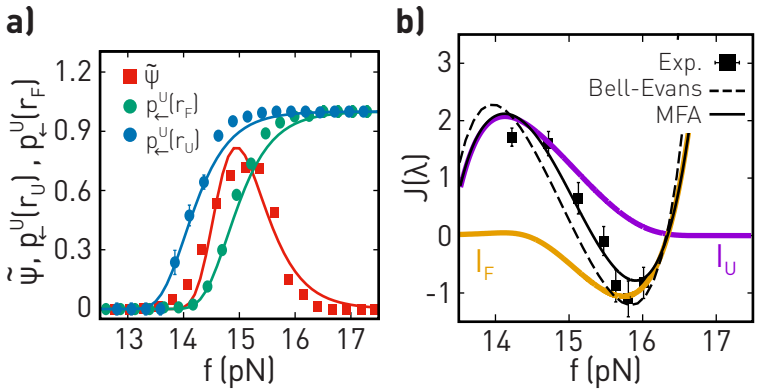

Figure 3.12.: **Partial thermodynamic information *J(λ)* for CTF.** (a) $p^{U}_{\leftarrow}(\lambda, r_U)$, $p^{U}_{\leftarrow}(\lambda, r_F)$ and last-folding density $\widetilde{\psi}(\lambda)$. Solid lines are fits to the Bell-Evans model. (b) Comparison between the $J(\lambda)$ derived from the detailed feedback-FT Eq.(3.24a) (squares) using the data from Fig.3.11 and the theoretical prediction using the Bell-Evans model (dashed line) and the MFA approximation with $\log(c'/c) \sim 0$ (solid line). $I_F(\lambda)$ (orange) and $I_U(\lambda)$ (purple) are the partial thermodynamic information of DTF using the Bell-Evans model and the experimental loading rates.

reduction is larger in CTF than in DTF, however this comes at the price of a larger k_BTY, leading to $\eta_M \cong 0$. Commonly, in non-feedback pulling experiments, dissipation reduction is achieved by simply reducing the loading rate (i.e., making the process less irreversible). However, this comes at the price of an increase in the average time per pulling cycle and a decrease of the total number of pulls per day of experiments, rendering free energy determination inefficient.

The interesting problem is to reduce dissipation with feedback while keeping the average time per pulling cycle equal or lower to the average time per pulling cycle without feedback. Here, we explore the possibility of modifying the feedback protocols in such a way that the dissipation reduction, $\Delta\langle W_d \rangle$, is maximized relative to k_BTY.

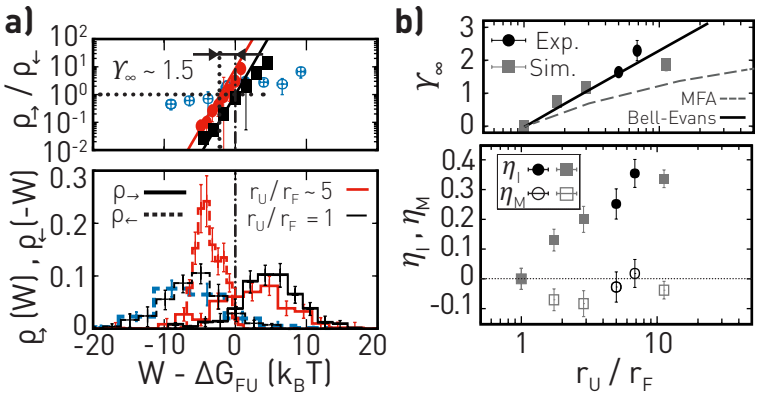

Figure 3.13.: **Thermodynamic information for CTF.** (a) Top: Test of the full feedback-FT Eq.(3.25a) (red circles) with shift Υ_∞ with respect to the non-feedback case (black squares). Blue circles show a negative test with an unweighed reverse work distribution. Bottom: Work distributions. (b) (Top) Thermodynamic information Υ_∞ and (Bottom) efficiencies η_I and η_M from the experimental data (circles) and simulations (squares) versus r_U/r_F. We also show the theoretical predictions for Υ_∞, Eq.(3.25b), using $p^U_{\leftarrow}(r_U)$, $p^U_{\leftarrow}(r_F)$, $\widetilde{\psi}$ derived from the Bell-Evans model (continuous lines) and the MFA Eq.(3.28b) (dashed line). Notice that $\eta_M < \eta_I$, as expected.

By definition, in the DTF protocol, the pulling rate is increased only when the molecule is found to be in U at λ_1, while no action is taken if the molecule is in F. In the single-hopping approximation (Sec. 3.2.2), dissipation reduction is the product of the fraction of trajectories that are in U at λ_1 ($p^U_{\rightarrow}(r_F)$), and the dissipated work reduction conditioned to the U-type trajectories (Eq. (3.17)). At high forces ($f > 15$pN) $p^U_{\rightarrow}(r_F)$ is large, whereas dissipation reduction is low (Fig. 3.15). Conversely, at intermediate forces ($13 - 14$pN), $p^U_{\rightarrow}(r_F)$ is small whereas dissipation reduction is larger. Maximal $\Delta\langle W_d\rangle$ is found close to the coexistence force where the terms $p^U_{\rightarrow}(r_F)$ and $[\langle W_d\rangle_{U\to F\to U}(r_U) - \langle W_d\rangle_{U\to F\to U}(r_F)]$ balance each other. To further reduce dissipation one might consider applying a feedback action also to the large set of F-type trajectories at λ_1.

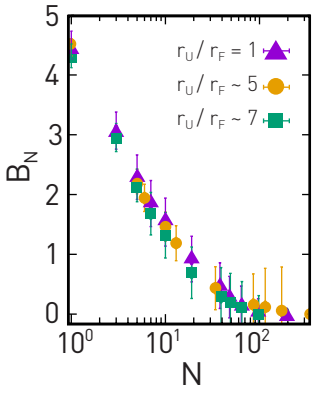

Figure 3.14.: **Experimental bias for CTF.** Bias for the studied molecules with CTF (squares, circle) compared with the non-feedback case (triangles). As for DTF, there is no improvement in free energy prediction.

For example, one might reduce the pulling rate after λ_1 for the F-type of trajectories to reduce the dissipated works for those trajectories.

To show that η_M can be optimized, we have implemented a feedback strategy by combining DTF and CTF protocols. In the DTF+CTF strategy, the molecule is initially pulled at r_F with DTF until λ_1, where a observation is made. If the outcome is U, the pulling rate is switched to $r_U > r_F$ between λ_1 and λ_{max}. Instead, if the outcome is F, the pulling rate is reduced to $r'_F < r_F$ and the CTF protocol turned on. In this case, at the first unfolding event after λ_1, the pulling rate is switched to $r_U > r_F > r'_F$ until λ_{max}. In the DTF+CTF protocol both U- and F-trajectories contribute to reduce the dissipated work. Moreover, the value of r'_F is chosen such that the average time per pulling trajectory is lower compared to the non-feedback case. In Figure 3.16a,b we show the results obtained from numerical simulations of hairpin L4 applying the DTF+CTF strategy with $r_F = 4 > r'_F = 1$pN/s and $r_U = 17$pN/s.

In the coexistence force region ($f_1 \simeq 14.5$pN) dissipated work is reduced by roughly 50% while Y remains unchanged with respect to the standard CTF protocol, leading to higher efficiencies $\eta_I \sim 0.6$, $\eta_M \sim 0.4$

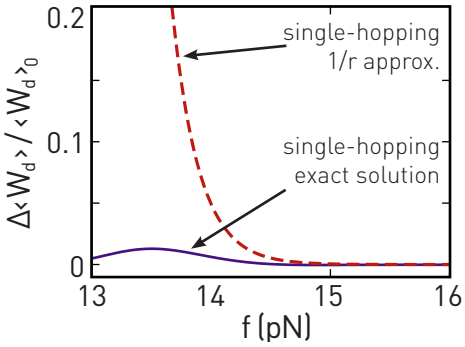

Figure 3.15.: **Dissipation reduction.** Dissipation reduction versus force from the exact single-hopping approximation solution Eq.(3.17) (dark blue line) and the $1/r$ approximation Eq.(3.19) (dashed red line).

(Fig.3.16b). Moreover, we find that B_1 decreases by $\sim 1k_BT$ for the studied parameters (Fig. 3.17a). In addition, we compared the bias as a function of the total experimental time for the four studied protocols (non-feedback, DTF, CTF, and DTF+CTF) using the same pulling rates. In Figure 3.17b, we show the time dependence of the bias, while in inset, we present the time dependence of the number N simulated trajectories. Notice that although CTF generates the largest number of trajectories, the DTF+CTF strategy is the most efficient one. Therefore, in DTF+CTF, dissipation has been reduced efficiently improving the free energy determination.

3.2.5 *Efficiency plot*

To present all results in perspective, we introduce the efficiency plot (Figure 3.18). In this plot, we show the dissipation reduction $\Delta\langle W_d\rangle = \langle W_d\rangle_0 - \langle W_d\rangle$ versus k_BTY, both normalized by the non-feedback dissipation value $\langle W_d\rangle_0$. We show the results for hairpin L4 obtained from experiments and simulations (yellow and green symbols), using DTF and CTF (squares and circles) and DTF+CTF strategy (red trian-

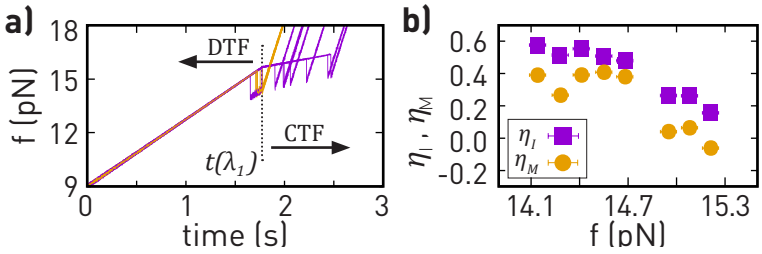

Figure 3.16.: **DTF + CTF strategy.** (a) Force versus time for the DTF + CTF strategy. The pulls start with DTF at $r_F = 4$pN/s and the measurement position is at λ_1. If the molecule is observed to be in U at λ_1, the loading rate is changed to $r_U = 17$pN/s (orange trajectories). If the molecule is observed to be in F at λ_1 the CTF is turned on with a loading rate $r'_F = 1$pN/s $< r_F$ (purple trajectories). At the first unfolding event, the loading rate is changed from r'_F to r_U. (b) Efficiencies η_I and η_M for DTF+CTF. The values of Y for this protocol were directly determined from the Jarzynski equality on Eq. (3.1) for B_1 because we do not have an analytical expression for the thermodynamic information (see Fig. 3.17).

gles). The black-dashed line $\Delta\langle W_d\rangle = k_BT\mathrm{Y}$ separates two regions: $\eta_M > 0$ (second law's weakening, yellow region) and $\eta_M < 0$ (second law's strengthening, white region). Remarkably, despite of dissipation reduction $\Delta\langle W_d\rangle > 0$, all results for DTF and CTF fall on the region $\eta_M \cong 0$ (squares and circles, dashed line) indicating that the second law is strengthened with feedback.

To evaluate the improvement in free energy determination, we express $\Delta\langle W_d\rangle/\langle W_d\rangle_0$ and $k_BT\mathrm{Y}/\langle W_d\rangle_0$ as sole functions of (η_I, η_M):

$$\frac{\Delta\langle W_d\rangle}{\langle W_d\rangle_0} = \eta_I \frac{\eta_M - 1}{\eta_I - 1} \tag{3.29a}$$

$$\frac{k_BT\mathrm{Y}}{\langle W_d\rangle_0} = \frac{\eta_M - \eta_I}{\eta_I - 1}\,, \tag{3.29b}$$

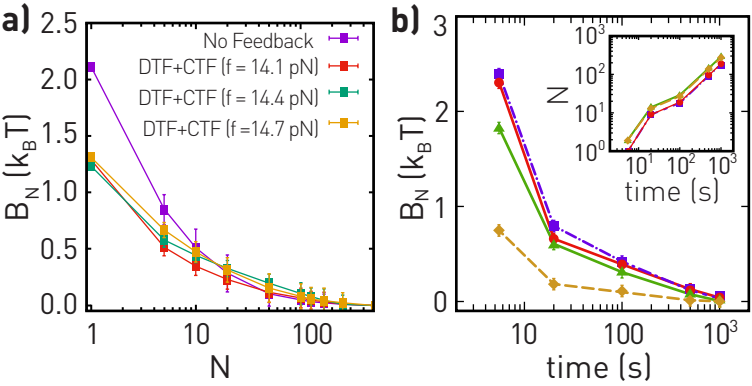

Figure 3.17.: **Jarzynski bias for the DTF + CTF strategy.** (a) Bias for the DTF+CTF strategy at four different λ_1 (or $f(\lambda_1)$ force is shown in the legend) compared with the bias for the case without Feedback. (b) Bias versus the total experimental times measured from numerical with no feedback (purple), DTF (red), CTF (green), and DTF+CTF (yellow) protocols using $r_F = 4\text{pN/s}$, $r_U = 17\text{pN/s}$, and $r'_F = 1\text{pN/s}$. Inset: Simulated number of trajectories (N) for a given total experimental time.

which for $\eta_M = 0$ gives $\Delta\langle W_d\rangle = k_B\text{TY}$ (black dashed line in Fig.3.18). Efficiencies η_I, η_M separately define two independent linear relations between $\Delta\langle W_d\rangle/\langle W_d\rangle_0$ and $k_B\text{TY}/\langle W_d\rangle_0$,

$$\frac{\Delta\langle W_d\rangle}{\langle W_d\rangle_0} = \eta_I + \eta_I \frac{k_B\text{TY}}{\langle W_d\rangle_0} \tag{3.30a}$$

$$\frac{\Delta\langle W_d\rangle}{\langle W_d\rangle_0} = \eta_M + \frac{k_B\text{TY}}{\langle W_d\rangle_0}\,. \tag{3.30b}$$

These relations are shown as dotted (Eq.(3.30a)) and dashed (Eq.(3.30b)) lines in Fig.3.18 with slopes equal to η_I and 1, and intersections with the y-axis equal to η_I, η_M, respectively. For a given point in the efficiency plot, we can read the values of η_I, η_M by drawing lines of slopes η_I and 1 to match the values η_I, η_M in the y-axis. As we can see, the DTF+CTF strategy yields the largest efficiencies for the largest $k_B\text{TY}/\langle W_d\rangle_0$ values measured in CTF (circled region). The efficiency plot shows there

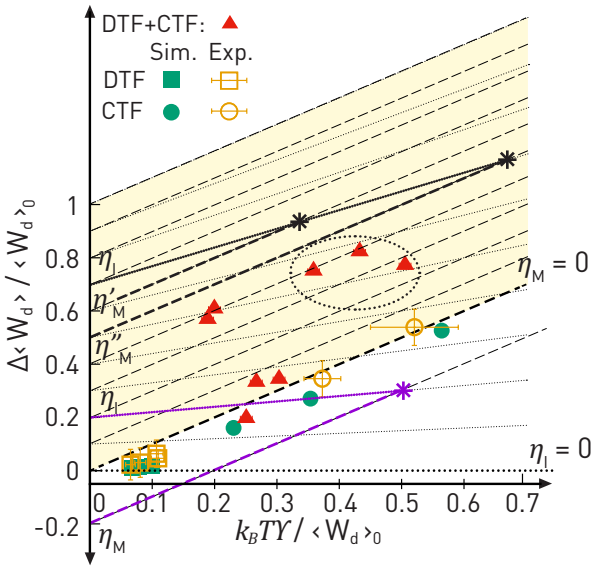

Figure 3.18.: **Efficiency plot.** Dissipation reduction, $\Delta\langle W_d\rangle$, versus thermodynamic information, k_BTY, normalized by the non-feedback dissipation value, $\langle W_d\rangle_0$, in all explored cases for hairpin L4: CTF (empty circles, experimental data; full circles, simulated data), DTF (empty squares, experimental data; full square, simulated data) and the DTF+CTF strategy (red triangles). Dotted line (Eq.(3.30a)) and dashed line (Eq. (3.30b)) have slopes η_I and I and intersect the y-axis at η_I, η_M, respectively. The thick dotted and dashed lines correspond to the case where $\eta_I = 0$ and $\eta_M = 0$.

is room for improved free energy prediction, opening the question of finding strategies that maximize η_M.

In addition, in the efficiency plot in Fig.3.18 we have added three hypothetical cases to show how to interpret this diagram. The first two examples are represented by two black asterisks corresponding to $\eta_I = 0.7$ in the feedback region where $\eta_M > 0$ (yellow region). Notice that for a given value of η_I, many η_M are compatible. The dotted line that passes through the two points intersect the Y-axis at $\Delta\langle W_d\rangle / \langle W_d\rangle_0 = \eta_I = 0.7$, whereas the dashed lines have slopes equal to 1 and intersect the Y-axis

at $\Delta\langle W_d\rangle/\langle W_d\rangle_0 = \eta'_M = 0.6$ and $\Delta\langle W_d\rangle/\langle W_d\rangle_0 = \eta''_M = 0.5$. These two points correspond to $\Delta\langle W_d\rangle/\langle W_d\rangle_0 = 0.93,\ 1.13$ and $k_BT\Upsilon/\langle W_d\rangle_0 = 1/3,\ 2/3$. At this η_I condition, an increase of $\sim 16\%$ in η_M requires doubling Υ and a reduction of 25% in $\langle W_d\rangle$. The last example is represented by the purple asterisk in the $\eta_M < 0$ white region in Fig. 3.18 falling in the purple doted line, $\eta_I = 0.2$, and purple dashed line, $\eta_M = -0.2$. This point indicates an experimental condition where the information-to-energy conversion is low efficiently ($\eta_I \gtrsim 0$) and a remarkable inefficiency at the level of information-to-measurement conversion ($\eta_M < 0$).

3.3 CONCLUSIONS

We have introduced cycle efficiencies η_I, η_M for information-to-work (dissipation reduction, $\Delta\langle W_d\rangle > 0$) and information-to-measurement (second-law inequality weakening, $\langle W_d\rangle + k_BT\Upsilon < \langle W_d\rangle_0$) in irreversible pulling experiments with discrete-time (DTF) and continuous-time feedback (CTF). These are particular cases of the first-time feedback (1^{st}TF) protocol where the pulling rate r_F switches to r_U the first time the molecule unfolds along a predetermined sequence of M trap positions. A detailed and full feedback-FT has been derived for such a protocol that is expressed in terms of the free energy difference, ΔG_{FU}, between the unfolded and folded states (Eqs.(3.7, 3.10)), and in terms of two new quantities, namely the partial information J_k and the full thermodynamic information Υ_M. For $M = 2$, 1^{st}TF maps onto DTF, Eqs.(3.12a, 3.12b), which applied to two-state molecules reduces dissipation by at most $k_BT\Upsilon_2 = k_BT\log 2$ (Landauer limit). In the opposite case, $M \to \infty$, we obtain a novel work-FT for CTF Eqs.(3.24a, 3.25a) for the partial ($J(\lambda)$) and full thermodynamic information (Υ_∞), which is amenable to experimental test. Note that Υ_∞ is finite and unbounded, a consequence of the fact that the information-content of the stored sequences diverges.

We have carried out experiments for DTF and CTF on hairpin L4 for pulling rates in the same range $r_F \sim 4 - 5$pN/s, $r_U \sim 17 - 23$pN/s. The experiments have been complemented with numerical simulations of

a phenomenological model, and theoretical estimates of the Bell-Evans two-state model in the mean-field and single-hopping approximations. We find that CTF leads to higher Y and η_I compared to DTF. Indeed, CTF profits on early and rare unfolding events during the pulling protocol, making Y and η_I larger. In contrast, both DTF and CTF are inefficient regarding η_M: $\langle W_d \rangle$ decreases by roughly $k_B T$Y leaving the second law inequality unweakened and the Jarzynski bias almost unchanged with feedback. In fact, by strategically combining DTF and CTF we can make information-to-measurement conversion efficient. The DTF+CTF strategy maximizes dissipated work reduction without increasing $k_B T$Y leading to high η_I, η_M values.

The explored feedback protocols are summarized in the efficiency plot, which demonstrates that efficient information-to-measurement conversion is obtained by maximizing the dissipation reduction, $\Delta\langle W_d \rangle / \langle W_d \rangle_0$, while minimizing the information, $k_B T\text{Y} / \langle W_d \rangle_0$. Our results show that feedback strategies (defined as a set of multiple-correlated feedback protocols) enhance the information-to-measurement efficiency, opening the door to find optimal strategies for improved free energy determination.

Part III

PROTEIN FOLDING

4

THERMODYNAMIC PROPERTIES OF BARNASE

MOTIVATION

How proteins fold into their native functional structure from a random coiled polypeptide chain configuration stands as one of the major open questions in science. This enigma is a fundamental problem in biophysics, and its answer could be crucial for protein design. In this chapter, we play our part in this problem by studying the paradigmatic protein Barnase (described in Chap. 2 in Sec. 2.1.2) using pulling experiments at different temperatures.

4.1 INTRODUCTION AND HISTORICAL CONTEXT

The first famous observation related to protein folding was in 1961 when Anfinsen introduced the thermodynamic hypothesis claiming that proteins spontaneously fold to a free energy minimum under appropriate conditions [52, 53]. Later, in 1969, Levinthal noticed that a polypeptide chain could not fold into the native state by random search in configurational space [54]. Protein folding is akin to finding a needle in a haystack and must be driven by molecular forces [55]. In an effort to solve the paradox, Ptitsyn proposed the molten globule hypothesis (MGH) where folding is similar to solid formation from a gas: a molten

globule state must precede protein folding, similarly to the metastable liquid phase preceding solid formation during gas deposition [56]. The molten globule is a necessary intermediate to form the native state (hereafter denoted as N) that is structurally similar to it, but with most bonds not yet formed. For years scientists have searched for folding intermediates, the most natural solution to Levinthal's paradox. While these have been identified in large proteins, many small globular proteins fold in a two-states manner, raising the question whether such molten globule intermediate does exist. Methods such as the phi-value analysis have shown that the transition state (hereafter referred to as TS) of two-state globular proteins is structurally similar to the native state [20, 57, 58]. The TS of two-state folders is a disguised molten globule of very short lifetime whose thermodynamic properties reflect those of the molten globule intermediate.

A new direction of thought emerged in the late 80's, Wolynes and collaborators proposed the energy landscape hypothesis (ELH): proteins fold in a funnel-like energy landscape (the native state being the deepest minimum in the energy landscape) by following different and productive folding trajectories [59, 60]. Albeit not excluded, intermediates are not obligatory folding steps. In both scenarios, MGH and ELH, the thermodynamics of the TS has generic and unique properties: on the one hand, a large energy barrier separates TS and N; on the other hand, there is a large configurational entropy loss upon forming the TS from the random coil or unfolded state (hereafter denoted as U).

To evaluate the different hypotheses computer simulations and experiments are employed [61, 62]. For the latter, it is crucial to have tools for accurately measuring the thermodynamics and kinetics of folding. Besides bulk techniques (e.g., NMR, mass spectrometry, calorimetry, etc.) single molecule fluorescence and force spectroscopy offer complementary insights on the protein folding problem. With these, individual proteins are manipulated and monitored with enough temporal resolution to observe short-lived intermediates [63–65], and to measure transition path times along kinetic barriers [66]. Key results are the demonstration

that the ribosome promotes the efficient folding of the nascent polypeptide chain [67], and the role of protein mechanical properties on nuclear translocation [68]. Single-molecule evidence of protein folding intermediates has been reported for RNAseH [69,70], the coiled-coil leucine zipper [71,72] and calmodulin [73]. Recently, the molten globule of apomyoglobin has been shown to be highly deformable under force [74] and an off-pathway molten globule has been observed in apoflavodoxin [75]. In other cases proteins fold in a two-states manner without detectable intermediates (e.g., PrP protein [72]) and a molten globule of very short lifetime transiently forms along the folding pathway [76,77].

Over the past decades, there has been much effort in determining the thermodynamic properties of transition states in two-state globular proteins. How much the enthalpy and entropy of the TS do differ from the native state? What is the heat capacity change (ΔC_p) between the TS and the N and U conformations? How the TS properties change by varying the external conditions (e.g., temperature, ionic strength, and pH)? Answering these questions is essential to understand the features of the different hypotheses (e.g., the *liquid-like* properties of the molten globule in the MGH or the funnel's shape the ELH) and the nature of the folding process itself.

Calorimetry techniques have been used to address such questions. Upon heating, proteins melt at a characteristic temperature T_m at which the heat capacity at constant pressure, C_p, shows a peak [78,79]. The heat capacity change, ΔC_p, between the native and unfolded states is directly related to the change in the number of degrees of freedom, Δn, across the transition, $\Delta C_p = \Delta n \cdot k_B/2$, with k_B being the Boltzmann constant. Therefore, ΔC_p quantifies the configurational entropy loss.

As mentioned in Chap 1, optical tweezers are suitable for calorimetric measurements, however most studies have been carried out at ambient temperature 25°C due to the difficulty of controlling temperature [80,81]. For many years this limitation has challenged direct enthalpy and entropy measurements over a wide range of temperatures, rendering ΔC_p inaccessible to single molecule assays. Here, we investigate the fold-

ing thermodynamics and kinetics of protein barnase using recently built temperature-jump optical trap (see Chap. 2, Sec.2.2) suitable for single-molecule force spectroscopy above and below 25°C. The high solubility and stability of barnase makes it an excellent model to investigate the folding kinetics of globular proteins. In addition, from bulk experiments it has been proved that Barnase reversibly folds in a two-state manner between U and N.

Here, we pull barnase in the range 7°C - 37°C and derive the folding free energy (ΔG), entropy (ΔS), and enthalpy (ΔH) by combining fluctuation theorems and kinetics. From the temperature dependence of ΔS and ΔH, we derive ΔC_p. Finally, we also determine the entropy, enthalpy and ΔC_p of the TS relative to the native and unfolded states.

4.2 RESULTS

Figure 4.1a shows five unfolding and folding selected force-distance curves (FDCs) for the six investigated temperatures, i.e., 7, 14, 18, 25, 32, and 37°C. It is apparent that the lower the temperature, the higher the unfolding force and the FDC hysteresis. In Figure 4.1b we show a single pulling cycle at 25°C. During stretching (red curve), barnase unfolding is observed as a sudden force rip ($\Delta f \sim 2$pN) in the FDC at forces $\sim 25 - 30$pN. Upon force release (blue curve), a folding transition is detected as a small force jump (~ 0.5pN) at forces $\lesssim 5$pN. The left encircled zoomed inset shows the two force branches: where barnase is folded in its native state (N-branch, black dashed line) and unfolded (U-branch, grey solid line). The relative trap position (λ) in the two branches contains the trap bead displacement plus the handles extension and barnase molecular extension. The difference between both branches at a given force, $\Delta\lambda$ (right circled zoomed inset), is the difference of molecular extensions between the polypeptide chain and the projection on the force axis of the dipole formed by the N- and C- termini of barnase.

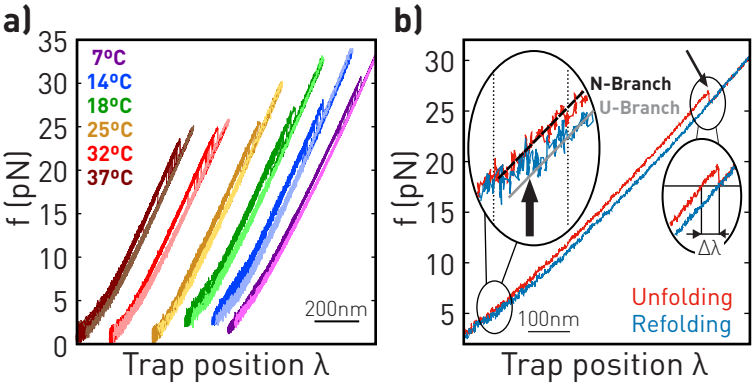

Figure 4.1.: **Temperature-dependent force-distance curves of barnase.** (a) Unfolding (dark) and folding (light) force-distance curves (FDCs) at different temperatures, 7°C (purple), 14°C (blue), 18°C (green), 25°C (yellow), 32°C (red) and 37°C (brown). FDCs at each temperature have been shifted along the x-axis for clarity. (b) Unfolding (red) and refolding (blue) trajectories at 25°C. Left Zoom: Refolding event (arrow) from the U-branch (gray solid) to the N-branch (black dashed). Right Zoom: Unfolding event (arrow) highlighting $\Delta\lambda$.

To determine the folding entropy and enthalpy of barnase, we proceeded in two different ways. First, we derived the temperature-dependent folding free energy at zero force, $\Delta G_0(T)$, using the fluctuation theorem (see section 4.2.2). Second, we derive $\Delta G_0(T)$ from the measured kinetic rates using the Bell-Evans model (see section 4.2.3). After we use a Clausius-Clapeyron like equation to derive ΔS (mathematical details in App. E). To use this approach it is necessary to determine the elastic response of protein barnase (see section 4.2.1).

4.2.1 *Elastic response of the polypeptide chain*

The temperature-dependent elastic properties of the polypeptide chain were determined using the molecular extension $x_p(f,T)$ derived from $\Delta\lambda(f,T)$ (Fig. 4.1b),

$$x_p = \Delta\lambda + x_d \,. \tag{4.1}$$

where $x_p(f,T)$ is the dipole extension for the folded barnase.

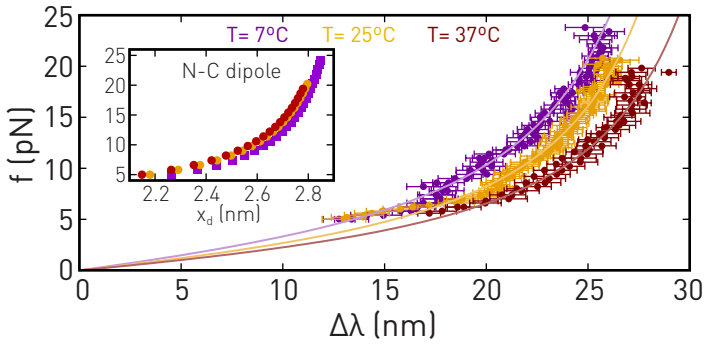

Figure 4.2.: **Force versus difference in trap position.** Force versus difference in trap position ($\Delta\lambda$) at three temperatures: 7ºC (purple), 25ºC (yellow) and 37ºC (brown). Inset: Elastic response of the folded protein modelled as a dipole of 3nm with the FJC model.

Figure 4.2 shows the force versus $\Delta\lambda$ response measured at three selected temperatures (7, 25 and 37ºC). To extract $x_p(f,T)$ from Eq.(4.1) we model $x_d(f,T)$ with the Freely-Jointed Chain elastic model (Fig. 4.2, inset), assuming that the distance between the N- and C-termini for the folded barnase (the dipole length taken equal to 3μm) is constant with temperature. By comparing Figure 4.2 (inset and main) we observe that $x_d(f,T) \ll \Delta\lambda(f,T)$, as expected since the dipole length is much shorter than the polypeptide extension. Therefore, $x_p(f,T)$ increases with T at a given f, making the polypeptide chain stiffer with temperature. $x_p(f,T)$

is well described by the inextensible Worm-Like Chain (WLC) model and its interpolation formula [82],

$$f = \frac{k_B T}{4L_p}\left(\left(1 - \frac{x_p(T)}{N_{aa}\cdot d_{aa}}\right)^{-2} + 4\cdot\frac{x_p(T)}{N_{aa}\cdot d_{aa}} - 1\right) \tag{4.2}$$

where L_p is the persistence length, N_{aa} is the number of residues (110 for barnase), and d_{aa} is the distance between consecutive amino acids.

The data relative to each investigated temperature were fit to Eq.(4.2) with L_p and d_{aa} as free parameters. The fits to the inextensible WLC model at each temperature are shown in Fig. 4.3.

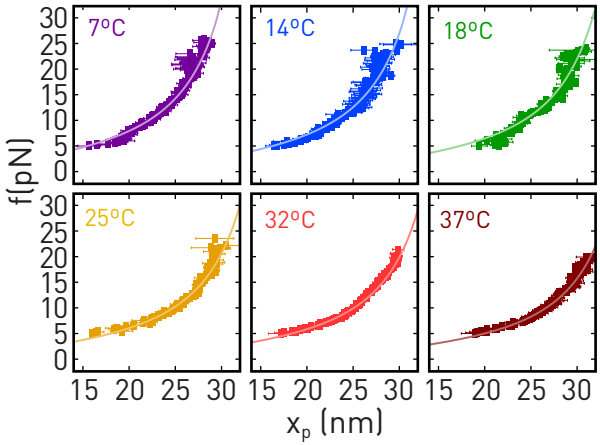

Figure 4.3.: **Elastic response of unfolded barnase.** Fits of the measured force versus polypeptide chain extension (x_p) to the WLC model (solid lines).

Figure 4.4 shows the best fitting parameters at each temperature. While L_p shows a strong T dependence, which is well approximated by a linear function of slope 0.011 ± 0.001 nm/K (Fig. 4.4a), d_{aa} presents a weak T-linear dependence of slope 0.008 ± 0.002 Å/K (Fig. 4.4b), which is one order of magnitude smaller than for L_p. Therefore, d_{aa} can be taken as constant, ~ 3.7 Å or 0.37 nm.

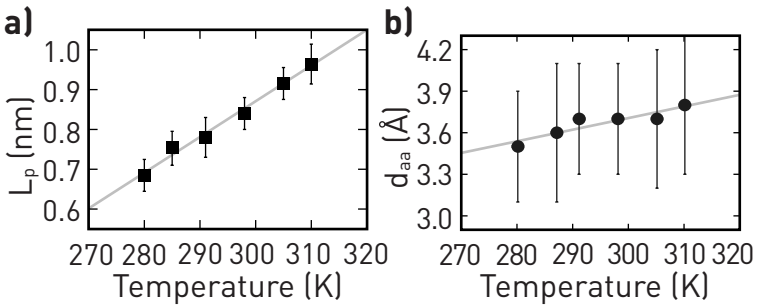

Figure 4.4.: **Elastic parameters of the polypeptide chain.** (a) Persistence length (L_p) and (b) amino acid distance (d_{aa}) of the polypeptide chain calculated from the fits in panel shown in Fig. 4.3. Solid lines are linear fits to the experimental points.

4.2.2 *Thermodynamic potentials from work measurements*

To derive folding free energy at zero force, ΔG_0, we use the thermodynamic relation that relates ΔG_0 with the elastic contributions of the polypeptide chain $\Delta G_{\rm p}$, the dipole $\Delta G_{\rm d}$, and the folding energy at force f, $\Delta G(f)$ (mathematical details in App. E),

$$\Delta G_0 = \Delta G_{\rm d}(0 \to f) + \Delta G(f) - \Delta G_{\rm p}(0 \to f) \tag{4.3}$$

The results of the previous section can be used to determine $\Delta G_{\rm p}$ and $\Delta G_{\rm d}$, however it remains unanswered how to measure $\Delta G(f)$ in such experiments. In optical tweezers experiments, the relative trap position λ is the control parameter, rather than the force which fluctuates depending on the molecular state. A thermodynamic relation similar to Eq.(4.3) holds by a Legendre transforming $f \to \lambda$ to the λ-ensemble (Appendix F). ΔG_0 is determined by measuring the free energy difference, ΔG_λ, between a minimum and a maximum trap positions where barnase is folded ($\lambda_{\rm min}$) and unfolded ($\lambda_{\rm max}$),

$$\Delta G_0 = \Delta G_\lambda - W_\lambda^{\rm elas} \tag{4.4}$$

where W_λ^{elas} stands for the elastic contributions of the setup (bead, handles, polypeptide chain and protein dipole) that must be subtracted to ΔG_λ (Appendix F).

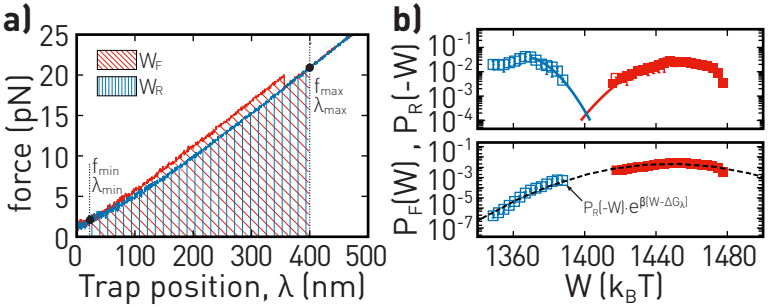

Figure 4.5.: **Work measurements.** (a) The unfolding (red) and folding (blue) work W are defined as the area below the FDCs limited by λ_{min} and λ_{max} (red and blue areas). (b) Top: Unfolding (red full squares) and folding (blue empty squares) work distributions at 25ºC. Bottom: $P_U(W)$ and $P_F(-W) \times \exp(W - \Delta G_\lambda / k_B T)$. The black dashed line is a Gaussian fit to determine ΔG_λ.

We used the work fluctuation theorem (work-FT) [83] to determine ΔG_λ from irreversible work (W) measurements by integrating the FDC between the minimum and maximum trap positions, $W = \int_{\lambda_{min}}^{\lambda_{max}} f d\lambda$ (Fig. 4.5a). Let $P_U(W)$ and $P_F(W)$ denote the unfolding and folding work distributions measured over many pulling cycles. The work-FT is given by,

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

where ΔG_λ equals the reversible work. Note the minus work sign in $P_F(-W)$ in the left-hand-side of Eq.4.5, which is a consequence of the fact that $W < 0$ in the folding process.

A corollary of Eq.4.5 is the Jarzynski equality,

$$\Delta G_\lambda = -k_B T \log\langle \exp\left(-\frac{W}{k_B T}\right)\rangle, \tag{4.6}$$

where $\langle(...)\rangle$ is the average over many (infinite) realizations. In practice, as shown in Chap. 3, the number of pulls is always finite and the Jarzynski equality is strongly biased. From Eq.4.5, work distributions cross at $W = \Delta G_\lambda$. However, the crossing point is not observed due to the high hysteresis of the FDCs, quantified by the dissipated work. The dissipated work is in the range $50 - 150 k_B T$, which is many times the value of ΔG_0 (see below). This makes the measurements of ΔG_λ unreliable because the tails of P_U and P_F do not cross.

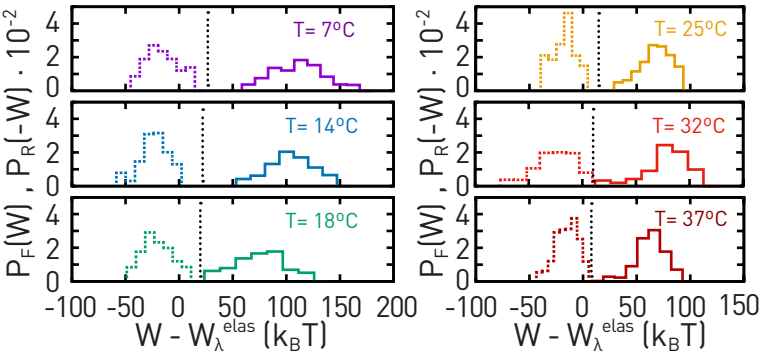

Figure 4.6.: **Work distributions at the six explored temperatures.** Unfolding (solid line) and folding (dashed lines) work distributions at different temperatures (with the elastic contribution W_λ^{elas} being subtracted from the total work). The dotted vertical line indicates ΔG_0 for each temperature.

In these limit conditions, one can still use the matching method [10], that provides reasonable free energy estimates and is comparably simpler than other mathematical approaches. In this method, the value of ΔG_λ is determined by matching the functions $P_U(W)$ and $P_F(-W) \times \exp(W - \Delta G_\lambda / k_B T)$. In practice, the leftmost and rightmost tails of $P_U(W)$ and $P_F(-W)$, respectively, are fitted to a generic form, such as $\exp\left(-|W - W_{\max}|^\delta / \Omega\right)$, to characterize its shape [84]. Figure 4.5b (top) shows $P_U(W)$, $P_F(-W)$ at 25ºC and the fitted tails. Rightmost (P_F) and leftmost (P_U) tails are well fitted with $\delta \sim 1.9$, indicating Gaussian-like tails ($\delta = 2$). Therefore, we simultaneously fitted $P_U(W)$

and $P_F(-W) \times \exp(W - \Delta G_\lambda / k_B T)$ to a single Gaussian distribution (black dashed line in Fig. 4.5b bottom) to find the value of ΔG_λ that best matches both distributions.

To determine ΔG_0 from ΔG_λ in Eq. (4.4) we determined the elastic contributions (W_λ^{elas} in Eq. (4.4)) as follows: the bead term was calculated for a linear optical trap of stiffness ~ 0.07pN/nm; the DNA handles term was calculated by integrating the WLC model with the temperature-dependent elastic parameters presented in Chap. 6; the ΔG_{p} and ΔG_{d} contributions were calculated using the elastic parameters presented in section 4.2.1. In Figure 4.6 we show the $P_U(W)$ and $P_F(-W)$ at different temperatures. Distributions are plotted versus $W - W_\lambda^{\text{elas}}$ instead of W, to directly determine ΔG_0 with the matching method (Fig. 4.6, dotted vertical lines). The values of $\Delta G_0(T)$ present a clear temperature dependence (Fig. 4.7) as expected from the relation, $\Delta G_0 = \Delta H_0 - T\Delta S_0$.

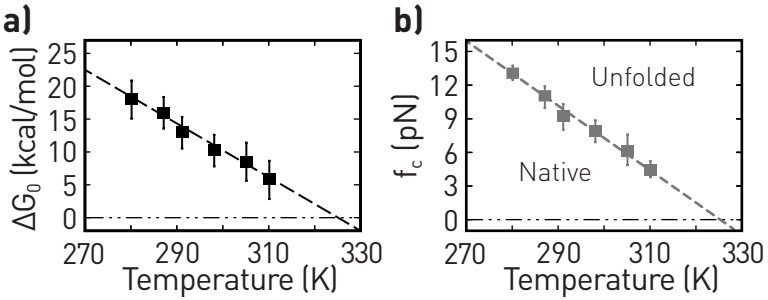

Figure 4.7.: **Folding free energy and coexistence force.** (a) Experimental values of ΔG_0 derived from Fig. 4.6 together with a fit to $\Delta G_0(T) = \Delta H_0 - T\Delta S_0$ (dashed line). (b) Coexistence force as a function of temperature. Dashed line corresponds to a linear fit to $f_c(T)$.

Entropy and Enthalpy derivation

To derive the temperature-dependent entropy difference between N and U from the measured values of $\Delta G_0(T)$, we use the following equation (mathematical derivation shown in App. E),

$$\Delta S_0(T) = -\frac{\partial f_c(T)}{\partial T}\Delta\lambda(f_c(T)) - \int_0^{f_c(T)} \frac{\partial \Delta\lambda(f', T)}{\partial T} df' \quad . \tag{4.7}$$

However, we do not have a direct measurement of f_c from pulling experiments. Instead, $f_c(T)$ is derived using the values of ΔG_0 and Eq. (4.3) evaluated at f_c at each temperature. Figure 4.7b shows the obtained values of $f_c(T)$. Notice that $f_c(T)$ decreases with T and it is well fitted by a linear trend, thus defining a $f - T$ phase diagram separating the native and unfolded states.

Finally, as mentioned before, we derived $\Delta S_0(T)$ using Eq. (4.7) and $f_c(T)$ (Fig. 4.8a). It changes by roughly 22% in the whole temperature range indicating a finite ΔC_p. Notice that the numerical T-derivative of $\Delta G_0(T)$ is roughly constant (Fig. 4.7a), which indicates that using Eq. (4.7) is the most reliable way to estimate $\Delta S_0(T)$. Folding enthalpies, $\Delta H_0 = \Delta G_0 + T\Delta S_0$, are shown in Fig. 4.8b.

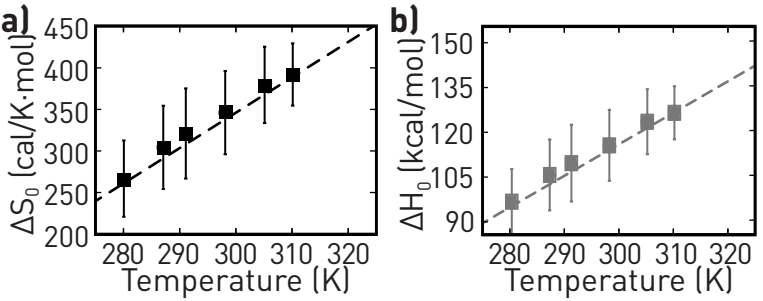

Figure 4.8.: **Entropy and Enthalpy.** (a) Entropy and (b) Enthalpy difference between N and U. Dashed lines are fits to Eq.(4.8)a and Eq.(4.8)b, respectively

Heat capacity change

Bulk assays have shown that barnase has a finite ΔC_p. The marked temperature dependence in ΔS_0 and ΔH_0 (Fig. 4.8) allows us to extract ΔC_p across the melting transition. To do so, we expand ΔH_0 and ΔS_0 around the melting temperature T_m,

$$\Delta S_0(T) = \Delta S_0^m + \Delta C_p \cdot \log\left(\frac{T}{T_m}\right) \tag{4.8a}$$

$$\Delta H_0(T) = \Delta H_0^m + \Delta C_p \cdot (T - T_m) \ , \tag{4.8b}$$

where ΔS_0^m and $\Delta H_0^m = T_m \Delta S_0^m$ are the entropy and enthalpy at T_m, and ΔC_p is the heat capacity change between N and U. $\Delta S_0(T)$ and $\Delta H_0(T)$ were fitted to Eqs. (4.8a) and (4.8b) (dashed lines in Fig. 4.8) with ΔC_p, ΔH_0^m, ΔS_0^m, and T_m as fitting parameters. The values that best match the experimental values are: $\Delta C_p = 1030 \pm 43$ cal/mol·K (average value), $\Delta S_0^m = 431 \pm 10$ cal/mol·K and $\Delta H_0^m = 140 \pm 6$ kcal/mol.

4.2.3 *Thermodynamic potentials from kinetic measurements*

Second, we derive $\Delta G_0(T)$, $\Delta S_0(T)$, and $\Delta H_0(T)$ from the unfolding and folding kinetic rates measured from the pulling experiments at different temperatures. Figure 4.9a shows the unfolding and folding force distributions ($\rho_\rightarrow(f)$, $\rho_\leftarrow(f)$) at three selected temperatures (7, 25 and 37ºC). We extract the unfolding and folding kinetic rates, $k_\rightarrow(f)$ and $k_\leftarrow(f)$, from the corresponding survival probabilities (Fig. 4.9b). For pulling experiments where f is ramped at constant loading rate $r = |df/dt|$, the following relations hold,

$$\frac{dP_N(f)}{df} = -\frac{k_\rightarrow(f)}{r} P_N(f) \ \Rightarrow \ k_\rightarrow(f) = -r\,\frac{\rho_\rightarrow(f)}{P_N(f)} \tag{4.9a}$$

$$\frac{dP_U(f)}{df} = \frac{k_\leftarrow(f)}{r} P_U(f) \ \Rightarrow \ k_\leftarrow(f) = r\,\frac{\rho_\leftarrow(f)}{P_U(f)} \tag{4.9b}$$

with $P_N(f) = 1 - \int_0^f \rho_{\rightarrow}(f)df$ and $P_U(f) = 1 - \int_f^{\infty} \rho_{\leftarrow}(f)df$ being the survival probabilities of N and U, respectively.

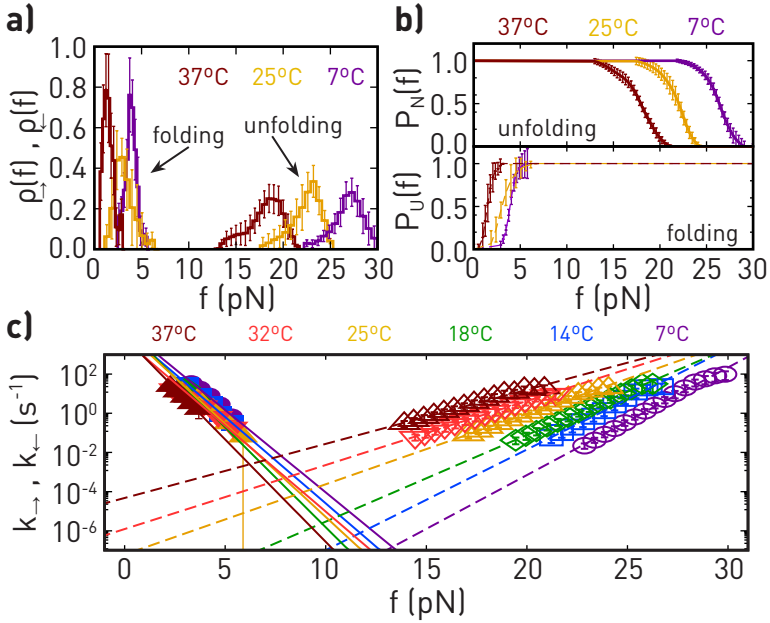

Figure 4.9.: **Temperature-dependent kinetic rates.** (a) Unfolding and folding force distributions at 7, 25, and 37ºC. (b) Top: Survival probability of N along unfolding trajectories. Bottom: Survival probability of U along folding trajectories. (c) Unfolding (empty symbols) and folding (solid symbols) kinetic rates for all temperatures. The solid (refolding) and dashed (unfolding) lines are fits to the Bell-Evans model.

The measured $k_{\rightarrow}(f)$ and $k_{\leftarrow}(f)$ are shown in Figure 4.9c in a log-normal scale, for all the studied temperatures. Kinetic rates are well described by the Bell-Evans model [85,86],

$$k_{\rightarrow}(f) = k^0_{\rightarrow} \exp\left(\frac{f \cdot x^{\dagger}}{k_B T}\right) \; ; \; k^0_{\rightarrow} = k_a \exp\left(-\frac{\Delta G^{\dagger}}{k_B T}\right) \tag{4.10a}$$

$$k_{\leftarrow}(f) = k^0_{\leftarrow} \exp\left(-\frac{f \cdot x^{*}}{k_B T}\right) \; ; \; k^0_{\leftarrow} = k_a \exp\left(-\frac{\Delta G^{*}}{k_B T}\right) . \tag{4.10b}$$

Parameters are: $k^0_{\rightarrow}$ ($k^0_{\leftarrow}$) the unfolding (folding) rate at zero force; k_a the attempt rate; $\Delta G^{\dagger}$ (ΔG^*) the height of the kinetic barrier relative to N (U) extrapolated to zero force; and $x^{\dagger}$ (x^*) the distance of the TS with respect to N (U).

The measured kinetic rates shown in Fig. 4.9c are fitted to the left hand side of Eqs. (4.10a) and (4.10b) to extract $x^{\dagger}$, x^*, $k^0_{\rightarrow}$, and $k^0_{\leftarrow}$. From these data we determined f_c as the force value at which $k_{\rightarrow} = k_{\leftarrow}$ (Fig. 4.10a). We use the detailed balance condition, $\Delta G_0 = -k_B T \log(k^0_{\leftarrow}/k^0_{\rightarrow})$, to determine ΔG_0 from the kinetic rates extrapolated at zero force, $k^0_{\leftarrow}$, $k^0_{\rightarrow}$ (Fig. 4.10b).

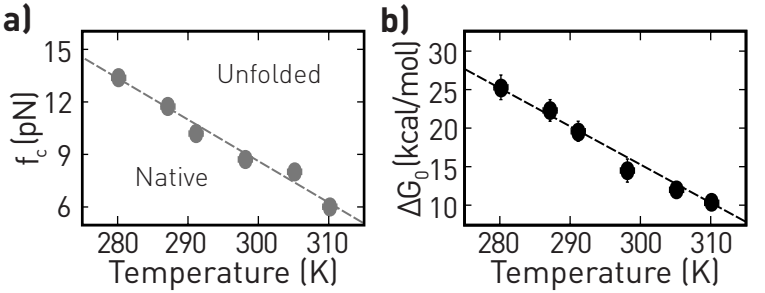

Figure 4.10.: **Temperature-dependent folding free energy.** (a) Coexistence force determined as the crossing point where $k_{\rightarrow} = k_{\leftarrow}$. (b) Folding free energy as a function of temperature derived from the detailed balance condition.

The temperature dependence of the TS position, $x^{\dagger}$ and x^*, is investigated in Chap. 5 in more detail.

Finally, as in the previous section, we have used the obtained values of $f_c(T)$, $\Delta G_0(T)$, and $\Delta\lambda(f,T)$ to derive the entropy across unfolding using Eq. (4.7). The obtained $\Delta S_0(T)$ and $\Delta H_0(T)$ derived from the kinetic rates are presented in Fig. 4.11. These values agree with those obtained from the FT. Fitting $\Delta S_0(T)$ and $\Delta H_0(T)$ to Eqs. (4.8a) and (4.8b) gives: $\Delta C_p = 990 \pm 56$ cal/mol·K (average value), $\Delta S^m_0 = 529 \pm 12$ cal/mol·K, and $\Delta H^m_0 = 172 \pm 8$ kcal/mol.

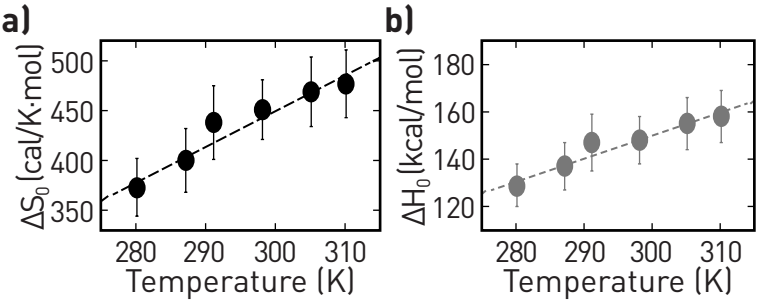

Figure 4.11.: **Temperature-dependent entropy and enthalpy.** (a) Entropy and (b) Enthalpy difference between N and U. Dashed lines are fits to Eq.(4.8)a and Eq.(4.8)b, respectively

4.2.4 *Transition state's entropy and enthalpy*

We determine the enthalpy and entropy of the TS relative to N ($\Delta S^{\dagger}$, $\Delta H^{\dagger}$), and U (ΔS^{*}, ΔH^{*}), to characterize the molecular free energy landscape governing the folding of barnase. Here, we follow the criteria $\Delta^{\dagger} = \Delta^{TS} - \Delta^{N}$ and $\Delta^{*} = \Delta^{TS} - \Delta^{U}$. To this aim, we rewrite $k^0_{\leftarrow}$, $k^0_{\rightarrow}$ in Eqs. (4.10a) and (4.10b) in terms of the TS entropies and enthalpies,

$$k^0_{\rightarrow}(T) = k_a \exp\left(\frac{\Delta S^{\dagger}}{k_B}\right) \exp\left(-\frac{\Delta H^{\dagger}}{k_B T}\right) \tag{4.11a}$$

$$k^0_{\leftarrow}(T) = k_a \exp\left(\frac{\Delta S^{*}}{k_B}\right) \exp\left(-\frac{\Delta H^{*}}{k_B T}\right). \tag{4.11b}$$

We performed a simultaneous fit of $k^0_{\rightarrow}$ and $k^0_{\leftarrow}$ to Eqs. (4.11a) and (4.11b) to derive the values of $\Delta S^{\dagger}$, ΔS^{*}, $\Delta H^{\dagger}$, ΔH^{*}. Interestingly, we found that $k^0_{\rightarrow}$ is strongly T-dependent, while $k^0_{\leftarrow}$ depends weakly, hinting at an entropy-driven folding process (Fig. 4.12).

The four-parameters fit was done by imposing two constraints: $\Delta S^{*} = \Delta S^{\dagger} - \Delta S_0$; and $\Delta H^{*} = \Delta H^{\dagger} - \Delta H_0$. For the fits to Eq. (4.11a) and (4.11b) the values of $\Delta S_0(T)$ and $\Delta H_0(T)$ have been taken as the mean values obtained from the FT (Fig. 4.8) and kinetics (Fig. 4.11). Moreover, we

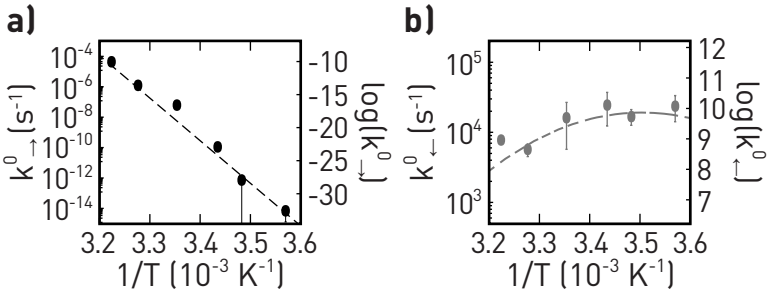

Figure 4.12.: **Kinetic rates at zero.** (a) Unfolding (b) and refolding kinetic rates at zero force versus $1/T$. Dashed lines are simultaneous fits to Eqs. (4.11a) and (4.11b). While $k^0_{\rightarrow}$ varies over ten order of magnitude, $k^0_{\leftarrow}$ varies over one order of magnitude.

used the attempt rate previously obtained on the same molecular system in similar experimental conditions [87], $k_a \sim 150\text{s}^{-1}$. Fits are shown as dashed lines in Fig. 4.12. *TS* entropies and enthalpies are shown in Figure 4.13. Fitting them to Eqs. (4.8a) and (4.8b) permits us to extract the heat capacity change between *N* and *TS* ($\Delta C_p^{N-TS} = C_p^{TS} - C_p^N$) and between *TS* and *U* ($\Delta C_p^{TS-U} = C_p^U - C_p^{TS}$). Notice that the criteria for the sign of ΔC_p^{TS-U} is the opposite of ΔH^* and ΔS^*. We obtain $\Delta C_p^{N-TS} \sim 100$ cal/mol·K and $\Delta C_p^{TS-U} \sim 900$ cal/mol·K, which gives the folding $\Delta C_p \sim 1000$ cal/mol·K.

4.2.5 *The folding funnel*

Tables 4.1, 4.2, and 4.3 summarize the main results: the energy differences between states *N* and *U* (ΔG_0, ΔH_0, and ΔS_0 in Tab. 4.1); the barrier energies to unfold, *N-TS* ($\Delta G^{\dagger}$, $\Delta H^{\dagger}$, and $\Delta S^{\dagger}$ in Tab. 4.2); and the barrier energies to fold, *U-TS* (ΔG^*, ΔH^*, and ΔS^* in Tab. 4.3).

The measured free energy landscape of barnase is illustrated in Fig. 4.14. Results show that barrier entropies, enthalpies and free energies

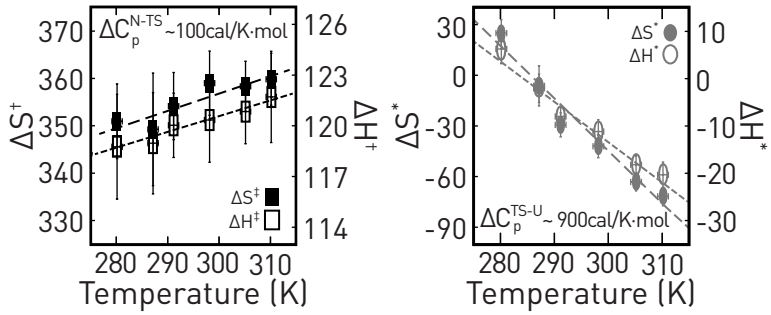

Figure 4.13.: **Thermodynamics's of *TS*.** Entropy (full symbols) and enthalpy (empty symbols) differences between N and TS (†, black); and between U and TS (*, gray). We follow the sign criteria $\Delta^\dagger = ^{TS} - ^N$ and $\Delta^* = ^{TS} - ^U$.

T	ΔG_0	ΔH_0	ΔS_0
7	21 ± 2	112 ± 10	326 ± 34
14	19 ± 2	121 ± 11	355 ± 37
18	16 ± 2	128 ± 13	383 ± 40
25	11 ± 2	131 ± 11	401 ± 35
32	10 ± 2	139 ± 11	421 ± 12
37	8 ± 2	142 ± 10	432 ± 31
50	0	156 ± 8	479 ± 10

Table 4.1.: Folding thermodynamic properties of barnase: T in ^{o}C; ΔG and ΔH in kcal/mol; ΔS in cal/mol·K.

to fold ($U \rightarrow TS$) are one order of magnitude smaller than the corresponding barriers to unfold ($N \rightarrow TS$): $|\Delta S^*| \ll |\Delta S^\dagger|$, $|\Delta H^*| \ll |\Delta H^\dagger|$ and $|\Delta G^*| \ll |\Delta G^\dagger|$. This difference suggests a folding process in two steps. In a first step, the unfolded protein reaches a TS with a few bonds ($\sim$ 20%) formed relative to the fully denatured state. In a second step, the protein collapses into N by forming the rest of native bonds ($\sim$ 80%). These bond percentages are estimated from the different enthalpy values for the TS relative to N and U ($\Delta H^\dagger \sim 120$ kcal/mol and $\Delta H^* \sim -30$ kcal/mol at T_m).

T	$\Delta G^{\dagger}$	$\Delta H^{\dagger}$	$\Delta S^{\dagger}$
7	22 ± 3	118 ± 3	351 ± 8
14	19 ± 3	119 ± 3	349 ± 12
18	17 ± 2	120 ± 2	354 ± 7
25	13 ± 3	120 ± 2	359 ± 7
32	11 ± 2	121 ± 2	358 ± 5
37	10 ± 3	122 ± 3	360 ± 6
50	4 ± 2	123 ± 2	364 ± 3

Table 4.2.: Thermodynamic properties of barnase between N and TS: T in °C; ΔG and ΔH in kcal/mol; ΔS in cal/mol·K. We follow the sign criteria $\Delta^{\dagger} = {}^{TS} - {}^{N}$.

T	ΔG^{*}	ΔH^{*}	ΔS^{*}
7	1 ± 3	6 ± 6	25 ± 9
14	0 ± 3	-2 ± 4	-6 ± 11
18	1 ± 3	-8 ± 6	-29 ± 9
25	2 ± 3	-11 ± 6	-42 ± 9
32	1 ± 3	-18 ± 6	-63 ± 8
37	2 ± 3	-20 ± 6	-72 ± 9
50	4 ± 2	-33 ± 5	-115 ± 13

Table 4.3.: Thermodynamic properties of barnase between TS and U: T in °C; ΔG and ΔH in kcal/mol; ΔS in cal/mol·K. We follow the sign criteria $\Delta^{*} = {}^{TS} - {}^{U}$.

A remarkable difference is found in ΔC_p between TS and N or U (Fig. 4.13). The main contribution to the total $\Delta C_p = 1010 \pm 47$ cal/mol·K comes from the formation of the TS coming from U ($\Delta C_p^{\mathrm{TS-U}} = 900 \pm 80$ cal/mol·K), which is ~ 9 times larger than between N and TS ($\Delta C_p^{\mathrm{N-TS}} = 105 \pm 26$ cal/mol·K). The value of ΔC_p is directly proportional to the change in the number of degrees of freedom (Δn), $\Delta C_p = \Delta n \cdot k_B/2$ which gives $\Delta n \sim 1$ per cal/mol·K unit in ΔC_p. This gives $\Delta n^{TS-U} \sim 900 \gg \Delta n^{N-TS} \sim 100$ showing that the main configurational entropy loss occurs upon forming the TS from U. This result depicts the TS as a molten globule of high free energy ($\Delta G^{\dagger} \sim \Delta G_0$) and low configurational entropy ($\Delta C_p^{N-TS} \ll \Delta C_p$), which is structurally similar to the native state: the major change in ΔC_p and Δn occurs between U and TS.

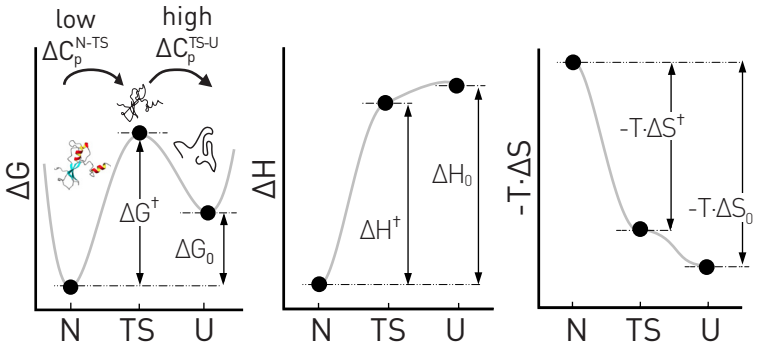

Figure 4.14.: **Molecular energy landscape.** Energy (left), enthalpy (center), and entropy (right).

4.3 CONCLUSIONS

We have used calorimetric optical tweezers to measure the FDCs of protein barnase in the range 7-37ºC and derived the folding thermodynamics at the single molecule level. An analysis based on the Clausius-Clapeyron equation (Eq. (4.7)) was used to extract the temperature- dependent values of $\Delta S_0(T)$ and $\Delta H_0(T)$ and ΔC_p. The obtained results agree with those obtained from bulk experiments in our conditions of ionic strength (20 mM monovalent salt) and neutral pH (7.0). The mean values for the enthalpy, entropy and T_m derived from the work-FT and kinetic analysis ($\Delta H_0^m = 156 \pm 8$ kcal/mol, $\Delta S_0^m = 479 \pm 10$ cal/mol·K and $T_m = 50 \pm 2$ºC) agree with those reported in the literature and collected in [88], $\Delta H_m \sim 115 - 145$ kcal/mol and $\Delta S_m \sim 400$ cal/mol·K. Our estimation of $\Delta C_p = 1010 \pm 47$ cal/mol·K also agrees with values obtained from differential scanning and isothermal titration calorimetry assays, as well as with recent atomistic numerical simulations [89]. Measurements of ΔC_p in calorimetric experiments often require determining the temperature dependence of ΔH with pH, ionic strength, or the denaturant concentration. In contrast, with calorimetric force spectroscopy

we directly measure thermodynamic potentials and kinetics at a given temperature.

Remarkably, our results feature a TS with the properties of a molten-globule: entropies and enthalpies between TS and U ($\Delta S^*, \Delta H^*, \Delta G^*$) are $\sim$ 9 times lower than the corresponding ones between TS and N, ($\Delta S^\dagger, \Delta H^\dagger, \Delta G^\dagger$). In fact, the low value of ΔG^* correlates with the high compliance of the molten globule upon stretching, as has been shown for apomyoglobin [74]. A general feature of free energies $(0, \dagger, *)$ in Tables 4.1, 4.2, and 4.3 is the compensation observed between entropy ($T\Delta S$) and enthalpy (ΔH) contributions, i.e., $\Delta G = \Delta H - T\Delta S \ll |\Delta H|, |T\Delta S|$. In contrast, the major contribution to ΔC_p occurs between U and TS ($\Delta C_p^{TS-U} \gg \Delta C_p^{N-TS}$). The value of ΔC_p^{TS-U} is proportional to the reduction ($\Delta n^{TS-U} \sim 900$) in the number of degrees of freedom (dof) between TS and U, at a rate of ~ 1 dof/(cal/mol·K).

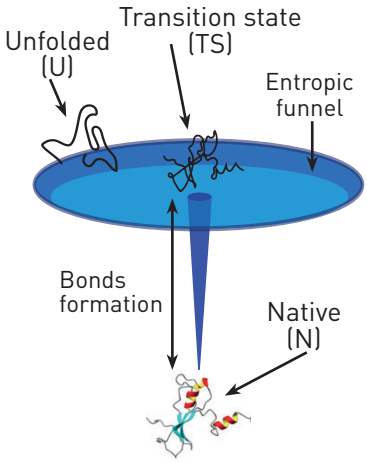

Figure 4.15.: **Golf-hole course folding free energy landscape.**

Our results agree also with the folding funnel scenario of the energy landscape hypothesis (ELH) (Fig. 4.14), where a large entropy loss ($\sim$ 90%) occurs upon forming the TS from U. Such entropy loss is accompanied by a low enthalpy change ($\sim$ 20% of the total folding en-

thalpy). These results suggest that in the TS the protein already presents a structure similar to N with 20% of the native bonds formed. We hypothesize that the protein structure in the TS is guided by interactions between the polypeptide chain and the surrounding water molecules. The formation of a water-layer clathrate limits the accessible volume of the protein facilitating the collapse from TS to N and the stabilization of the protein core. The TS is a molten globule structurally similar to N, stabilized by an outer hydration shell ($\Delta H^* < 0$) but with the internal contacts not formed. The large configurational entropy loss between U and TS demonstrates that folding is an entropically driven process in a golf-course energy landscape, where the TS is the native hole. The collapse from TS to N forms most of the native bonds accounting for most of the entropy and enthalpy of folding ($\Delta S_0 \simeq \Delta S^\dagger, \Delta H_0 \simeq \Delta H^\dagger$) (see Fig. 4.15). Overall, our results also validate the main predictions of the ELH. Three thermodynamic inequalities summarize our results: $|\Delta S^*| \ll \Delta S^\dagger, |\Delta H^*| \ll \Delta H^\dagger$ and $\Delta C_p^{TS-U} \gg \Delta C_p^{N-TS}$. These are key inequalities for molecular folding in line with predictions of the molten globule and energy landscape hypotheses.

5

FORCE DEPENDENCE OF PROTEINS' TRANSITION STATE POSITION

MOTIVATION

Single-molecule force spectroscopy has opened a new field of research in molecular biophysics and biochemistry. Pulling on single proteins permits us to manipulate and monitor molecular conformational transitions with high temporal resolution. In single protein pulling experiments, force-extension curves often present large hysteresis, with higher unfolding than refolding forces. To date, most experiments based on pulling experiments have reported a discrepancy between the sum of the folding and unfolding transition state distances using the kinetic Bell-Evans model and the full molecular extension. Here, we investigate this discrepancy by using the two-state folder protein barnase from the Bell-Evans model and the worm-like chain elastic model.

5.1 INTRODUCTION AND HISTORICAL CONTEXT

As mentioned in Chap. 4, one of the most prominent open questions in science is how proteins fold. Two of the most accepted hypotheses to explain protein folding based on experimental observations are the *energy landscape* and *foldon* hypotheses. The first describes the protein

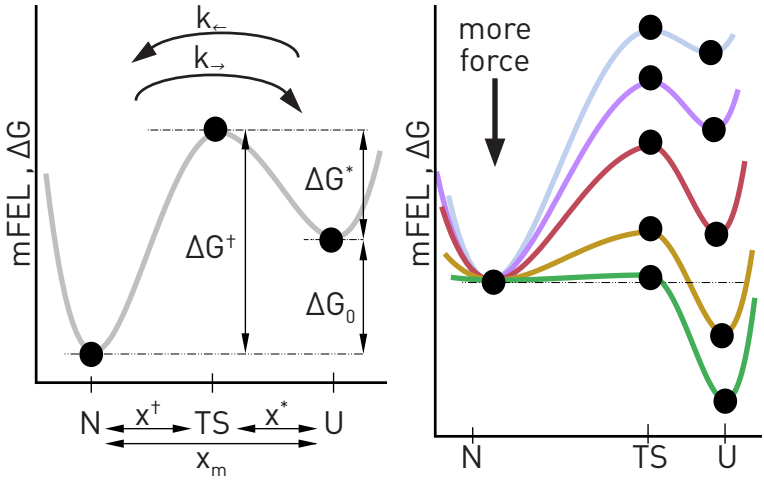

Figure 5.1.: **Illustration of the free energy landscape for the Bell-Evans model.** Left: Scheme of the mFEL at a given force highlighting the position of the transition states, as well as the height of the kinetic barrier. Right: Force effect on the mFEL. As we increase the force, the barrier decreases and the unfolded state becomes more stable.

folding process as a thermally activated transition between states on a funneled energy landscape with the native state at the bottom of the funnel [59]. The funnel can present secondary minima establishing intermediate states with short lifetimes where the polypeptide chain is partially folded. Hence, the unfolded protein can fold into the native state following arbitrary trajectories connecting intermediates. In the late '80s, bulk hydrogen exchange, NMR, and mass spectrometry studies and theoretical models mostly by Englander and collaborators consistently observed intermediates upon folding [90]. Based on these results, the foldon hypothesis claims that, as a result of evolution, proteins fold along a unique path in the energy landscape via a sequence of well-defined intermediates or foldons [91].

Single-molecule techniques have been used to investigate the thermodynamics of individual proteins with high temporal resolution [92, 93]. In such experiments, force is used to mechanically break the bonds that stabilize the tertiary structure of proteins. The force required to unfold proteins is exerted in the N- and C- termini of the polypeptide chain, defining a proper reaction coordinate, the end-to-end distance or molecular extension, useful to describe the molecular free energy landscape (mFEL) guiding protein folding [71, 72, 94]. Over the years, single-molecule experiments have permitted the reconstruction of the free energy landscape of a wide variety of proteins, both characterized by a two-state manner folding/unfolding process [72, 95–97] and in the presence of intermediates states [69–71, 73, 98]. The energy landscape governing the unfolding and folding reactions is widely investigated using the of Bell-Evans (BE) model [85, 86, 99, 100]. For a two-state system, the mFEL is simplified by two wells representing the native (N) and unfolded (U) states separated by a kinetic barrier placed in-between at the transition state (TS) (Fig. 5.1). The Bell-Evans model assumes that the position of the barrier relative to the native ($x^{\dagger}$) and unfolded (x^{*}) states is fixed, whereas its height is reduced (increased) relative to N (U), proportionally to the external applied force, upon increasing it (Fig. 5.1). Moreover, the kinetics to unfold and fold, $k_{\rightarrow}$ and $k_{\leftarrow}$, are proportional to the exponential of the kinetic barrier, being force-dependent. Two different experimental approaches are usually exploited to investigate the folding and unfolding transitions using single-molecule techniques: hopping and pulling experiments.

On the one hand, hopping experiments are used to derive $k_{\rightarrow}$ and $k_{\leftarrow}$ from force-time traces by determining the average residence times at each state. It is known that hopping experiments provide good estimations for the molecular extension $x_m = x^{\dagger} + x^{*}$, and the coexistence force f_c (i.e., the force at which $k_{\rightarrow}(f_c) = k_{\leftarrow}(f_c)$) for molecules with residence times within the accessible experimental range. Thus, proteins presenting a high hysteresis, i.e. the unfolding and folding events occurring at different force ranges, have kinetic rates that are often inaccessible to hopping experiments. On the other hand, in pulling experiments,

which are commonly used to derive the folding free energy and the elastic properties of the unfolded polypeptide chain, the kinetic rates are derived from the survival probabilities of N and U. Interestingly, in these experiments, the timescales of the molecule do not represent an inconvenience to derive $k_{\rightarrow}$ and $k_{\leftarrow}$, thus they can be used to investigate high mechanically stable proteins. However, it has been shown that the molecular extension x_m derived by using the Bell-Evans model underestimates the predictions based on the elastic properties of the proteins [87]. The following questions arise: Can we use the BE model to study the kinetics of mechanically stable proteins in irreversible pulling experiments? Which information can we derive from such experiments?

To answer these questions, we investigate the unfolding/folding kinetic rates of the two-state folder protein barnase using pulling experiments. In fact, as shown in Chap. 4, barnase presents strong irreversibility upon mechanical folding and unfolding. Therefore, this protein is a good candidate to discuss the potentiality, reliability and limitations of pulling experiments. By performing experiments at different temperatures we also derive the temperature-dependence of the unfolding and folding kinetic rates, and the kinetic barrier position.

5.2 RESULTS

To derive the temperature- and force-dependent unfolding and folding kinetic rates of barnase, we have considered the pulling experiments described in Chap. 4 where barnase was pulled with optical tweezers in the temperature range of 7°-37°C. In pulling experiments, the optical trap is repeatedly moved away and towards the micro-pipette at constant loading rate r. At the same time, the force exerted on the ends of the molecular construct is recorded, generating the force-distance curves (FDCs) shown in Fig. 5.2. Along the stretching or unfolding trajectories (dark color curves in Fig. 5.2), the protein is pulled from an initial force, ~ 1 pN, where it is in N, to a final force, ~ 30 pN, where it is in U. The unfolding event is observed as a sudden force drop in the FDCs (black

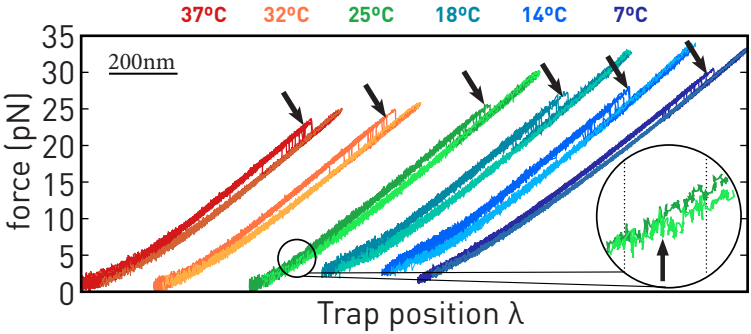

Figure 5.2.: **Pulling experiments at different temperatures.** Measured unfolding (dark color) and folding (light color) force-distance curves (FDCs) at different temperatures, 7°C (dark blue), 14°C (blue), 18°C (turquoise), 25°C (green), 32°C (orange) and 37°C (red). The FDCs are shifted along the x-axis for clarity. The unfolding events at the studied temperatures are indicated with black arrows. The folding event at 25°C is shown in a zoom (circle) highlighting the folding event (black arrow). These FDCs are the ones presented in Fig. 4.1 of Chap. 4.

arrows in Fig. 5.2). When the force is relaxed in the folding process (light color curves in Fig. 5.2), the protein folds back into its native state and the event is observed as a force rip of about 0.5 pN at lower forces, < 5 pN (see zoom in Fig. 5.2). The hysteresis between the forces at which the protein unfolds and folds is remarkable, indicating that barnase folding is highly irreversible when unfolded and folded by mechanical forces.

From the FDCs, the first unfolding and folding force distributions $\rho_{\rightarrow}(f)$ and $\rho_{\leftarrow}(f)$ is derived. As a way of example, in Fig. 5.3a the force distributions obtained at 7, 25, and 37°C have been reported. The plots clearly show that the unfolding (folding) force distributions shift to higher (lower) forces as the temperature is decreased. This is due to the fact that lower temperatures make the protein structure more stable, and a higher force is required to break the bonds that stabilize the native structure. Furthermore, in Fig. 5.3a the irreversibility of barnase

unfolding clearly emerges, being represented by the wide force gap between the unfolding and folding force distributions. This effect can also be seen from the survival probability of the native and unfolded states along the unfolding and folding trajectories, $P_N^{\rightarrow}(f)$ and $P_U^{\leftarrow}(f)$. The survival probabilities are derived by using equations

$$P_N^{\rightarrow}(f) = 1 - \int_0^f \rho_{\rightarrow}(f')df' \tag{5.1a}$$

$$P_U^{\leftarrow}(f) = 1 - \int_f^{\infty} \rho_{\leftarrow}(f')df' \,. \tag{5.1b}$$

Figure 5.3b shows the survival probabilities $P_N^{\rightarrow}(f)$ and $P_U^{\leftarrow}(f)$ at three selected temperatures (7, 25, and 37°C). As expected, the transitions from 1 to 0 of $P_N^{\rightarrow}(f)$ and $P_U^{\leftarrow}(f)$ do not cross in a wide force gap ($6 - 13$pN, highlighted area in Fig. 5.3b) in the temperature range studied in this work, due to the irreversibility of the mechanically induced unfolding/folding process of barnase.

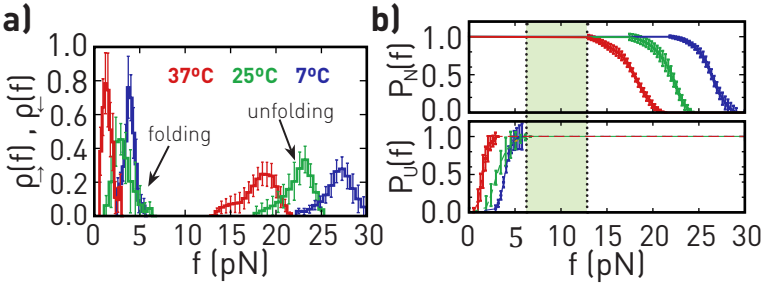

Figure 5.3.: **Survival probabilities at different temperatures.** (a) Unfolding and folding force distributions at 7°C (dark blue), 25°C (green), and 37°C (red). (b) Top (bottom): Survival probability of the native (unfolded) state along the unfolding (folding) trajectory at 7°C (dark blue), 25°C (green), and 37°C (red). The wide force gap ($6 - 13$pN) is indicated by the green shaded area.

Concerning the unfolding and folding kinetics of barnase, in pulling experiments where force is increased/reduced at a constant loading rate

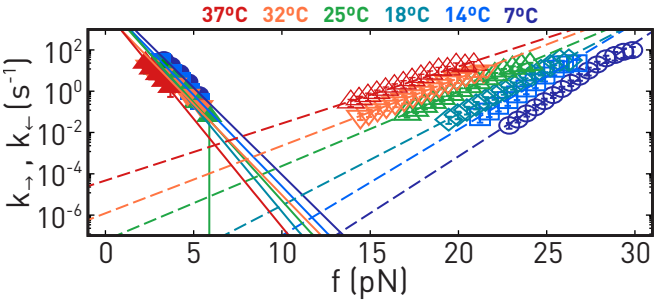

Figure 5.4.: **Temperature-dependent kinetic rates.** Unfolding (empty symbols) and folding (solid symbols) kinetic rates at different temperatures, 7°C (dark blue), 14°C (blue), 18°C (turquoise), 25°C (green), 32°C (orange) and 37°C (red). The dashed and solid lines are fits to Eqs. (5.3a) and (5.3b), respectively.

$r = |df/dt|$, the kinetic rates are derived from the force distributions via the following equations,

$$\rho_{\rightarrow}(f) = \frac{k_{\rightarrow}(f)}{r} P_N^{\rightarrow}(f) \;\Rightarrow\; k_{\rightarrow}(f) = r\,\frac{\rho_{\rightarrow}(f)}{P_N(f)} \tag{5.2a}$$

$$\rho_{\leftarrow}(f) = \frac{k_{\leftarrow}(f)}{r} P_U^{\leftarrow}(f) \;\Rightarrow\; k_{\leftarrow}(f) = r\,\frac{\rho_{\leftarrow}(f)}{P_U^{\leftarrow}(f)}\,. \tag{5.2b}$$

We derived $k_{\rightarrow}(f)$ and $k_{\leftarrow}(f)$ from the FDCs recorded at different temperatures using Eqs. (5.2a) and (5.2b), which are shown in a log-normal plot in Fig. 5.4 for all the investigated temperatures. The effect of the temperature is more evident in the unfolding process, where unfolding kinetic rates are strongly affected by T, in contrast to the folding process where kinetics is weakly temperature dependent. The BE model defines the unfolding and folding kinetic rates as,

$$k_{\rightarrow}(f) = k_a \exp\left(-\beta\Delta G^{\dagger}\right) \cdot \exp\left(\beta f x^{\dagger}\right) \tag{5.3a}$$

$$k_{\leftarrow}(f) = k_a \exp\left(-\beta\Delta G^{*}\right) \cdot \exp\left(\beta(\Delta G_0 - f x^{*})\right) \tag{5.3b}$$

being k_a the attempt rate, ΔG_0 the free energy difference between N and U, and $\Delta G^{\dagger}$ and ΔG^{*} the kinetic barriers measured relative to N and

U, respectively. $\beta = 1/k_B T$, with k_B the Boltzmann constant and T the temperature. A schematic illustration of a general mFEL highlighting these parameters is shown in Fig. 5.1. We fitted Eqs. (5.3a) and (5.3b) to the force behaviour of the experimental values of $k_{\rightarrow}(f)$ and $k_{\leftarrow}(f)$ (dashed and solid lines in Fig. 5.4) to derive the position of TS and the height of the energetic barrier.

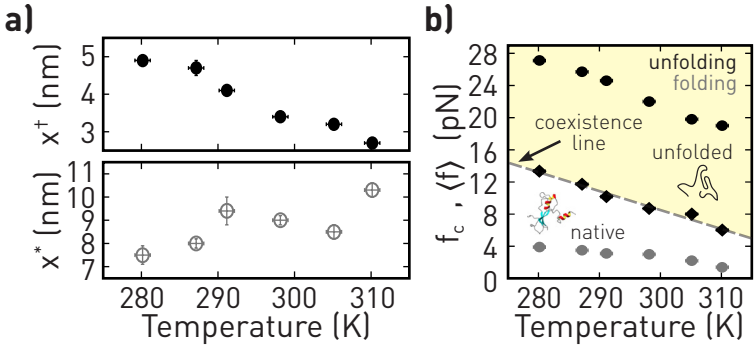

Figure 5.5.: **Bell-Evans fits.** (a) Transition state position relative to N (top) and U (bottom). (b) Unfolding (black circles), folding (gray circles), and coexistence (black diamonds) forces. The coexistence force separates the native and unfolded state in the force-temperature plane.

Figure 5.5a shows the obtained $x^{\dagger}$ (top panel) and x^* (bottom panel) as a function of the temperature. Notice that when the temperature increases, $x^{\dagger}$ becomes shorter and x^* becomes larger, meaning that TS moves towards N. Nevertheless, it is necessary to remark that the fits to Eqs. (5.3a) and (5.3b) are done around the most probable first unfolding and folding force (black and gray circles in Fig. 5.5b), meaning that $x^{\dagger}$ and x^* are evaluated at different forces. We have derived the coexistence force f_c by matching $k_{\rightarrow}$ and $k_{\leftarrow}$ using the values that best fits the BE model to the experimental data (black diamonds in Fig. 5.5b). The temperature dependence of f_c defines the coexistence line and it is well represented with a linear function. The coexistence line separates the native and unfolded states in the force-temperature plane. Above the

coexistence line the most probable state is U (yellow shaded region in Fig. 5.5), whereas below the line the most probable is N. Looking at Fig. 5.5b, we can notice that the force needed to unfold the protein is higher than the coexistence force, while the forces at which the protein folds are below the coexistence force. In addition, at the coexistence force N and U have the same free-energy, meaning that the barrier to unfold or fold are equal.

5.2.1 *Temperature-dependent kinetic barrier*

First, we determined the height of the kinetic barrier from the average unfolding and folding forces using the Bell-Evans model,

$$\langle f_{\rightarrow(\leftarrow)} \rangle = \frac{k_B T}{x^{\dagger(*)}} \log \left(\frac{x^{\dagger(*)} r \cdot e^{(\Delta G^{\dagger(*)}/k_B T)}}{k_a \, k_B T} \right) . \tag{5.4}$$

To derive $\Delta G^{\dagger}$ and ΔG^{*} using Eq. 5.4 it is necessary to know k_a. The attempt rate for protein barnase has been determined previously, $k_a \sim 150$ s^{-1}, using a miniaturized optical tweezers instrument like the one employed in this work but that only operates at room temperature [87]. The derived temperature-dependent $\Delta G^{\dagger}$ and ΔG^{*} are shown in Fig. 5.6a. As known from thermodynamics, the temperature-dependent Gibbs free energy difference between two states is given by $\Delta G = \Delta H - T\Delta S$, where ΔH and ΔS are the enthalpy and entropy difference between states. The measured $\Delta G^{\dagger}$ and ΔG^{*} vary roughly 60% and 50%, respectively, in the explored temperature range permitting us to determine the enthalpy and entropy of the transition state. The derived enthalpy and entropy at the melting temperature in the assumption of zero heat capacity change are: $\Delta H^{\dagger} = 144 \pm 7$ kcal/mol, $\Delta S^{\dagger} = 437 \pm 22$ cal/mol·K, $\Delta H^{*} = 30 \pm 5$ kcal/mol, and $\Delta S^{*} = 89 \pm 15$ cal/mol·K. The measured enthalpy and entropy for the transition state at the melting temperature agree with the values shown in Chap. 4. Notice that ΔG^{*} presents a slight curvature with temperature, whereas $\Delta G^{\dagger}$ is linear. This result evidences that the heat capacity change, ΔC_p, between TS and U (ΔC_p^{TS-U}) is higher

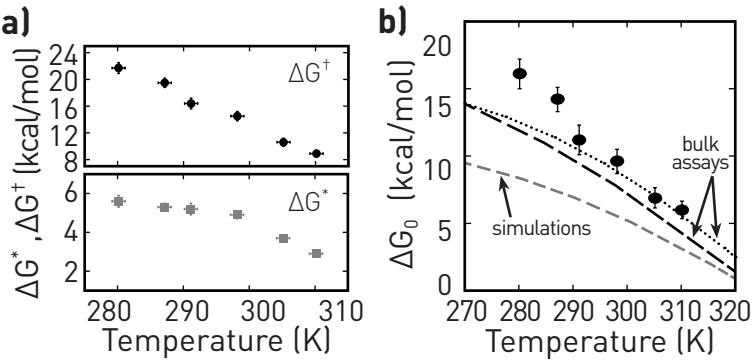

Figure 5.6.: **Temperature-dependent free energy.** (a) Height of the kinetic barrier relative to N (top) and U (bottom). (b) Folding free energy as a function of temperature compared with bulk assays [101,102] (black lines) and numerical simulations [89] (gray line).

than the one between N and TS (ΔC_p^{N-TS}). This has been demonstrated in the previous chapter by directly deriving the temperature-dependent values of $\Delta H^\dagger$, $\Delta S^\dagger$, ΔH^*, and ΔS^*. We measured the folding free energy at each temperature by subtracting ΔG^* to $\Delta G^\dagger$, $\Delta G_0 = \Delta G^\dagger - \Delta G^*$. The measured folding free energy are shown in Fig. 5.6b. We have compared our results with the protein stability curve of barnase reported in calorimetric studies [101,102] and numerical simulations [89]. Again, the measured values for ΔG_0 agree with the values shown in the previous chapter.

5.2.2 *Transition state's position*

Second, we have carried out an in-depth analysis about the position of the kinetic barrier, $x^\dagger$ and x^*, derived from the fits to the BE model to the kinetic rates. As a comparison, we determined the molecular extension, $X(f)$, at the coexistence force as described in Sec. 4.2.1, Chap. 4. The

extension $X(f)$ for the semi-flexible polypeptide chain is well described by inextensible WLC model [82],

$$f = \frac{k_B T}{4L_p}\left[\left(1-\frac{X(f)}{Nd_{aa}}\right)^{-2} + 4\frac{X(f)}{Nd_{aa}} - 1\right] \tag{5.5}$$

where L_p is the persistence length of the polypeptide chain, N is the number of residues, 110 for the case of barnase, and d_{aa} is the distance between consecutive amino acids. The parameters L_p and d_{aa} are derived in Chap. 4, Sec. 4.2.1 by fitting Eq. (5.5) to the measured molecular extension from the FDCs.

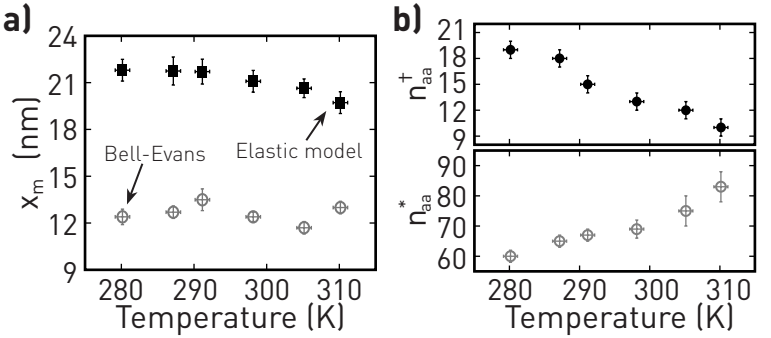

Figure 5.7.: **Molecular extension.** (a) Molecular extension derived from the BE model (empty circles) and from the WLC and FJC models (solid circles) derived from the FDCs measurements in shown in Sec. 4.2.1, Chap. 4. (b) Number of released amino acids at the transition relative to N (top) and U (bottom).

By definition, the total extension $X(f)$ must be equal to $x^{\dagger} + x^{*}$ plus the extension of the native state (x_d), equal to the extension of a dipole oriented along the force axis. This extension is commonly modeled using the Freely-Jointed Chain (FJC) model for a single Kühn segment of length equal the N-C-termini distance, 3 nm for barnase [21]. Figure 5.7a shows the experimental values $x_m = x^{\dagger} + x^{*}$ (solid symbols) together with the expected ones using the WLC and FJC models at the coexistence force, $x_m = X(f_c) - x_d(f_c)$ (empty symbols). There is a sys-

tematic underestimation of ~ 9 nm for the experimental values derived from the BE model (circles) relative to the expected behavior using the temperature-dependent elastic properties (squares).

The discrepancy arises from the strong hysteresis between the unfolding and folding FDCs of barnase. Indeed, as mentioned before, $k_{\rightarrow}$ and $k_{\leftarrow}$ have been adjusted in a range of forces close to the most probable first unfolding and folding force, respectively. This means that $x^{\dagger}$ is evaluated at a force higher than f_c and x^{*} is evaluated at a force below f_c, both forces are far from the estimated coexistence force. It is at the latter force that, the molecular extension has been calculated using the elastic models. Therefore, to properly compare the experiments with the expected behavior, we need to consider the force-dependent elastic response of $x^{\dagger}$ and x^{*}. This can be achieved by deriving the number of released amino acids (aa), n_{aa}, at $x^{\dagger}$ and x^{*}. To do so, we have used Eq. (5.5) to derive n_{aa} at a given force and extension. The derived number of unfolded amino acids relative to N ($n^{\dagger}_{aa}$, Fig. 5.7b-top), and relative to U (n^{*}_{aa}, Fig. 5.7b-bottom) are shown in Fig 5.7b. Notice that the sum of $n^{\dagger}_{aa}$ and n^{*}_{aa} ranges between $80-90$, and does not recover the total number of amino acids of the polypeptide chain (110). This result might be interpreted in terms of the presence of intermediate states in the unfolding/folding pathway. However, calorimetric studies [101, 102] and numerical simulations [89] have demonstrated that barnase folds in a two-state manner. In addition, no evidence of the intermediate has been reported in other single-molecule pulling experiments [87]. Therefore, the most probable explanation is that the TS moves with force. This result does not contradict the BE model, which assumes that the distance between N and TS ($x^{\dagger}$) is force-independent. In other words, as force increases $n^{\dagger}_{aa}$ decreases but the extension $x^{\dagger}$ remains constant. This is evidenced by the nearly perfect linearity of $\log(k_{\rightarrow}(f))$ ($\log(k_{\leftarrow}(f))$) versus force in Fig. 5.4. Thus, we used Eq. (5.5) to derive the number of released amino acids at TS as a function of force during the unfolding process assuming that at a given temperature $x^{\dagger}$ is constant in force.

Figure 5.8a shows the force-dependence of the estimated $n^{\dagger}$ at all the studied temperatures. Interestingly, *TS* moves towards the native state (i.e., $n^{\dagger}_{aa}$ decreases) when the force or the temperature increase. This result correlates with the Leffler-Hammond postulate for chemical reactions [103, 104]. According to such postulate, upon external perturbations, such as the force or temperature employed in our study, the transition state counteracts the increased thermodynamic stability of the unfolded state by approaching the native state.

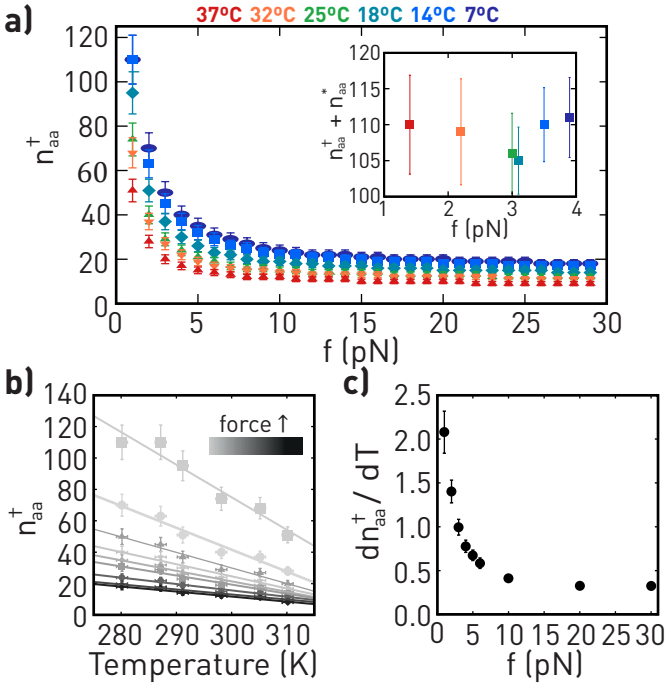

Figure 5.8.: **Force dependence of transition state position.** (a) $n^{\dagger}_{aa}$ as a function of force. Inset: $n^{\dagger}_{aa}$ plus n^{*}_{aa} evaluated at the folding force. (b) $n^{\dagger}_{aa}$ at different temperatures for fixed values of force (from 1pN (light gray) to 30pN (black)). (c) Derivative of $n^{\dagger}_{aa}$ relative to T measured at different forces.

Taking into account that $n^{\dagger}_{aa}$ moves when force is applied, it is evident that $n^{\dagger}_{aa}$ and n^{*}_{aa} can be summed to give the total number of aa (100) only if they have been evaluated at the same force. Hence, we have calculated the position of the barrier $n^{\dagger}_{aa}$ at the most probable folding force $\langle f_{\leftarrow} \rangle$ and summed it to the value of n^{*}_{aa} derived experimentally at the same force. Fig. 5.8a-inset presents the values of $n^{\dagger}_{aa} + n^{*}_{aa}$ evaluated at $\langle f_{\leftarrow} \rangle$ for the different temperatures, showing that the exact number of 110 amino acid is recovered. This analysis confirms the two-state behavior of barnase folding and, above all, the effective movement of TS in amino acid units. TS's motion counteracts the external perturbation as predicted by the Leffler-Hammond postulate.

Finally, to better highlight the role played by force and temperature in promoting the protein unfolding/folding, we have studied the number of released amino acids at fixed forces as a function of the temperature. Fig. 5.8b shows how $n^{\dagger}_{aa}$ varies roughly linearly with T at at a constant force, presenting a slope which is strongly affected by the applied force. Moreover, Fig. 5.8c shows the derivative of $n^{\dagger}_{aa}$ with respect to T at fixed forces. In this case, $dn^{\dagger}_{aa}/dT$ exhibits a significant decay at low forces, whereas at forces above 10 pN, $dn^{\dagger}_{aa}/dT$ reaches a plateau.

5.3 CONCLUSIONS

Here, we employed the optical tweezers technique to pull barnase protein at the single-molecule level, studying the influence of force and temperature on its molecular free energy landscape (mFEL). The interest in barnase stems from the high hysteresis observed in force spectroscopy measurements and from its behaviour as a two-state system upon folding and refolding, as confirmed by bulk experiments [101, 102] and simulations [89]. These features make the protein barnase the ideal example to illustrate the underestimation of the molecular extension that has been reported in literature when pulling experiments and the Bell-Evans model are combined to investigate dissipative proteins.

First, we have characterized the mFEL of the protein by using the Bell-Evans model, overcoming the apparent inconsistency for highly mechanically stable proteins. Using non-equilibrium pulling experiments, we derived the force-dependent unfolding and folding kinetic rates ($k_{\rightarrow}$ and $k_{\leftarrow}$), in a wide temperature range ($7 - 37^{o}$C). The kinetic barrier ($\Delta G^{\dagger}$ relative to N or ΔG^{*} relative to U) has been studied as a function of temperature, by using the Bell-Evans approach and the average unfolding and folding forces determined experimentally. According to the temperature-dependent Gibbs energy definition, we have also derived the entropy and enthalpy of the transition state at the melting temperature relative to N, $\Delta H^{\dagger} = 144 \pm 7$ kcal/mol, $\Delta S^{\dagger} = 437 \pm 22$ cal/mol·K, and relative to U, $\Delta H^{*} = 30 \pm 5$ kcal/mol, and $\Delta S^{*} = 89 \pm 15$ cal/mol·K. We found a good agreement between these results and the ones derived using the detailed balance condition and work fluctuation theorems in Chap. 4. In addition, from $\Delta G^{\dagger}$ and ΔG^{*}, we have estimated the folding free energy of barnase, which match with the ones derived in Chap. 4 and the ones derived from bulk assays and numerical simulations [89, 101, 102].

Second, we have focused our attention on the study of the TS position, relative to the native state N and to the unfolded state U, $x^{\dagger}$ and x^{*} respectively. These positions have been calculated from the kinetic rates and, as expected for highly irreversible unfolding and folding proteins, we found that the sum $x_m = x^{\dagger} + x^{*}$ underestimates the total molecular extension derived from elastic models, like the WLC and FJC ones. To better understand this discrepancy, we derived the number of released amino acids at the transition state, $n^{\dagger}_{aa}$ and n^{*}_{aa}, by considering $x^{\dagger}$ and x^{*} and the WLC equation. We have found that the position of TS in amino acid units moves toward N when the force or the temperature is increased, while its distance remains fixed as stated by the Bell-Evans model. This result agrees with the Leffler-Hammond postulate which states that for chemical reactions the TS moves toward the reactants when an external agent favors the products [103, 104].

Finally, the variation of $n_{aa}^{\dagger}$ with the force and the temperature (Fig.5.8) clearly pointed out that force and temperature induce a movement of TS towards the native state. However, while the effect of the temperature is mainly relevant at low forces, as confirmed by the behaviour of $dn_{aa}^{\dagger}/dT$ (Fig. 5.8c), on the contrary the force reveals to be in general a more efficient denaturant, capable to induce a more relevant movement of TS towards N. This result is not surprising, after all force regulates a wide variety of biological processes and playing a key role in the conformational changes experienced by many proteins in the cells.

Part IV

NUCLEIC ACIDS FOLDING

6

TEMPERATURE-DEPENDENT ELASTIC PROPERTIES OF DNA

MOTIVATION

The accurate knowledge of the elastic properties of single-stranded and double-stranded DNA is critical to characterize the thermodynamics of molecular reactions studied through single-molecule techniques. Here, we examine the temperature dependence of the elastic properties of single-stranded and double-stranded DNA molecules.

6.1 INTRODUCTION AND HISTORICAL CONTEXT

DNA is the biomolecule in charge of storing the genetic information of living organisms. In *in vivo* conditions, DNA is commonly found in its double-stranded (dsDNA) conformation forming the double helix that James Watson and Francis Crick discovered in 1953. However, DNA dissociates into its single-stranded (ssDNA) conformation in many biological processes, such as DNA replication, reparation, or transcription [105, 106].

Since the end of the XX century after the invention of single-molecule techniques, such as optical tweezers, the elastic properties of dsDNA and ssDNA molecules have focused the interest of researchers [107].

Moreover, in force spectroscopy experiments where mechanical forces are used to unfold DNA, a detailed characterization of the elastic properties is crucial to derive valuable information, i.e., the folding free energy, binding energies, or bending energies [108–110]. The elasticity of semi-flexible biopolymers, such as DNA, is commonly studied using the Worm-Like Chain (WLC) model [107, 111]. This model characterizes the elastic response using two parameters: L_p and d_b. L_p is the persistence length, which is the length over which correlations in the direction of the tangent of the polymer decays. d_b is the inter-phosphate distance, i.e., the distance between consecutive phosphates in the nucleotide chain.

Previous studies investigated L_p and d_b varying the length of the molecules at standard temperature, 25°C. First, the inter-phosphate distance reported from single-molecule experiments (SMEs) match with the reported ones from X-ray crystallography, $d_b \sim 6$ Å for ssDNA, and $d_b \sim 3$ Å for dsDNA, without important effects when varying the length of the polymer. Second, the persistence length exhibits different behavior when varying the length of the polymer for ssDNA and dsDNA. On the one hand, the persistence length of long ssDNA chains, more than 100 bases, are roughly equal to d_b, $L_p \sim 0.7$ nm, whereas for short chains, less than 100 bases, $L_p \sim 1.3$ nm [111, 112]. Finally, for dsDNA molecules, $L_p \sim 50$ nm for long chains (> 50 bp) and $L_p \sim 2$ nm for short chains (< 50 bp) [107, 113].

Here, we investigate the temperature dependence of the elastic properties of ssDNA and dsDNA molecules. To do so, we measure the force-dependent molecular extension from pulling experiments. We have investigated three DNA hairpins formed by 24, 32, and 44 bases, and a dsDNA molecule formed by 24 kbp in a wide temperature range $5 - 50$°C. The information of the elastic properties of such molecules would provide the information required to derive the folding entropy and enthalpy.

6.2 RESULTS

6.2.1 *Single-stranded DNA*

Three studied hairpins with different GC content have been used to investigate the ssDNA temperature-dependent elastic response and establish if there is a sequence dependence or not (sequences shown in Fig. 6.1). Moreover, the hairpins have been designed with 5′-Pyrimidine - Purine-3′ motifs to prevent stacking interaction between consecutive bases. The same hairpins are used to study the temperature-dependent folding energy, entropy, and enthalpy in Chap. 7.

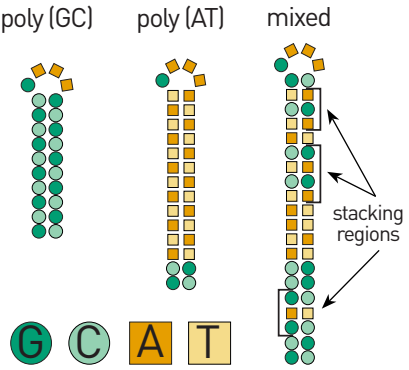

Figure 6.1.: **Hairpins' sequence.** Sequence of the three used hairpins, poly(GC), poly(AT), and mixed, to derive the elastic properties of the ssDNA.

We have carried out pulling experiments to determine the force- and temperature-dependent molecular extension of ssDNA, x_{ssDNA}, and subsequently derive its elastic properties. Briefly, in these experiments, the optical trap is repeatedly displaced away from and towards a reference point (in our case, the bead in the micro-pipette) to mechanically unfold/fold the DNA hairpins (Fig. 6.2). At low force, the three hairpins are in their native state (N) or dsDNA conformation, while at high force,

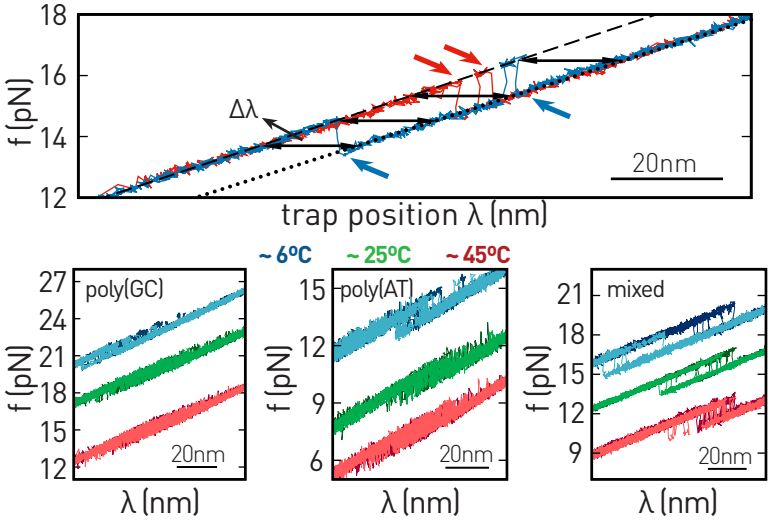

Figure 6.2.: **Pulling experiments at different temperature.** (a) Unfolding (red) and folding (blue) force-distance curves. The red and blue arrows point out the unfolding and folding events, respectively. The black dashed and dotted lines correspond to the N- and U- force branches. (b) Force-distance curves measured at room (green), high (red), and low (blue) temperatures for poly(GC) (left), poly(AT) (center), and mixed (right) hairpins. Dark (light) curves correspond to unfolding (folding) trajectories.

the hairpins are unfolded (U) in their ssDNA conformation. The unfolding ($N \rightarrow U$) and folding ($U \rightarrow N$) events are observed as sudden force rips in the force-distance curves, FDCs, (red and blue arrows in Fig. 6.2-top). In the FDCs, we define two force branches: when the molecule is in N, N-branch, and when the molecule is in U, U-branch. Therefore, by measuring the difference in trap position at fixed force values $\Delta\lambda(f)$, we derive the difference in molecular extension between the ssDNA and dsDNA conformations of the hairpins (Fig. 6.2-top),

$$\Delta\lambda(f) = x_{ssDNA}(f) - x_d(f) \; . \tag{6.1}$$

To extract x_{ssDNA} from Eq.6.1, we have modeled the extension of the native (dsDNA conformation) state, x_d, as the extension of a rigid dipole projected along the force axis. The extension x_d has been determined using the Freely-Jointed Chain model assuming that the Kühn and contour lengths are equal to 2 nm and constant with temperature. Moreover, by measuring $\Delta\lambda$ at different values of f between the last folding and unfolding events, we reconstructed the force-extension curve of the ssDNA. Taking profit from the fact that each hairpin unfolds/refolds at different force regimes (Fig. 6.2-bottom), we have demonstrated that the elasticity of ssDNA does not present effects due to the sequence. Commonly, the force versus x_{ssDNA} curves are known as force-extension curves (FECs).

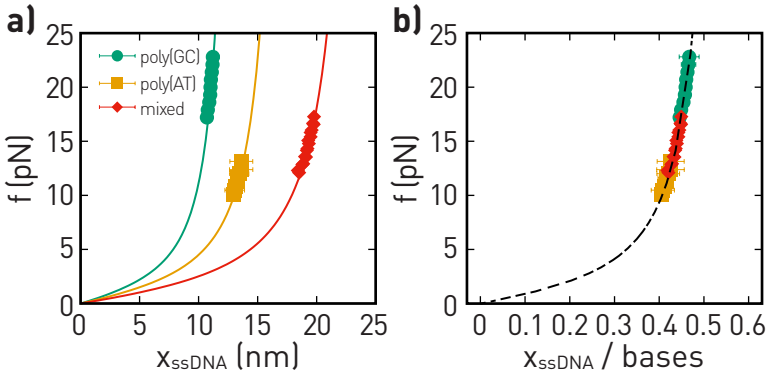

Figure 6.3.: **Force-extension curves of ssDNA.** (a) Molecular extension at 25ºC of poly(GC) (green), poly(AT) (yellow), and mixed (red) hairpins. Solid line is the expected behavior using WLC model with the elastic properties reported in the literature [111]. (b) Normalized molecular extension to demonstrate that the ssDNA does not present sequence effects.

Figure 6.3a shows the measured $x_{ssDNA}(f)$ at 25ºC for the three DNA hairpins: poly(GC) (green), poly(AT) (yellow), and mixed (red). The solid lines are the expected behavior considering the elastic response reported in the literature [111]. As expected from the FDCs, each hairpin provides information on the elastic response of the ssDNA at different force regions. In addition, we have normalized the extension x_{ssDNA} by

the total number of bases to demonstrate that the elasticity of ssDNA does not present sequence effects (Fig. 6.3b).

To determine the temperature-dependent elastic properties of ssDNA, we have adjusted x_{ssDNA} to the inextensible Worm-Like Chain (WLC) model and its interpolation formula [82],

$$f = \frac{k_B T}{4L_p}\left(\left(1-\frac{x}{Nd_b}\right)^{-2} + 4\frac{x}{Nd_b} - 1\right) \tag{6.2}$$

being f the measured force, x the molecular extension, N the number of bases, k_B the Boltzmann constant, and T the temperature. The temperature-dependent molecular extension of the ssDNA has been directly determined from the FDCs recorded at every temperature (Fig. 6.2-bottom) using Eq. (6.1). We have determined the temperature dependence of L_p and d_b by fitting the measured molecular extension as a function of force to Eq. (6.2) keeping L_p and d_b as free parameters.

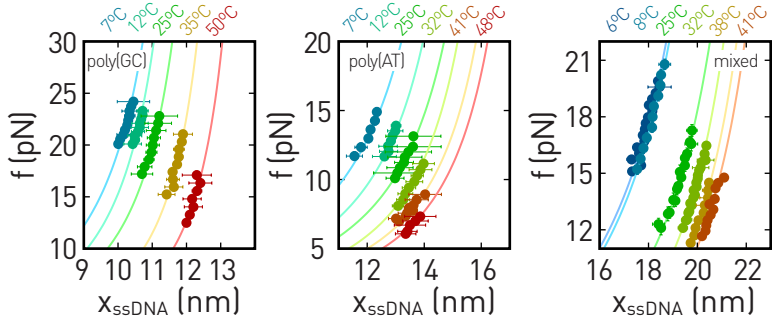

Figure 6.4.: **Temperature-dependent FECs.** Experimental values of x_{ssDNA} for the poly(GC) (left), poly(AT) (center), and mixed (right) hairpins at all the studied temperatures. The solid lines are the fits to Eq. 6.2 with L_p and d_b as free parameters.

Figure 6.4 shows the measured $x_{ssDNA}(f)$ (symbols) together with the fits to Eq. 6.2 (solid lines) leaving L_p and d_b as free parameters for the three hairpins at all the studied temperatures. Note that for a fixed

value of force, the extension of the ssDNA becomes larger as we increase the temperature. This fact indicates that if we assume that the distance between bases d_b does not vary with temperature, the elastic response of the ssDNA becomes stiffer as we increase the temperature.

We have found that L_p has a strong temperature dependence independently of the sequence (Fig. 6.5-bottom), whereas the inter-phosphate distance has a tiny temperature dependence (Fig. 6.5-top). Both parameters, d_b and L_p have been adjusted with a linear trend with slopes 0.016 ± 0.002 Å/K and 0.28 ± 0.01 Å/K, respectively. Comparing the slopes, we can see that $\partial L_p/\partial T$ is ten times larger than $\partial d_b/\partial T$, indicating that the inter-phosphate distance is roughly constant with T. Moreover, it has been reported that L_p has an electrostatic contribution that depends on the temperature-dependent screening Debye length, which is proportional to $T^{1/2}$ [114]. For the studied temperature range, we did not find a significant difference between a linear fit and a fit using $L_p(T) = aT^{1/2} + b$.

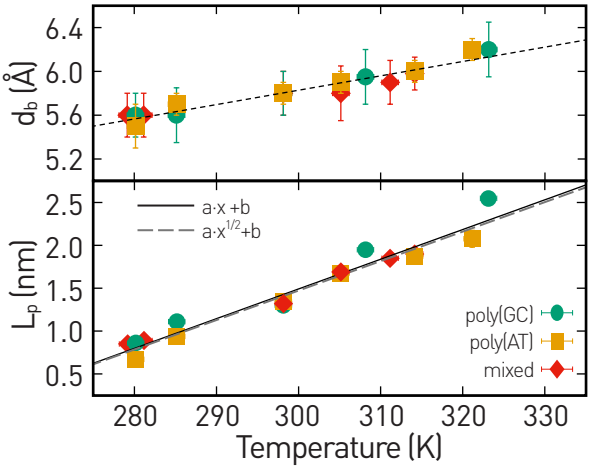

Figure 6.5.: **Elastic parameters.** Inter-phosphate distance (top) and persistence length (bottom) measured using the poly(GC) (green), poly(AT) (yellow), and mixed (red) hairpins.

6.2.2 *Double-stranded DNA*

In contrast with pulling experiments using DNA hairpins, in pulling experiments using dsDNA molecules, the force is exerted in the 3'- ends of each complementary ssDNA strand (Fig. 6.6). Hence, force does not induce a mechanical melting of the dsDNA but stretches it from its ends in this configuration. To study the elastic properties of dsDNA, we have carried pulling experiments using a segment of 24805 bp from the DNA of the bacterial virus *enterobacteria phage* λ. To obtain this specific segment, we utilized two specific restriction enzymes that cut the dsDNA in the desired recognition sites.

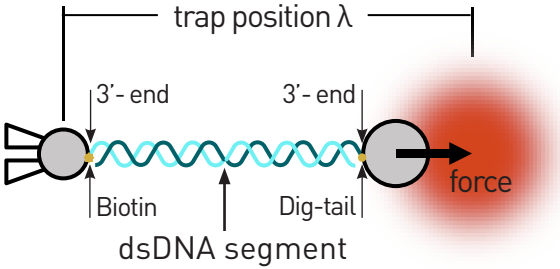

Figure 6.6.: **Experimental setup.** The dsDNA segment is tethered between two dielectric. The connection between the dsDNA molecule and the beads is done by specific interactions: Biotin - streptavidin and Digoxigenin - anti-digoxigenin.

In pulling experiments using dsDNA molecules, the molecule is repeatedly stretched and relaxed between a minimum force (typically ~ 1 pN) and a maximum force ($50 - 60$ pN). The measured FDCs present three different regimes. In the first regime, $f < 4$ pN, an entropic contribution dominates the elastic response of the dsDNA. At these forces, the randomly coiled dsDNA starts to be stretched, losing degrees of freedom. The second regime, $4 < f < 60$ pN, is dominated by the enthalpic contribution to stretch the hydrogen bonds of the dsDNA. In this regime, the FDCs are way steep, and small changes in the molecular extension

are translated into large changes in the measured force. The third regime is defined at high forces, $f \sim 67$ pN at 25ºC, where a force plateau is observed. This plateau corresponds to the overstretching transition of DNA [107].

In Figure 6.7a, we show a single stretching trajectory measured at different temperatures, 9 (dark blue), 15 (light blue), 25 (green), 38 (brown), and 45ºC (red). Notice that the entropic regime and the overstretching plateau are recorded for all the studied temperatures. Whereas the entropic regime is nearly temperature independent, the temperature strongly affects the overstretching transition. Figure 6.7b shows the average force along the plateau as a function of temperature. This force is approximately linear with the temperature. Finally, we can see that the elastic response between $4 - 40$ pN becomes less rigid with T, the slope of this regime is more pronounced at low temperatures.

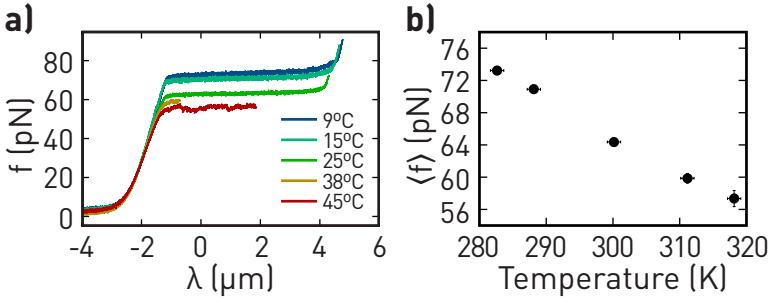

Figure 6.7.: **Stretching dsDNA.** (a) FDCs obtained at different temperatures using a 24kbp dsDNA molecule. (b) Average force at the measured overstretching plateau as a function of temperature.

To derive the elastic properties of dsDNA, we used the WLC model, Eq. (6.2). Nevertheless, at this point, we do not have a direct measurement of the force-dependent molecular extension of the dsDNA molecule, $x_{dsDNA}(f)$, in contrast to the previous case. To derive such

molecular extension, we have subtracted the bead displacement from the measured distance λ,

$$\lambda(f) = x_{dsDNA}(f) + \frac{f}{k_{trap}} + \lambda_0 \,, \tag{6.3}$$

where λ_0 is an arbitrary origin to the λ measurement, and k_{trap} is the stiffness of the optical trap.

First, we have derived the value of k_{trap} from the FDCs of the DNA hairpins. To do so, we have used a variant of Eq. 6.3,

$$\lambda(f) = x_{ssDNA}(f) + \frac{f}{k_{trap}} + x_{handle}(f) + \lambda_0 \,, \tag{6.4}$$

where the elastic response of the DNA handles ($x_{handles}(f)$) was derived using the elastic properties reported in the literature [113], and the x_{ssDNA} were derived as explained before. Using these parameters, we have fitted the *U*-force branch to find the value of k_{trap} that best matches Eq. (6.4) with the experiments. The resulting force-extension curves for dsDNA are shown in Fig. 6.8 together with the fits to the WLC model. Equation (6.2) have been fitted in a force range between 0 and 20 pN.

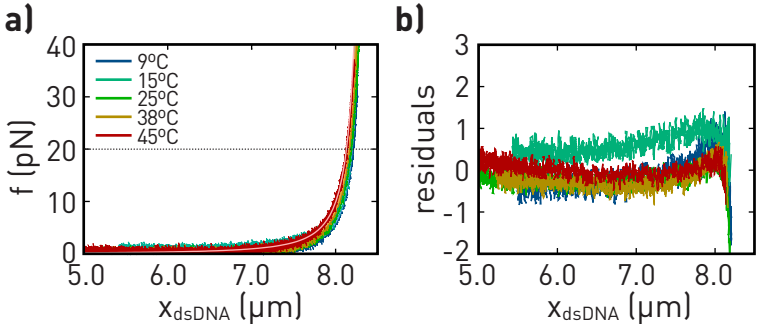

Figure 6.8.: **Force-dependent extension of dsDNA.** (a) Measured FECs at 9 (dark blue), 15 (light blue), 25 (green), 38 (brown), and 45ºC (red). The light lines correspond to fits to the WLC model. (b) Residuals defined as the difference between the measured force and the fitted one as a function of x_{dsDNA}.

The derived L_p and d_b are shown in Fig. 6.9. On the one hand, d_b is practically constant in temperature with an average inter-phosphate distance of 0.34 nm, which agrees with X-ray crystallographic measurements. On the other hand, L_p is strongly affected by the temperature. Like the ssDNA, the temperature response of L_p is well fitted with a linear function or a function like $aT^{1/2} + b$. However, the slope that best fits with the experimental data is negative and equal to -0.54 ± 0.05 nm/K. Contrary to ssDNA, where L_p increases with temperature, the dsDNA becomes softer when the temperature increases.

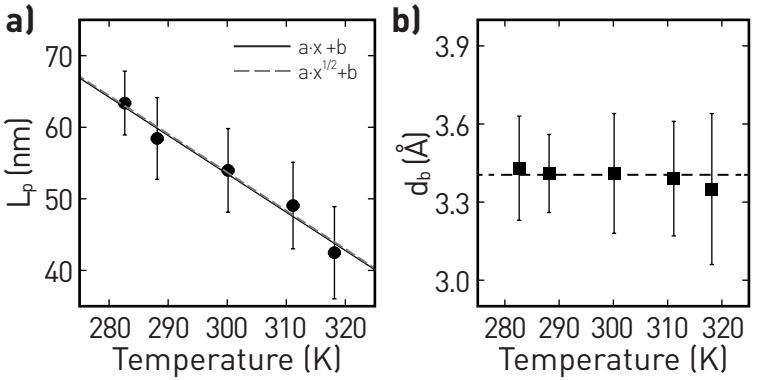

Figure 6.9.: **Elastic properties of dsDNA.** (a) Measured persistence length (symbols) together with a linear fit (black solid line) and a fit to $aT^{1/2} + b$ (gray dashed line). (b) Measured inter-phosphate distance. The dashed line denotes the average value.

6.3 CONCLUSIONS

Here, we have investigated the elastic response of ssDNA and dsDNA molecules at different temperatures using pulling and stretching experiments. We have unfolded and folded DNA hairpins to measure the molecular extension of ssDNA molecules. DNA hairpins are ssDNA

chains that fold into themselves, forming structures like a hairpin, i.e., a dsDNA stem ended with an ssDNA loop. To measure the force-dependent molecular extension of dsDNA molecules, we have stretched and relaxed a 24kbp DNA segment of the bacterial-virus λ-phage. To characterize the elastic response of both molecules, we have used the inextensible Worm-Like Chain (WLC) model. This model describes the elasticity of semi-flexible biopolymers using two parameters: the persistence length L_p, which gives information of the correlations along the tangential direction of the polymer, and the inter-phosphate distance d_b.

We have found that the persistence length of ssDNA and dsDNA is strongly affected by T in the explored temperature range, $5 - 50^{o}$C. Interestingly, whereas ssDNA becomes rigid (L_p larger) at high temperatures, dsDNA becomes softer (L_p smaller). The temperature response of L_p is well fitted with a $T^{1/2}$ behaviour where the slope of dsDNA is ten times bigger than the one of ssDNA. It means that the dsDNA is more affected by the temperature in the same temperature range. In addition, previous pulling experiments proved that L_p for ssDNA follows the expected behavior of $L_p \sim 1/[Ion]^{1/2}$ considering the Debye-Hückel theory [115]. Also, unzipping experiments have changed the ionic strength to simulate different temperature conditions to derive the temperature dependence of DNA folding energies [116]. The persistence length of a semi-flexible polymer, such as ssDNA or dsDNA, has an intrinsic persistence length and an electrostatic contribution that depends on the screening Debye length (λ_D):

$$\lambda_D = \left(\frac{\epsilon_r \epsilon_0 k_B T}{\sum_{j=0}^{N} n_j^0 q_j^2} \right)^{1/2} . \tag{6.5}$$

In Equation (6.5), N denotes the different species of charges, q_j is the charge of the jth species with concentration n_j. ϵ_r and ϵ_0 are the relative static permittivity and dielectric constant, respectively. In the explored temperature range, we have found that L_p is well fitted with a linear or the $aT^{1/2} + b$ function. A discrepancy of only $\sim 10\%$ between both functions is found close to 100^{o}C, the temperature where water boils.

Below this temperature, we do not appreciate any difference between both functions.

Finally, the inter-phosphate is slightly affected by the temperature for both DNA molecules in the explored temperature range. On the one hand, the inter-phosphate distance of the ssDNA chain increases linearly with T with a slope ten times smaller than the slope for L_p. Therefore, we can assume that d_b is constant with T with an average value equal to 0.58 ± 0.01 nm. On the other hand, d_b for the dsDNA molecule is practically constant with T with an average value of 0.34 ± 0.01 nm, which agrees with the inter-phosphate distance measured from X-ray crystallographic experiments.

7

SEQUENCE EFFECTS ON DNA HEAT CAPACITY CHANGE

MOTIVATION

The folding thermodynamics of biopolymers is a notoriously difficult topic. This problem is made especially complex by the presence of several actors: solvent, counterions and solutes that establish interactions of similar order of magnitude. This is why, still nowadays, it is important to push the resolution of measurements on the process by which two complementary DNA strands meet and form the well known double helix. Here, we investigate the folding process of DNA using calorimetric single-molecule experiments.

7.1 INTRODUCTION AND HISTORICAL CONTEXT

Two of the most relevant techniques in the study of DNA folding are differential scanning calorimetry [117] and isothermal titration calorimetry [118] (DSC and ITC). These techniques measure the melting temperature and the enthalpy difference between the native, N, and the denatured or unfolded conformation, U, at the melting temperature. Initial results based on DSC suggested that the dissociation enthalpy difference does not have a temperature dependence, i.e., no heat capacity change

(ΔC_p) between dsDNA and ssDNA was detected [119–121]. More recent studies combining ITC and DSC have shown that the melting process measured at low (ITC) and high (DSC) temperatures do not match each other. These results prove the fact that the formation enthalpy and entropy must depend on temperature. The temperature dependency of the enthalpy ($\Delta H_0 = H_0^U - H_0^N$) and entropy ($\Delta S_0 = S_0^U - S_0^N$) difference equals,

$$\Delta H_0(T) = \Delta H_0^m + \Delta C_p\,(T - T_m) \tag{7.1a}$$

$$\Delta S_0(T) = \Delta S_0^m + \Delta C_p \log\left(\frac{T}{T_m}\right) \tag{7.1b}$$

where $\Delta C_p = C_p^N - C_p^U$, denotes the heat capacity change between N and U, and ΔH_0^m (ΔS_0^m) is the enthalpy (entropy) change at the melting temperature (T_m). Depending on the sequence and length of the molecules employed, ΔC_p values range between $40 - 160$ cal/mol·K per base-pair (bp) [122–126]. Moreover, although entropy and enthalpy values at a given temperature are different for a GC or an AT base-pair, recent studies did not resolve the sequence dependence for heat capacity changes [125, 126].

Nowadays, Single-Molecule Experiments (SMEs) study folding and binding energies with an unprecedented resolution. In such experiments, force is used to modulate the relative stability of the folded and unfolded state. Force becomes a novel thermodynamic parameter to obtain a richer picture of the folding transition. Two advantages of SMEs are especially relevant in studying DNA folding thermodynamics. On the one hand, we can measure the force-dependent folding kinetics with high accuracy. This allows us to characterize the relative stability between N and U and the activation energy and the position of the transition state to map the molecular Free-Energy Landscape (mFEL).

On the other hand, we can explore folding thermodynamics on the two-dimensional plane defined by force and temperature, two independent parameters that we can control precisely. In this plane, we focus on the particular line, the coexistence line, where the probability of find-

ing the molecule in N or U is 50%. At coexistence, we can use the Clausius-Clapeyron equation to directly estimate the folding entropy at the corresponding coexistence force and temperature with tight control on experimental error (mathematical details in App. E). However, to compare results obtained at a given force with the zero-force calorimetric studies, we must subtract the elastic contributions of the different elements involved in the SMEs.

Here, we investigate three short DNA hairpins with different GC content, i.e., GC-rich, AT-rich, and mixed, to determine the free energy, entropy, enthalpy, and heat capacity change of a single GC (Guanine paired with a Cytosine) and AT (Adenine pared with a Thymine) base-pair. To do so, we carry out hopping experiments measuring the folding and unfolding kinetic rates in a wide temperature range, 5-50^{o}C. Contrary to protein barnase, where hopping experiments are unfeasible due to the high dissipation towards mechanical unfolding, we determine the coexistence force and folding free energy directly from equilibrium hopping experiments. To use the Clausius-Clapeyron like-equation, we use the elastic properties of ssDNA derived in Chap. 6. Moreover, we investigate two hairpins with the same stem sequence ($\sim$ 50 % GC content) with different loop lengths to study their impact on the folding free energy, entropy, and enthalpy.

7.2 RESULTS

The sequences of the three DNA hairpins used to derive the energy, entropy, and enthalpy per bp are shown in Fig. 7.1. Mention that these sequences are the same ones used in Chap. 6 to derive the elastic properties of ssDNA molecules. The first hairpin, which we named *poly(GC)*, has been designed with a stem formed by only five 5'-GC-3' dinucleotide steps (the stem has a total of 10 bp). We chose this sequence to prevent stacking interactions between first neighbors when the hairpin is folded. Poly(GC) hairpin allows us to determine the folding free energy of a single GC bp and its entropy and enthalpy difference. The second

hairpin, *poly(AT)*, has a stem formed by six 5′-AT-3′ dinucleotide steps. In addition, poly(AT) has a 5′-GC-3′ motif preceding the twelve AT bp to prevent breathing or thermal hopping at the beginning of the stem. Therefore, poly(AT) is formed by 14bp in the stem. As for poly(GC), poly(AT) has a purine followed by a pyrimidine to prevent stacking interactions between consecutive bases. For the poly(AT) hairpin, we derive the energy, entropy, enthalpy, and ΔC_p of an AT bp. Finally, the last studied hairpin, named *mixed*, has a stem formed by 20 bp with a 50% GC content. We use the last hairpin to test the validity of the results obtained with the other two hairpins. The mixed sequence, however, contains three non-nearest neighbour motifs containing 3 and 4 consecutive purines along the single strand DNA chain (GAG, AGA, and GAGA highlighted in Fig. 7.1). These trinulceotides and tetranucleotides motifs are expected to further stabilize the double helix beyond what is predicted by the nearest neighbour model. Therefore, the energy per bp derived by averaging results obtain from the poly(GC) and poly(AT) hairpins would set a lower bound to the energy per bp obtained in the mixed hairpin. The difference in energy would be related to the trinucleotide and tetranucleotide purine stacks. The total free energy, entropy, and enthalpy difference between the initial and final state, i.e., N and U, is the sum over the base-pairs forming the stem. The three hairpins have a tetra-loop equal to 5′-GAAA-3′.

7.2.1 *Unfolding/Folding kinetic rates*

We carried out passive hopping experiments (see Chap. 2, Sec. 2.2 for experimental details) at different trap positions in a wide temperature range, from 5°C to 50°C, to study the unfolding and folding force-dependent kinetic rates of the three DNA hairpins. In Figure 7.2, we show the first 5 seconds of a force-time trace measured at 6, 25, and 45°C for the three hairpins (poly(GC) in the top, poly(AT) in the middle, and mixed in the bottom). On the right, we show the force distribution of each force-time trace considering the entire trajectory. Notice that the

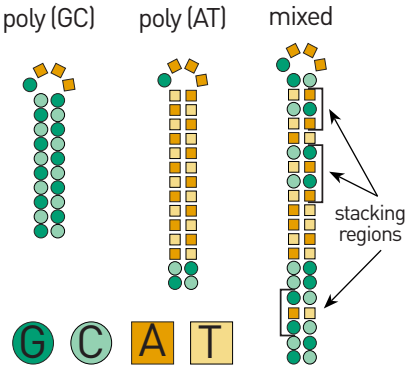

Figure 7.1.: **Hairpins' sequence.** Poly(GC) has only GC bps (left), Poly(AT) has two GC bp and twelve AT bp (center), and mixed has 50% GC content.

number of transitions per unit of time at a given temperature increases when reducing the hairpin length. Moreover, the number of hopping events increases when heating the media surrounding the molecules. In addition, the average unfolding and folding forces, i.e., the measured force when the hairpins are in N and U, respectively, are reduced when we increase the temperature.

To determine the unfolding ($k_{\rightarrow}$) and folding ($k_{\leftarrow}$) kinetic rates from the force-time traces, we calculated the average residence time or lifetime $\langle \tau \rangle$ of each state,

$$k_{\rightarrow} = 1/\langle \tau_N \rangle \;\; ; \;\; k_{\leftarrow} = 1/\langle \tau_U \rangle \tag{7.2}$$

Figure 7.3 shows the measured unfolding (solid symbols) and folding (empty symbols) kinetic rates for the poly(GC) (top), poly(AT) (middle), and mixed (bottom) hairpins. At first sight, we notice that the hairpin poly(GC) is the hairpin that folds/unfolds at higher forces. In comparison, the hairpin poly(AT) is the one that folds/unfolds at lower forces, and the mixed hairpin folds and unfolds in-between. This result indicates that the required energy to unfold the poly(GC) hairpin is higher than that required to unfold the poly(AT) hairpin. This agrees with the

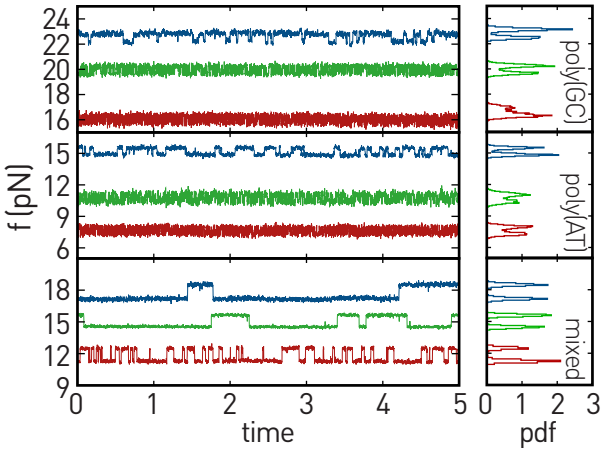

Figure 7.2.: **Passive hopping experiments at different temperatures.** Force-time traces measured at 6 (blue), 25 (green), and 45°C (red) for the poly(GC) (top), poly(AT) (middle), and mixed (bottom) hairpins. In the right is shown the force distribution of each force-time trace.

fact that a GC bp has an extra hydrogen bond relative to an AT bp. Moreover, as mentioned before, we can see that the force required to unfold/fold the hairpins is shifted at lower values when the temperature increases. Finally, from Fig. 7.3, we notice that the kinetic rate at coexistence where $k_{\rightarrow}(f)$ and $k_{\leftarrow}(f)$ cross increases upon increasing T.

To investigate the mFEL and derive the folding free energy from the measured kinetic rates, we use the two-state Bell-Evans (BE) model [85, 86]. This model describes the mechanically induced unfolding (folding) transitions as thermally activated processes over a transition state energy barrier $\Delta G_0^{\dagger}$ (ΔG_0^{*}) relative to N (U). The energy of the kinetic barrier is defined as $\Delta G_0^{\dagger} = G_0^{TS} - G_0^{N}$ ($\Delta G_0^{*} = G_0^{TS} - G_0^{U}$). The BE model assumes that, for a fixed transition state (TS) position, the height of the kinetic barrier to unfold (fold) decreases linearly with the applied force, $\Delta G^{\dagger} = \Delta G_0^{\dagger} - fx^{\dagger}$ ($\Delta G^{*} = \Delta G_0^{*} - fx^{*}$), being $x^{\dagger}$ (x^{*}) the distance from the native (unfolded) state to the transition state. The following

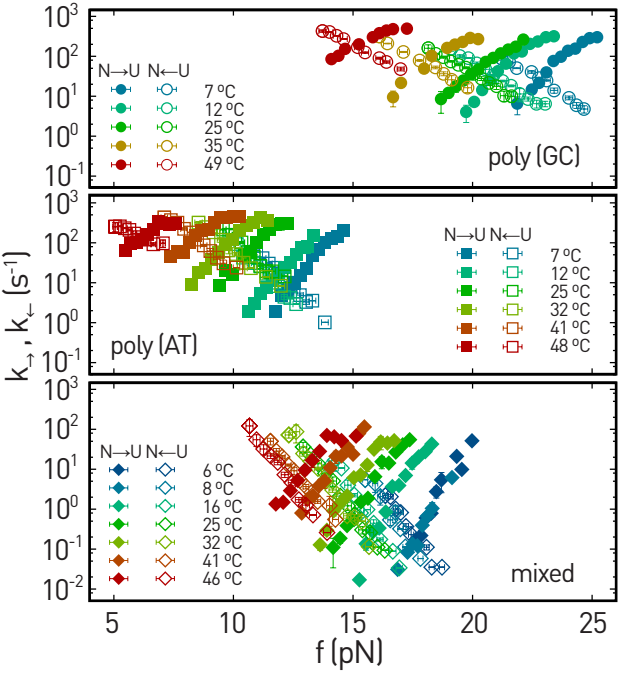

Figure 7.3.: **Force- and temperature-dependent kinetic rates.** Unfolding ($k_{\rightarrow}$, solid symbols) and folding ($k_{\leftarrow}$, empty symbols) for the poly(GC) (top), poly(AT) (middle), and mixed (bottom) hairpins.

equations describe the force-dependent unfolding and folding kinetic rates,

$$k_{\rightarrow}(f) = k_a \exp\left(-\beta \Delta G_0^{\dagger}\right) \cdot \exp\left(\beta f x^{\dagger}\right) \tag{7.3a}$$

$$k_{\leftarrow}(f) = k_a \exp\left(-\beta \Delta G_0^{*}\right) \cdot \exp\left(-\beta f x^{*}\right) \tag{7.3b}$$

being $\beta = k_B T$, k_B the Boltzmann constant, T the temperature, and k_a the attempt rate of the system.

According to Eqs. (7.3a) and (7.3b), the kinetic rates $k_{\rightarrow}$ and $k_{\leftarrow}$ satisfy the detailed balance condition,

$$\log\left(\frac{k_{\rightarrow}}{k_{\leftarrow}}\right) = \beta(f x_m - \Delta G_0) \tag{7.4}$$

where $\Delta G_0 = \Delta G_0^{\dagger} - \Delta G_0^{*} = G_0^{U} - G_0^{N}$ is the free energy difference between N and U, and $x_m = x^{\dagger} + x^{*}$ the extension between N and U, which equals the extension of U minus the extension of N, i.e., $x_m = x_U - x_N$.

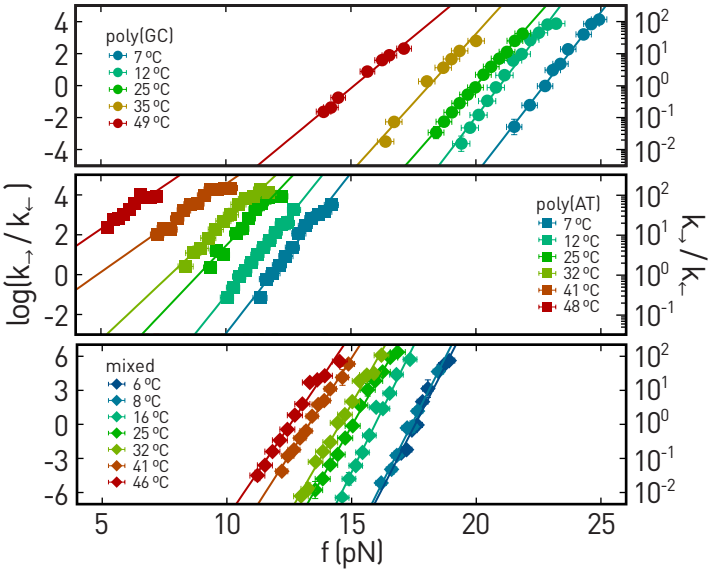

Figure 7.4.: **Detailed balance condition.** Measured values of $\log(k_{\rightarrow}/k_{\leftarrow})$ for the poly(GC) (top), poly(AT) (middle), and mixed (bottom) hairpins (symbols) together with fits to Eq. (7.4) (solid lines).

Figure 7.4 shows the measured $k_{\rightarrow}/k_{\leftarrow}$ in a log-normal plot (right Y-axis) and normal-normal plot (left Y-axis) for the three hairpins at all the studied temperatures as a function of the *apparent force*. The apparent force is defined as the average force considering the elastic response of the system when the hairpin is in N and U. This force is helpful to

determine the $k_{\rightarrow}/k_{\leftarrow}$ at a given value of force [113]. As can be seen, the behavior of $\log(k_{\rightarrow}/k_{\leftarrow})$ is approximately linear in the explored force range, which agrees with Eq. (7.4). The solid lines in Fig. 7.4 are fits to Eq. (7.4).

7.2.2 *Coexistence condition*

We have derived the coexistence force f_c following two analogous approaches. First, we determined the force where $k_{\rightarrow}$ crosses $k_{\leftarrow}$, and second, we determined the force at which $\log(k_{\rightarrow}/k_{\leftarrow}) = 0$. Figure 7.5a shows the average f_c as a function of temperature. As expected for a thermally activated process, f_c decreases when the temperature increases. The coexistence force for the three molecules defines the line where the probability of being in N and U is equal. It is well adjusted by a linear trend (dashed lines in Fig. 7.5a). Second, we have determined the coexistence kinetic rate ($k_{\rightarrow}(f_c) = k_{\leftarrow}(f_c) := k_c$). We have found that $k_c \propto \exp(1/T)$ as expected from Eqs. (7.3a) and (7.3b) (Fig. 7.5b).

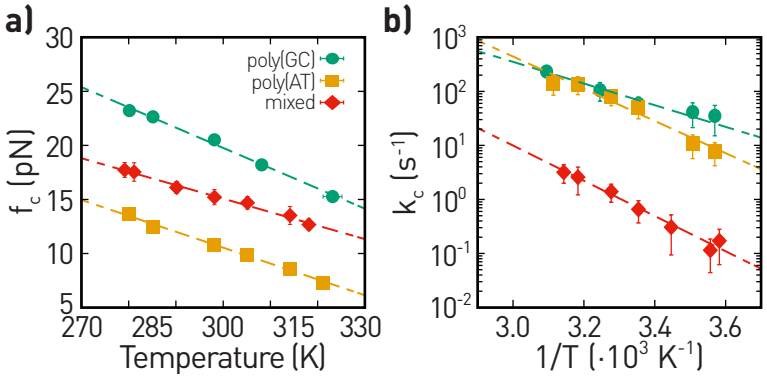

Figure 7.5.: **Coexistence condition.** Temperature-dependent coexistence force (a) and kinetic rate (b). Dashed lines in panels (a) and (b) are linear and exponential fits, respectively.

7.2.3 *Transition state's position*

The position of the transition state relative to N, $x^\dagger$, and U, x^*, have been determined from the slopes of the $\log(k_\rightarrow(f))$ and $\log(k_\leftarrow(f))$ close to the coexistence point (f_c, k_c). According to Eqs. (7.3a) and (7.3b), $x^\dagger + x^*$ equals the extension difference between N and U at coexistence force, which we denote as x_m. Figure 7.6a shows x_m normalized by the total number of bases to compare the three DNA hairpins (24 bases for the poly(GC), 32 bases for the poly(AT), and 44 bases for the mixed hairpin).

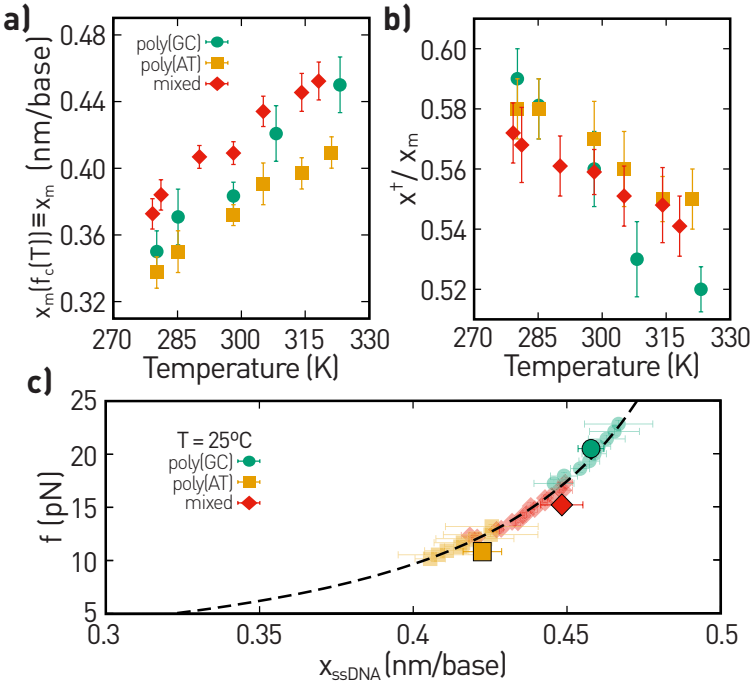

Figure 7.6.: **Molecular extension at the coexistence force $f_c(T)$.** (a) Extension difference between N and U at $f_c(T)$ normalized by the number of bases. (b) Transition-state distance normalized by x_m. (c) Extension per base measured in Chap. 6 (semitransparent symbols) and extension measured from the kinetic rates (solid symbols) at 25ºC.

According to the Leffler-Hammond postulate [103, 104], the number of released base pairs at the transition state at a given force is reduced as we increase the temperature. Here, we have determined the fraction of unfolded bp at the transition state by dividing $x^\dagger$ by the difference in extension between N and U, i.e., x_m, both extensions evaluated at the coexistence force. Figure 7.6b shows the fraction of unfolded base pairs at TS relative to N as a function of temperature. As expected, at high temperatures, the fraction is smaller, in agreement with the Leffler-

Hammond postulate. Finally, in Fig. 7.6c, we compare the molecular extension per base at 25ºC derived in Chap. 6, with the ones using the Bell-Evans model. According to Fig. 7.6c, the extension x_m per base for the mixed hairpin would be between the one of poly(GC) and poly(AT). However, in Fig. 7.6a, the extension x_m for the mixed hairpin is larger than the other two. This result arises from the fact that the dipole contribution for the mixed hairpin represents roughly the 8% of the full extension. In contrast, this contribution represents the 20% and 14% for the poly(GC) and poly(AT), respectively. This difference in percentages arises from the different number of base pairs of the hairpins (20bp for the mixed, 10bp for the poly(GC), and 14bp for the poly(AT).

7.2.4 *Folding free energy*

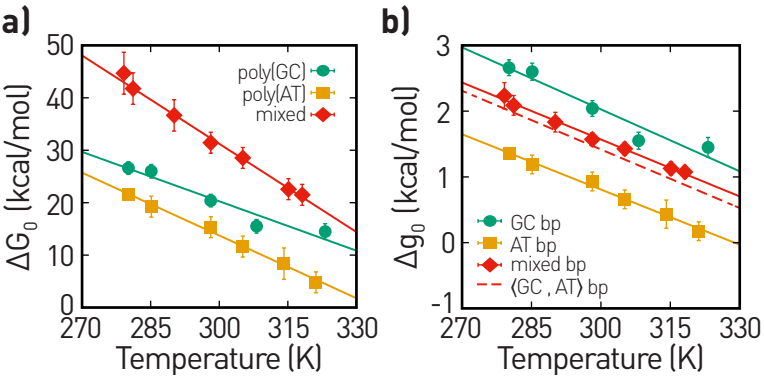

Figure 7.7.: **Temperature-dependent folding free energy.** (a) Total folding free energy for the three studied hairpins, poly(GC) (green), poly(AT) (yellow), and mixed (red). (b) Folding free energy per bp derived from the experiments: GC bp (green circles), AT bp (yellow squares), and mixed (red diamonds). The dashed line corresponds to the average value considering the experimental results per GC and AT bp.

We have derived the folding free energy ΔG_0 as a function of temperature using two different approaches: from the fits to Eqs. (7.3a) and (7.3b), and from the fit to Eq. (7.4). In figure 7.7a, we show the derived folding free energy as a function of temperature. As expected from thermodynamics, ΔG_0 is strongly affected by T. In fact, from Gibbs energy definition we know,

$$\Delta G_0(T) = \Delta H_0 - T\Delta S_0 \tag{7.5}$$

where ΔH_0 and ΔS_0 are the enthalpy and entropy difference between N and U, respectively. Therefore, by adjusting Eq. (7.5) to the experimental values, we have derived ΔH_0 and ΔS_0, as well as the melting temperature T_m, which is defined as $T_m = \Delta H_0/\Delta S_0$. The obtained values are shown in Tab. 7.1-top rows.

Finally, from the measured ΔG_0 values, we have extracted the folding energy of a GC and AT bp using the poly(GC) and poly(AT) hairpins and the average folding energy per bp from the mixed hairpin. We have divided the total energy of the poly(GC) hairpin by the ten bp forming the stem of the hairpin to derive the folding energy of a single GC bp. The obtained results are shown as green circles in Fig. 7.7b. We have subtracted the energy of the two GC bp in poly(AT) and divided the rest by the twelve AT bp to derive the average energy of a single AT bp. The resulting values are shown as yellow squares in Fig. 7.7b. Finally, the average energy per bp has been obtained by dividing the total energy of the mixed hairpin by the 20 bp forming the stem (red diamonds in Fig. 7.7b). As before, we have fitted Eq. 7.5 to the experimental values to derive the entropy, enthalpy, and melting temperature of a single bp. The results are summarized in Tab. 7.1-bottom rows. In addition, using the enthalpy and entropy per GC and AT bp, Δh_0 and Δs_0, we have calculated the expected average energy per bp. As can be seen in Fig. 7.7b, the average energy per bp using the derived energies for a single GC and AT bp (dashed red line) matches the observed behavior obtained from the mixed hairpin.

The enthalpy and entropy in Tab. 7.1 are measured in the assumption of zero heat capacity change, i.e., H and S do not depend on T. Therefore, the values in the tables would match the values measured at the melting temperature. According to Eqs. (7.1a) and (7.1b), the temperature dependence of ΔG_0 is defined as

$$\Delta G_0(T) = \Delta H_0^m - T\Delta S_0^m + T\Delta C_p \left[1 - \log\left(\frac{T}{T_m}\right)\right] - T_m \Delta C_p \,. \quad (7.6)$$

Commonly, equation (7.6) is known as the stability curve of a protein or DNA/RNA molecule. According to Eq. (7.6), the measured energy per base pair should exhibit a negative curvature in the explored temperature range due to the $\Delta C_p > 0$. Nevertheless, the energy per base pair shown in Fig. 7.7b does not have the expected curvature. Therefore, to derive the heat capacity change for each bp type, we derived the temperature-dependent entropy and enthalpy.

	ΔH_0 (kcal/mol)	ΔS_0 (cal/mol·K)	T_m (°C)
poly(GC)	115 ± 14	314 ± 48	93 ± 5
poly(AT)	133 ± 5	400 ± 16	59 ± 2
mixed	200 ± 7	560 ± 21	82 ± 2
GC bp	12 ± 1	31 ± 5	93 ± 5
AT bp	9.2 ± 0.4	28 ± 1	56 ± 2
mixed bp	10.2 ± 0.3	29 ± 1	82 ± 2

Table 7.1.: Derived enthalpy, entropy, and melting temperature

7.2.5 *Temperature-dependent entropy and enthalpy change*

We have determined the temperature-dependent entropy change across unfolding/folding using the Clausius-Clapeyron like-equation derived in App. E,

$$\Delta S_0(T) = -\frac{\partial f_c(T)}{\partial T} x_m(f_c(T)) - \int_0^{f_c(T)} \frac{\partial x_m(f', T)}{\partial T} df' \quad . \tag{7.7}$$

In equation (7.7), x_m is the molecular extension difference between N and U. The value of $x_m(f_c(T))$ (Fig. 7.6a) has been determined from the kinetic rates and the Bell-Evans model, while the force and temperature dependence of x_m in the integral of the rhs of Eq. (7.7) has been derived using the elastic parameters of ssDNA presented in Chap. 6. The entropy and entropy per bp measured using Eq. (7.7) are shown in Fig. 7.8. The entropy per GC and AT bp has been estimated following the same procedure as the one used to derive the folding free energy per bp.

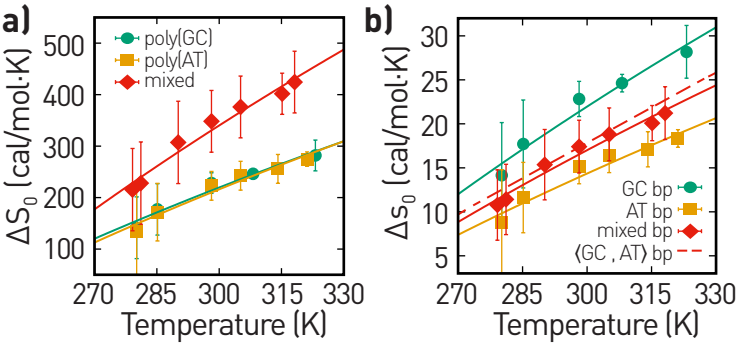

Figure 7.8.: **Temperature-dependent entropy change.** (a) Total entropy change for the three studied hairpins, poly(GC) (green), poly(AT) (yellow), and mixed (red). (b) Entropy difference per bp derived from the experiments: GC bp (green circles), AT bp (yellow squares), and mixed (red diamonds). The dashed line corresponds to the average value considering the experimental results per GC and AT bp. The solid lines in panel (a) and (b) are fits to Eq. (7.1b).

The heat capacity change across folding has been derived using Eq. (7.1b), which is an expansion of the temperature dependence of ΔS_0 close to the melting temperature T_m. Equation 7.1b is also used to derive the entropy difference per bp. The heat capacity change and Δs_0^m per bp are summarized in the Tab. 7.2.

	Δs_0 (cal/mol·K)	ΔC_p (cal/mol·K)	Δh_0 (kcal/mol)	ΔC_p (cal/mol·K)
GC bp	41 ± 2	95 ± 11	14 ± 1	85 ± 10
AT bp	21 ± 1	66 ± 8	6.5 ± 0.2	70 ± 6
mixed bp	30 ± 2	77 ± 6	11 ± 1	76 ± 7

Table 7.2.: Melting entropy, enthalpy and heat capacity change per bp measured using Eqs. (7.1b) and (7.1a).

Finally, to derive the temperature-dependent enthalpy difference between N and U, we have used the thermodynamic definition of Gibbs free energy, i.e., $G = H - TS$. To derive the melting enthalpy and heat capacity change from the measured values, we have adjusted the experimental trend to Eq. (7.1a).

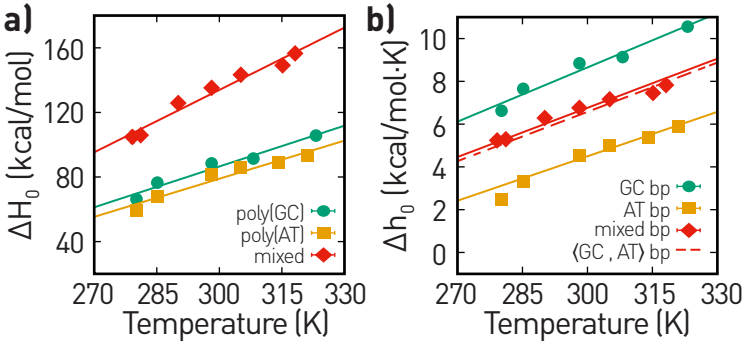

Figure 7.9.: **Temperature-dependent enthalpy change.** (a) Total enthalpy change for the three studied hairpins. (b) Enthalpy difference per bp derived from the experiments: GC bp (green circles), AT bp (yellow squares), and mixed (red diamonds). The dashed line corresponds to the average value considering the experimental results per GC and AT bp. The solid lines in panel (a) and (b) are fits to Eq. (7.1a).

The experimental results of ΔH_0 and Δh_0 are shown in Fig. 7.9. The solid lines in panels (a) and (b) are fits to Eq. (7.1a) to estimate ΔH_0^m, Δh_0^m, and ΔC_p. The estimated Δh_0^m and ΔC_p are presented in Tab. 7.2. In addition, using the values of Δh_0^m and ΔC_p for a GC and AT bp, we have estimated the enthalpy for the mixed bp (dashed red line in Fig. 7.9b), which agrees with the experimental results obtained from the mixed hairpin.

7.2.6 *Loop's length effects*

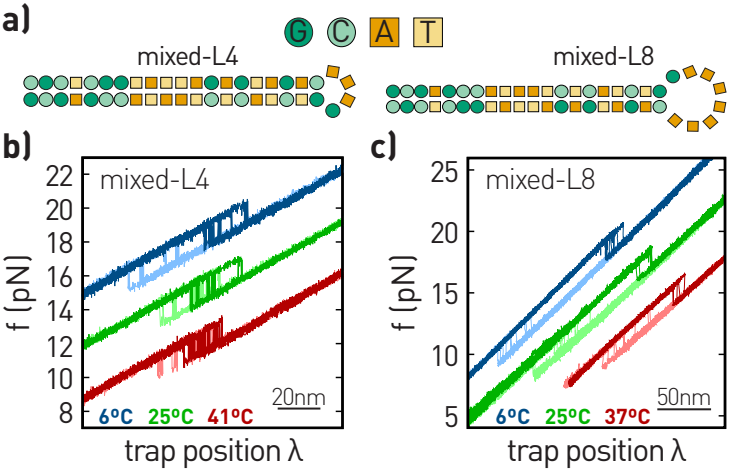

Figure 7.10.: **Pulling experiments with different loops.** (a) Sequence of the mixed hairpins. (b) and (c) Pulling experiments at low (blue), room (green), and high (red) temperatures using the mixed-L4 (a) and mixed-L8 (b) hairpins. In both panels, dark (light) colors denote the unfolding (folding) trajectories.

To conclude this investigation, we have studied the folding thermodynamics of a modified mixed hairpin (Fig. 7.10a) to scrutinize the effects of the loop length in the folding entropy, enthalpy, and heat capacity

change. The modified mixed hairpin, which we called *mixed-L8*, has the same stem as the original mixed hairpin ended with an octa-loop 5'-GAAAAAAA-3'. In this section, the original mixed hairpin is denoted as mixed-L4 to make the comparison with mixed-L8 explicit.

The increase in loop size raises the entropic barrier that the unfolded ssDNA chain must overcome. As a result, the time scales of the unfolding/folding transitions are not in experimentally accessible timescales. Therefore, to investigate the kinetic rates of mixed-L8, we have carried out pulling experiments and compared them with those for the mixed-L4 original hairpin (Fig. 7.10b and c).

To derive the unfolding and folding kinetic rates of mixed-L8, we have proceeded as for protein barnase (Chap. 4). First, we have determined the first unfolding and folding force distributions. Second, we have derived the survival probabilities of N and U from the unfolding and folding force distributions, respectively. The resulting $k_{\rightarrow}$ and $k_{\leftarrow}$ are shown in Fig. 7.11.

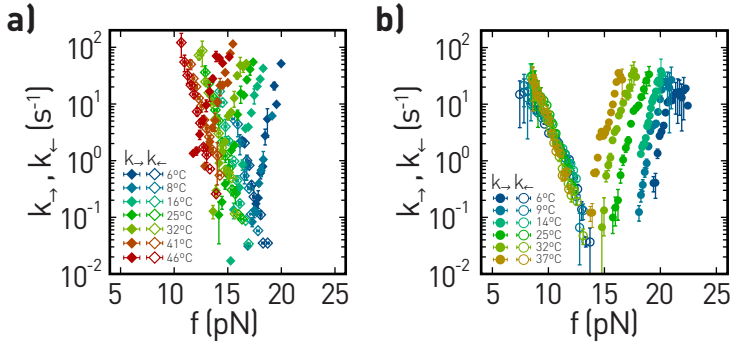

Figure 7.11.: **Force- and Temperature-dependent kinetic rates.** (a) Unfolding (solid symbols) and folding (empty symbols) kinetic rates of mixed-L4 hairpin (same results as the ones shown in Fig. 7.3-bottom) (b) Unfolding (solid symbols) and folding (empty symbols) kinetic rates of mixed-L8 hairpin.

Interestingly, we have found that the refolding kinetic rates of mixed-L8 are virtually unaffected by the temperature. Moreover, the kinetic rates of mixed-L8 hairpin do not cross, in contrast with the mixed hairpin where $k_{\rightarrow}$ and $k_{\leftarrow}$ cross around f_c. This result suggests that the folding process of mixed-L8 is mainly entropic. To determine the folding energy, entropy, and enthalpy of mixed-L8, we have used Eq. (7.7) and the coexistence force derived using the Bell-Evans model.

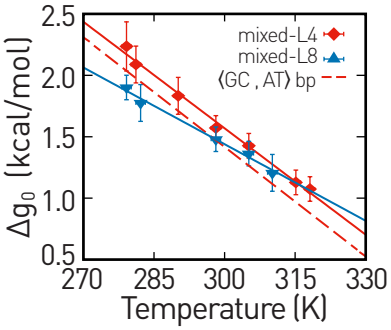

Figure 7.12.: **Folding free energy per bp.** Δg_0 derived using the mixed-L4 (red) and mixed-L8 (blue) hairpins. The red dashed line correspond to the expected energy using the Δh_0^m and Δs_0^m values using the GC and AT hairpins.

Figure 7.12 shows the derived folding free energy per bp using the mixed-L4 and mixed-L8 hairpins. The solid lines in Fig. 7.12 are linear fits to the experimental values to derive the melting temperature of the hairpins. The measured melting temperature for the mixed-L8 hairpin is around 96°C. The red dashed line in Fig. 7.12 correspond to the expected energy using the values of Δh_0^m and Δs_0^m derived from the GC and AT bps. To derive the heat capacity change per bp, we have investigated the temperature-dependent entropy and enthalpy. In figure 7.13, we show the measured Δs_0 (panel a) and Δh_0 (panel b) in symbols (red: mixed ; blue: mixed-L8) together with the fits to Eqs. (7.1b) and (7.1a), respectively. The measured total and per bp ΔS_0^m, ΔH_0^m, and average ΔC_p for both hairpins are summarized in Tab. 7.3.

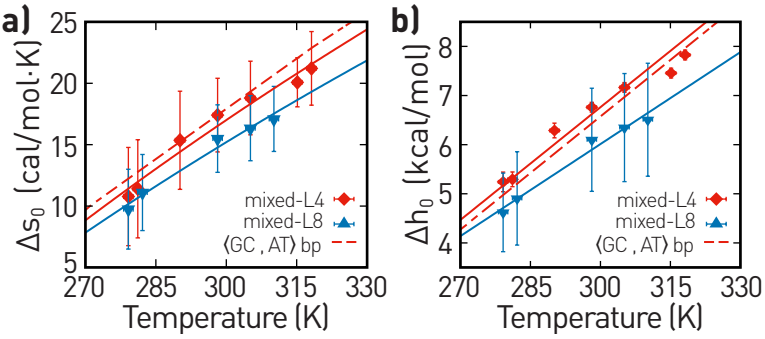

Figure 7.13.: **Entropy and enthalpy difference per bp.** (a) Δs_0 and (b) Δh_0 derived using the mixed-L4 (red) and mixed-L8 (blue) hairpins.

	ΔS_0 (cal/mol·K)	ΔH_0 (kcal/mol)	ΔC_p (cal/mol·K)
mixed-L8	587 ± 20	212 ± 8	1400 ± 120
mixed-L4	600 ± 40	220 ± 20	1540 ± 140
mixed-L8 (per bp)	29 ± 1	10.3 ± 0.4	70 ± 6
mixed-l4 (per bp)	30 ± 2	11 ± 1	77 ± 7

Table 7.3.: Melting entropy, enthalpy and heat capacity change

We have compared our results with estimations based on the unified oligonucleotide database, also known as *Mfold* server [127, 128], and previous experiments based on long DNA unziping experiments [116, 129] (denoted as HU) to validate the values of ΔG_0, ΔS_0^m, and ΔH_0^m. These database assume that the DNA hybridization does not exhibit a significant heat capacity change, i.e., ΔS and ΔH are constant with T. Here, we have demonstrated that this assumption is not true. Nevertheless, we can see if the entropy and enthalpy differences between both hairpins agrees with the that predicted by Mfold and HU. These databases state that the difference in entropy and enthalpy between mixed and mixed-L8 is around 11 cal/mol·K and 4 kcal/mol, respectively. Our differences are 13 ± 2 cal/mol·K and 8 ± 2 kcal/mol, the error-bars are the statistical

error considering all the studied molecules. Therefore, what causes the difference in the folding kinetic rates between the two hairpins?

Entropy and enthalpy of TS

To try to answer this question, we have derived the entropy and enthalpy of the transition state relative to N ($\Delta S^{\dagger}$ and $\Delta H^{\dagger}$, with $\Delta^{\dagger} = {}^{TS} - {}^{N}$) and U (ΔS^{*} and ΔH^{*}, with $\Delta^{*} = {}^{TS} - {}^{U}$). The procedure to derive $\Delta S^{\dagger}$, $\Delta H^{\dagger}$, ΔS^{*}, and ΔH^{*} is analogous to the one used in Chap. 4 for protein barnase. The main difference is that the kinetic attempt k_a of both hairpins has been derived using the Continuous Effective Barrier Approach (CEBA) [130]. In this approach, the attempt rates k_a are derived by matching the experimental values of $-\log k_{\rightarrow}(f)$ and the barriers predicted using the nearest-neighbor base-pair energies [127, 128],

$$\frac{B_{NU}(f)}{k_B T} = \log\left(\sum_{m=0}^{M}\sum_{m'=0}^{m} e^{\left(\frac{\Delta G_m(f) - \Delta G_{m'}(f)}{k_B T}\right)}\right) . \tag{7.8}$$

In equation (7.8), the double sum runs over all hairpin configurations, labeled by m and m', and M being the total number of base pairs. In Chap. 9, we provide more details of CEBA and investigate the mFEL of different DNA hairpins exhibiting several intermediates states along the folding pathway.

The values of k_a for the mixed and mixed-L8 have been measured by matching the experimental $\log(k_{\rightarrow})$ to the expected barrier using Eq. (7.8) (see Chap. 9 for more details). The measured values are 8×10^3 s^{-1} and 5.6×10^3 s^{-1} for the mixed and mixed-L8, respectively. Using these values we have simultaneously fitted the measured kinetic rates at zero force ($k^0_{\rightarrow}$ and $k^0_{\leftarrow}$) to the following equations,

$$k^0_{\rightarrow} = k_a \exp\left(-\frac{\Delta G^{\dagger}}{k_B T}\right) \tag{7.9a}$$

$$k^0_{\leftarrow} = k_a \exp\left(-\frac{\Delta G^{*}}{k_B T}\right) , \tag{7.9b}$$

imposing $\Delta S_0 = \Delta S^\dagger - \Delta S^*$ and $\Delta H_0 = \Delta H^\dagger - \Delta H^*$.

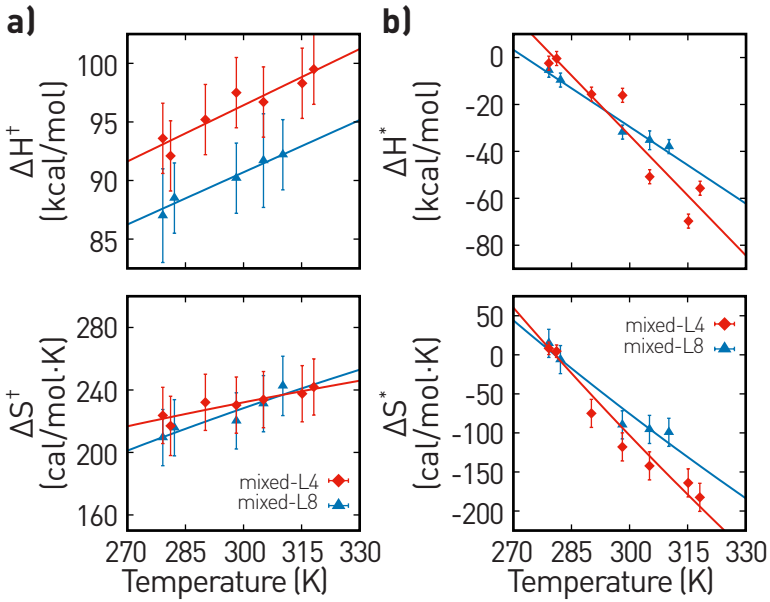

Figure 7.14.: **Entropy and enthalpy difference at the *TS*.** ΔS and ΔH relative to N (a) and U (b). Solid lines are fits to Eqs. (7.1b) and (7.1a).

The derived entropy and enthalpy differences at TS are shown in Fig. 7.14. Here, we want to see if we can see a noticeable difference in the entropic barrier induced by the large loop of the mixed-L8 hairpin. Comparing the values of ΔS^* (Fig. 7.14 panels (b)) for the mixed-L4 and mixed-L8 hairpins, we can see that the entropic barrier to fold for both hairpins is approximately equal.

Finally, we have determined the heat capacity change between N and TS ($\Delta C_p^\dagger = C_p^{TS} - C_p^N$) and between U and TS ($\Delta C_p^* = C_p^U - C_p^{TS}$) by fitting the temperature behavior of the experimental values of $\Delta H^\dagger$, ΔH^*, $\Delta S^\dagger$ and ΔS^* to Eqs. (7.1a) and (7.1b), respectively. Here, we define the heat capacity change with this sign criteria to have positive values. The

measured heat capacity change, and entropy and enthalpy at the melting temperature are summarized in Tab. 7.4. Interestingly, as for the case of protein barnase, $\Delta H_m^\dagger \gg \Delta H_m^*$ and $\Delta S_m^\dagger \gg \Delta S_m^*$.

	$\Delta S_m^\dagger$ (cal/mol·K)	$\Delta H_m^\dagger$ (kcal/mol)	$\Delta C_p^\dagger$ (cal/mol·K)
mixed-L8	282 ± 13	101 ± 5	204 ± 40
mixed-L4	256 ± 5	105 ± 3	152 ± 26
	ΔS_m^* (cal/mol·K)	ΔH_m^* (kcal/mol)	ΔC_p^* (cal/mol·K)
mixed-L8	-311 ± 38	-105 ± 9	1114 ± 135
mixed-L4	-364 ± 24	-127 ± 15	1428 ± 180

Table 7.4.: Melting entropy, enthalpy and heat capacity change at the transition state relative to N and U.

Moreover, $\Delta C_p^\dagger \ll \Delta C_p^*$ with $\Delta C_p^* \sim \Delta C_p$, as seen for protein barnase. These results suggest that the mechanisms governing the DNA folding are similar to those in protein folding. In addition, from Fig. 7.14a,b, we can see that $\Delta H^\dagger \gg \Delta H^*$ and $\Delta S^\dagger \gg \Delta S^*$ in all the explored temperature range. We interpret this result in a way where the two single-strands of the DNA hairpin are properly oriented and facing each other but with the hydrogen bonds not yet formed at the transition state. In the collapse from TS to N, we the hydrogen bonds forming the stem of the hairpin are stabilized. From this perspective, the molecular free-energy landscape governing DNA folding looks similar to the one for protein barnase. The energetic, enthalpic and entropic contributions are depicted in Fig. 7.15.

7.3 CONCLUSIONS

In this chapter, we have used the calorimetric optical tweezers to measure the unfolding and folding kinetic rates of three different hairpins (poly(GC), poly(AT), and mixed) from equilibrium hopping experiments.

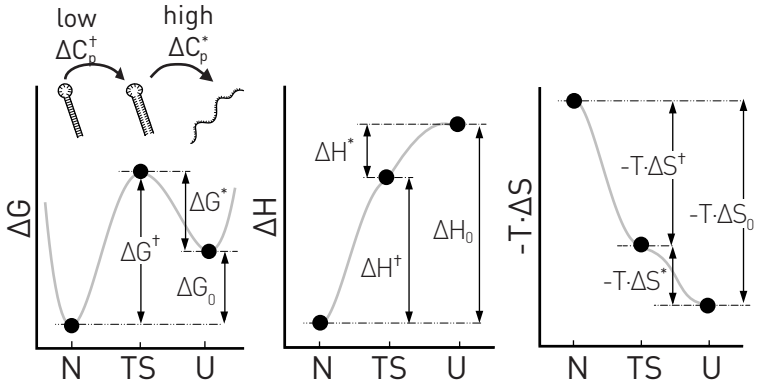

Figure 7.15.: **Free-energy, enthalpy and entropy landscape.** Illustrative depiction the energetic, enthalpic, and entropic landscape.

The three investigated hairpins are specifically chosen to derive the folding energy, entropy, enthalpy, and heat capacity change of a GC and AT base-pair (bp) and the average value per bp. Moreover, we have studied the effect of the loop size in the folding energy, entropy, and enthalpy difference by using two hairpins with the same stem ended with a tetra- and an octa-loop (mixed and mixed-L8 hairpins).

First, we derived the folding energy ΔG, entropy ΔS, enthalpy ΔH, and heat capacity ΔC_p changes from the kinetic measurements carried out in a wide temperature range, $5 - 50^{o}$C. We have used the well-known Bell-Evans model (Eqs. (7.3a) and (7.3b)) and the Detailed Balance Condition (Eq. (7.4)) to derive the temperature-dependent ΔG. An analysis based on the Clausius-Clapeyron equation (Eq. (7.7)) was used to extract the temperature-dependent values of ΔS. In addition, using the thermodynamic definition of Gibbs free-energy, we have determined ΔH as a function of temperature. We have derived the thermodynamic potentials for a GC and AT base pair assuming that that a 5'-GC-3' (5'-AT-3') or a 5'-CG-3' (5'-TA-3') motif has the same energy. Then, to derive the energy (Δg), entropy (Δs), and enthalpy (Δh) per GC bp, we have divided the total ΔG, ΔS, and ΔH obtained from poly(GC) by the number of base

pairs. To derive Δg, Δs, and Δh per AT bp, we have used the values obtained from the poly(AT) hairpin and subtracted the contributions of the first two GC bps. The rest has been divided by the total number of AT bps. Finally, the values obtained from the mixed hairpin have been divided by the total number of bps to measure the average Δg, Δs, and Δh per bp. Remarkably, the expected averaged energy, enthalpy, and entropy per bp obtained using the experimental values per GC and AT bp derived from the poly(GC) and poly(AT) hairpins underestimate the values derived from the mixed hairpin. We hypothesize that the difference in energy arises from the purine-purine stacking contributions of the trinucleotide and tetranucleotide motifs present in the mixed hairpin. The energy, entropy, and enthalpy per purine-purine stacking motifs are $\Delta\Delta G \sim 0.3$kcal/mol, $\Delta\Delta H \sim 0.3$kcal/mol, and $\Delta\Delta S \sim 0.6$cal/mol·mol. The contributions per purine stacking are calculated as the difference between the values measured for the mixed hairpin, and the total energy using the Δg, Δh, and Δs values per GC and AT bp results divided by the seven purine-purine stacking interactions (highlighted in Fig. 7.1). Our results pave the way to study the nucleic acid stability beyond the nearest-neighbor model by considering the purine-purine stacking interactions overcoming the first neighbor.

Second, we have derived the heat capacity change ΔC_p per GC, AT, and mixed bp. According to previous results based on bulk assays, depending on the sequence and length employed, ΔC_p values range between $40 - 160$ cal/mol·K per bp [122, 123]. Here, taking profit of the single-molecule resolution, we have measured the ΔC_p per bp with the sufficient resolution to confirm that ΔC_p for a GC or AT bp is different. The measured ΔC_p are 90 ± 11 cal/mol·K, 68 ± 8 cal/mol·K, and 77 ± 7 cal/mol·K per GC, AT and mixed bp, respectively.

Third, we have extended our investigation to a modified mixed hairpin (in this last section, we name the original mixed hairpin as mixed-L4), which we named mixed-L8, that has the same stem as the original hairpin, but with doubled loop size. Unexpectedly, the folding kinetic rates of mixed-L8 are temperature independent. To validate these results,

we have compared our experimental results with the unified oligonucleotide database, commonly known as the Mfold server. The measured difference in enthalpy ($\sim$ 13 cal/mol·K) and entropy ($\sim$ 8 kcal/mol) between the mixed-L4 and the mixed-L8 hairpins follows the trend of the predictions based on Mfold, 10.3 cal/mol·K and 3 kcal/mol, respectively. The measured differences represent only $\sim$ 2 % and $\sim$ 4 % of the total entropy and enthalpy change. Therefore, why the folding kinetic rates of mixed-L8 do not change with temperature? To answer this question, we have derived the entropic and enthalpic contributions of the kinetic barrier mediating transitions between N and U. We have observed that the entropic barrier of mixed-L8 is slightly higher than the one measured for the mixed hairpin. Moreover, we have determined the heat capacity at the transition state relative to N ($\Delta C_p^\dagger$) and U (ΔC_p^*). The measured heat capacity changes, $\Delta H^\dagger$, ΔH^*, $\Delta S^\dagger$, and ΔS^* follow the trend observed using the protein barnase (Chap. 4), $\Delta C_p^* \ll \Delta C_p^\dagger$, $\Delta H^* \gg \Delta H^\dagger$, and $\Delta S^* \gg \Delta S^\dagger$. This fact suggests that the proteins' biological mechanisms governing the folding transitions are the same as those of DNA molecules. ΔC_p^* quantifies the change of degrees of freedom Δn upon bending the loop to align the two ssDNA chains opposite each other. In the subsequent collapse from TS to N [131], the hydrogen bonds that stabilize the double helix are formed. $\Delta C_p^\dagger$ quantifies the change on the number degrees of freedom of this collapse.

8

DNA UNZIPPING EXPERIMENTS

MOTIVATION

DNA hybridization is still nowadays a hot topic in molecular biology with many applications. The nearest-neighbor (NN) model for nucleic acids predicts DNA thermodynamics using energy values for the different base-pair motifs. In the past, single-molecule techniques have been used to derive the energy of the NN motifs varying salt conditions. Here, we want to study the unzipping of long DNA hairpins to infer the temperature-dependent folding energy of the NN motifs.

8.1 INTRODUCTION AND HISTORICAL CONTEXT

DNA hybridization is the process when two complementary ssDNA chains meet each other, forming a dsDNA molecule. DNA mainly appears in vivo conditions with the well-known double-helix conformation, primarily stabilized by the hydrogen bonds connecting the nucleotide bases of the ssDNA chains. In *in vivo* conditions, DNA is in charge of storing and transmitting genetic information for future generations. Therefore, accurate knowledge of the melting or opening of DNA is critical to understand, e.g., the DNA replication [132]. The replication of DNA molecules is widely used in industry, e.g., polymerase chain reac-

tions (PCR) amplify DNA sequences for clinical purposes. In this technique, the temperature of the sample is repeatedly raised and lowered to favor a DNA replication enzyme copy the target DNA sequence [133]. A possible approach to enhance the PCR cycles requires appropriately adjust the temperature limits to improve the efficiency of the enzymatic activity. Therefore, a precise prediction of the free energy formation and the melting temperature is required.

The folding free energy of a DNA duplex is predicted by the nearest-neighbor (NN) model for nucleic acids [134, 135]. Briefly, the NN model defines the total energy as the addition of all energy contributions of adjacent nearest-neighbor base-pairs stacks along the sequence. To determine the folding energy of an arbitrary ssDNA sequence is necessary to search for the minimum energetic level among all possible secondary structures. Hence, it requires the implementation of an optimization algorithm such as the one used in the oligonucleotide database, also known as *Mfold* [127, 128]. This database uses the NN model to minimize the energy of a given sequence of ssDNA to estimate its native or minimum energetic state.

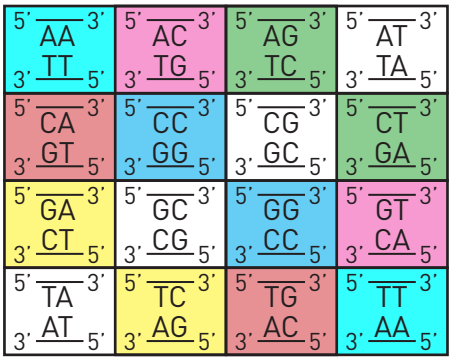

Figure 8.1.: **The sixteen NN base-pairs.** The twelve highlighted motifs are symmetric with respect to the anti-diagonal (each color correlates with its symmetric).

In the NN model, sixteen different motifs define all possible combinations of first neighbors. The sixteen motives are summarized in Fig.

8.1, and each motif has a different folding energy. However, as can be seen in Fig. 8.1, we have some symmetries between motifs concerning the anti-diagonal (white boxes). These symmetries are highlighted with different colors. So, the NN model is defined with only ten different folding energies, i.e., the ten symmetric motifs plus the ones in the anti-diagonal.

In the last years, single-molecule experiments (SMEs) carried out using optical tweezers have been used to derive the folding energy of the ten different NN DNA motifs (the white and colored NN base-pairs in Fig. 8.1) at different ionic strengths [116, 129]. Here, we carry out unzipping experiments under different temperature conditions to derive the temperature-dependent folding energy of the ten NN DNA motifs. Moreover, we compare the measured folding energies with the previously shown tests using shorts hairpins with different GC content.

8.2 RESULTS

8.2.1 *Force distance curves*

To derive the folding energies, we have carried pulling experiments using a long DNA hairpin at different temperature conditions. A DNA hairpin is a specific secondary structure where an ssDNA folds into themselves, forming an individual hairpin, i.e., a dsDNA stem ended with an ssDNA loop. In particular, we have used a DNA hairpin formed by a 3594 base-pairs (bp) stem connected with a tetra-loop 5'-ACTA-3'.

The long DNA hairpin has been synthesized between two identical 29bp dsDNA handles to connect the hairpin and the dielectric beads used to exert and measure the external force done by the optical trap. In unzipping experiments, the two ends of the molecular construct are moved up and down to unzip the DNA hairpin mechanically. In contrast with short hairpins, where the unzipping of the bp forming the stem occurs in a single force rip (see, e.g., Chap. 3), long hairpins exhibit

a serrated pattern. Each force rip in the force-distance curves (FDCs) corresponds to the unzipping of many bp. In figure 8.2, we show the average unzipping FDCs at different temperature conditions: 7, 13, 16, 19, 22, 25, 30, 36, and 42°C. The average unzipping force is strongly affected

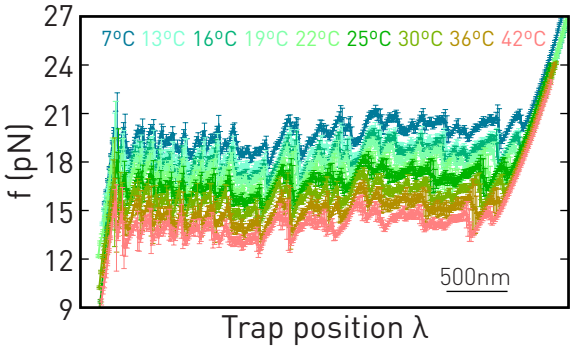

Figure 8.2.: **Measured FDCs at different temperatures.** Blue and red trajectories correspond to the limit temperatures, i.e., 7 and 42°C. The temperatures in-between are represented by the colors between blue and red.

by the temperature. More force is required to break the hydrogen bonds stabilizing the hairpin at low temperatures (blue trajectory) than high temperatures (red trajectory). This fact agrees with the results measured from hopping experiments using short DNA hairpins.

8.2.2 *Elastic response*

From the measured FDCs, we have derived the molecular extension of the DNA hairpin as a function of force. At a given force, the trap position λ equals

$$\lambda(f) = \lambda_0 + x_h(f) + x_b(f) + x_m(f) \,, \tag{8.1}$$

where λ_0 is the initial value relative to the center of the trap position detector, x_h is the extension of the handles, $x_b(f)$ is the position of the

bead relative to the center of the optical trap, and $x_m(f)$ is the molecular extension.

The contributions x_h and x_b have been calculated following the methodology used to determine the molecular extension for short DNA hairpins [112],

$$x_b(f) + x_h(f) = \frac{f}{k^N_{eff}} \tag{8.2}$$

where k^N_{eff} is the effective stiffness of the system bead plus handles plus DNA molecule when the molecule is in its native state, i.e., all the bp formed. k^N_{eff} has been calculated doing a linear fit at the beginning of the FDCs before the first force rip.

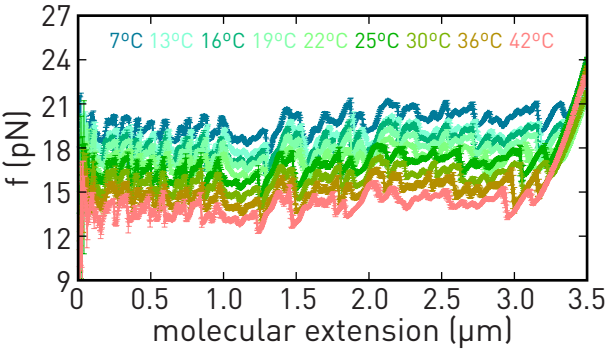

Figure 8.3.: **Measured FECs at different temperatures.** Blue and red trajectories correspond to the limit temperatures, i.e., 7 and 42°C. The temperatures in-between are represented by the colors between blue and red.

The measured force-molecular extension curves (FECs) are shown in Fig. 8.3. From the last part of the trajectories, i.e., when the molecule is completely unzipped, we have derived the elastic properties of the long ssDNA chain at different temperatures. To characterize the elasticity of the ssDNA, we have used the Worm-Like Chain (WLC) model and its interpolation formula,

$$f = \frac{k_B T}{4L_p}\left[\left(1 - \frac{x_m}{Nd_b}\right)^{-2} + 4\frac{x_m}{Nd_b} - 1\right] \tag{8.3}$$

where L_p is the persistence length, d_b is the inter-phosphate distance, N is the number of bases ($N = 3594 \cdot 2 + 4$ when the hairpin is totally unfolded), k_B is the Boltzmann constant and T is the temperature. The

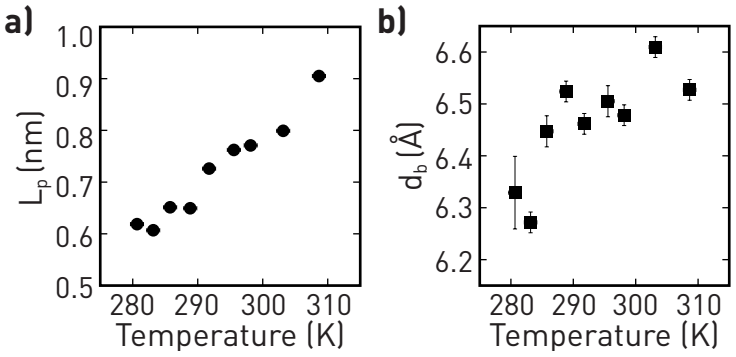

Figure 8.4.: **Elasticity of long ssDNA chains.** (a) Persistence length and (b) inter-phosphate distance as a function of temperature.

extracted temperature-dependent L_p and d_b are shown in Fig. 8.4. We have derived these values by adjusting the last part of the FECs to Eq. (8.3) with L_p and d_b as fitting parameters. The reported L_p and d_b for long ssDNA chains exhibit the same temperature trend than the results for short DNA hairpins shown in Chap. 6.

8.2.3 *Unzipped base-pairs*

The experiments have been carried out applying low pulling rates r ($r \sim 3$ pN/s). With these conditions, the unzipping and rezipping trajectories do not exhibit hysteresis, i.e., the system is stretched quasi reversibly. Therefore, the average unzipping trajectory can be treated as an

equilibrium unzipping curve. The equilibrium curve is estimated from the partition function, Z, at a total molecular extension x_m,

$$Z = \sum_n \exp\left(-\frac{G(x_m, n)}{k_B T}\right) \tag{8.4a}$$

$$f_{eq} = -k_B T \frac{\partial}{\partial x_m} \ln Z(x_m) \, . \tag{8.4b}$$

In equation (8.4a), $G(x_m, n)$ is the free energy at a given molecular extension calculated over all possible intermediate state given n unfolded bp. f_{eq} in Eq. (8.4b) corresponds to the expected force under equilibrium conditions.

We have determined the number of unzipped DNA bp at a given x_m using the elastic properties shown in Fig. 8.4. In figure 8.5, we illustrate the methodology to determine the number of released bp at a given x_m and f. In figure 8.5, we show a single FEC measured at 25°C

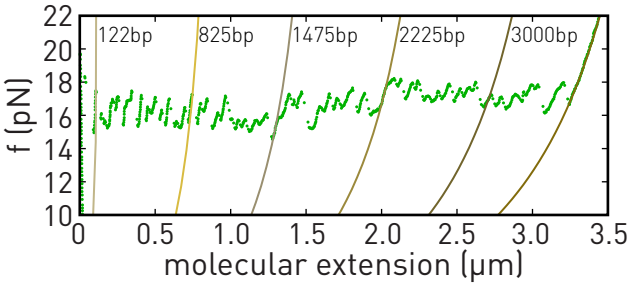

Figure 8.5.: **Unzipped base-pair determination.** Dots correspond to a single unzipping trajectory measured at 25°C. At the same time, solid lines are fits to Eq. (8.4) with L_p and d_b fixed, leaving N as a fitting parameter.

using green dots. As can be seen, the trajectory has different clouds of points corresponding to the stretching of the released ssDNA. Each point has been fitted to Eq. (8.4) leaving N as a free parameter with L_p and d_b fixed. We have recovered the distributions of released base pairs by repeating this procedure for several molecules and trajectories at a given temperature. The derived bp distributions are shown in Fig. 8.6.

Interestingly, the distributions of unzipped base pairs is not affected by the temperature.

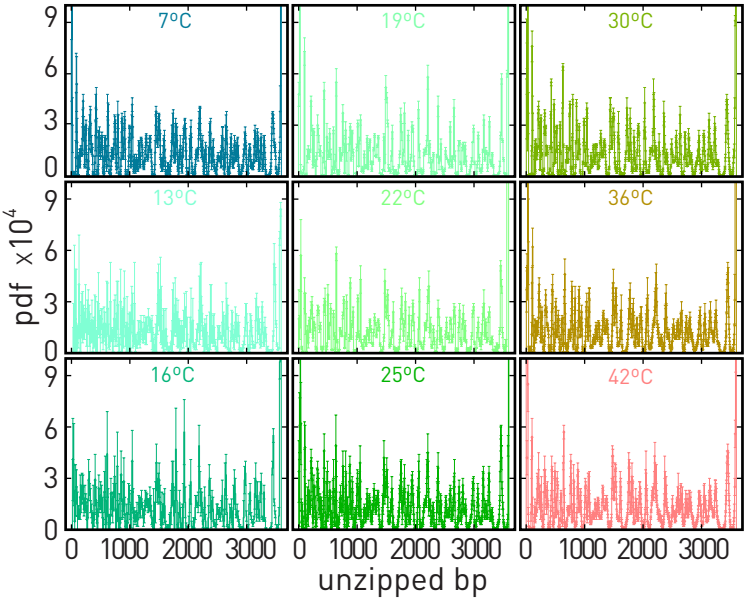

Figure 8.6.: **Unzipped base-pair distributions.** Unzipped base pairs distributions at all studied temperatures.

8.2.4 *Folding free energy*

Next, we have adjusted Eqs. (8.4a) and (8.4b) to derive the folding energy of the ten different NN DNA motifs, Δg_i ($i = 1, 2, ..., 10$). First, mention that the energy $G(x_m, n)$ term has two contributions. On one side, we have the energy cost to stretch the unzipped molecule at a given force f. On the other hand, we have the stacking and hydrogen-bond energies of the bases, which in the NN model considers the first neighbor, i.e., sixteen different energy contributions.

To adjust Eq. (8.4b) to the average experimental FECs, we have minimized the difference between the experimental and theoretical curves using a Monte Carlo based approach. The error E that needs to be minimized is defined as,

$$E(\Delta g_1, ..., \Delta g_{10}) = \frac{1}{N} \sum_{j}^{N} \left(f - f_{eq}(\Delta g_1, ..., \Delta g_{10}) \right)^2 \tag{8.5}$$

where f is the experimental values of force, f_{eq} is the theoretical force, and N is the total number of points per trajectory.

In Figure 8.7 the average experimental curves at 7 (blue), 25 (green), and 42ºC (red) are shown together with the theoretical curve that best fits the experimental curves (solid lines).

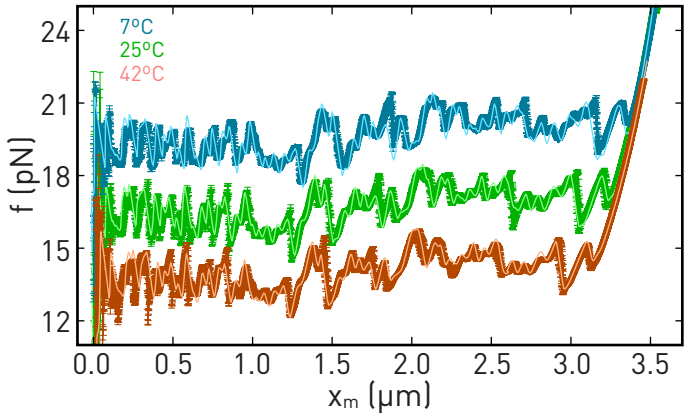

Figure 8.7.: **Equilibrium curves.** Experimental average unzipping curves (dots) at 7 (blue), 25 (green), and 42ºC (red). Solid lines are the theoretical equilibrium curve that best fits the experimental data.

The temperature-dependent folding free energy of the six symmetric NN DNA motifs are shown in Fig. 8.8, whereas the the energy of the anti-diagonal NN motifs is shown in Fig. 8.9.

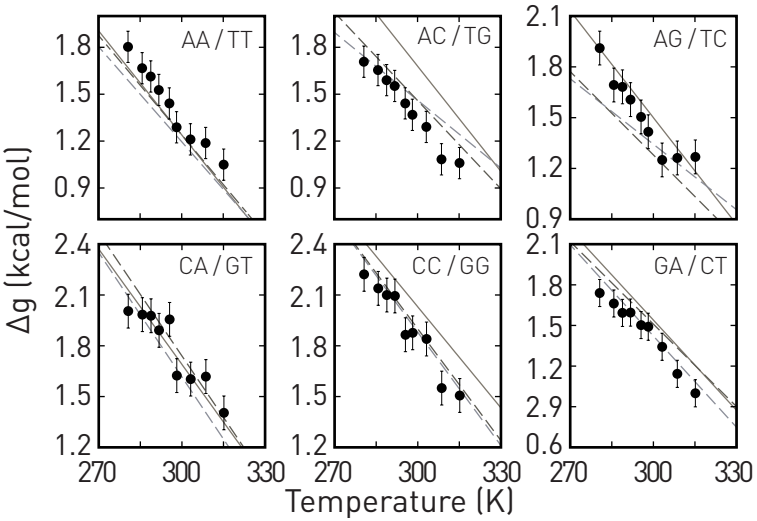

Figure 8.8.: **Folding free energy.** Experimental folding free energy (symbols) for the six symmetric NN DNA motifs derived from the FECs. The solid lines are the expected behavior from the oligonucleotide database [127, 128], while the dashed lines are the expected behavior using the data from [116, 129].

From figure 8.8, we can see that the measured Δg has a strong temperature dependence. According to the Gibbs free energy definition ($G = H - TS$), we have determined the entropy (Δs) and enthalpy (Δh) difference for each motif in the assumption of zero heat capacity change along melting. Besides, in Fig. 8.8, we have represented the expected folding energy per motif considering the entropy and enthalpy provided by the oligonucleotide database Mfold [127, 128] (solid line). Moreover, we have compared our temperature behavior with experiments carried out using optical tweezers varying the ionic strength to obtain different melting temperatures [116, 129] (dashed lines). The derived entropy and enthalpy per motif of DNA are shown in Tabs. 8.1 and 8.2, respectively.

Our experimental values agree with the tabulated ones in Mfold and those derived from unzipping experiments varying the salt condition.

	Our data Δs (cal/mol·K)	Mfold Δs (cal/mol·K)	Ref [70] Δs (cal/mol·K)	Ref [83] Δs (cal/mol·K)
AA/TT	22.3 ± 1.6	22.2 ± 2.2	20.28	21.20
AC/TG	20.9 ± 1.2	22.4 ± 2.4	14.46	18.80
AG/TC	19.7 ± 2.6	21.0 ± 2.1	12.89	16.40
CA/GT	18.5 ± 2.8	22.7 ± 2.7	24.48	24.20
CC/GG	22.3 ± 1.9	9.9 ± 2.0	22.30	22.00
GA/CT	21.7 ± 1.6	22.2 ± 2.2	22.46	20.30
AT/TA	15.5 ± 2.3	20.4 ± 2.2	11.62	20.40
TA/AT	15.7 ± 3.2	21.3 ± 2.1	25.06	15.30
GC/CG	27.3 ± 1.5	24.4 ± 2.4	25.96	25.00
CG/GC	29.5 ± 2.1	27.2 ± 2.7	24.43	24.30

Table 8.1.: Entropy per DNA motif

	Our data Δh (kcal/mol)	Mfold Δh (kcal/mol)	Ref [70] Δh (kcal/mol)	Ref [83] Δh (kcal/mol)
AA/TT	8.0 ± 0.5	7.9 ± 0.8	7.28	7.60
AC/TG	7.6 ± 0.4	8.4 ± 0.8	5.80	7.10
AG/TC	7.3 ± 0.8	7.8 ± 0.8	5.21	6.20
CA/GT	7.3 ± 0.8	8.5 ± 0.8	8.96	9.00
CC/GG	8.5 ± 0.6	8.0 ± 0.8	8.57	8.50
GA/CT	7.9 ± 0.5	8.2 ± 0.8	8.16	7.60
AT/TA	5.8 ± 0.7	7.2 ± 0.7	4.63	7.10
TA/AT	5.7 ± 0.9	7.2 ± 0.7	8.31	5.45
GC/CG	10.6 ± 0.4	9.8 ± 0.9	10.1	9.80
CG/GC	11.0 ± 0.6	10.6 ± 1.0	9.66	9.60

Table 8.2.: Enthalpy per DNA motif

In Figure 8.9, we show the obtained folding energies for the NN motifs in the anti-diagonal of Fig. 8.1. Notice that these motifs are the ones present in the short DNA hairpins investigated in Chap. 7. Therefore, we can compare the results obtained from the unzipping experiments and the results obtained from hopping experiments.

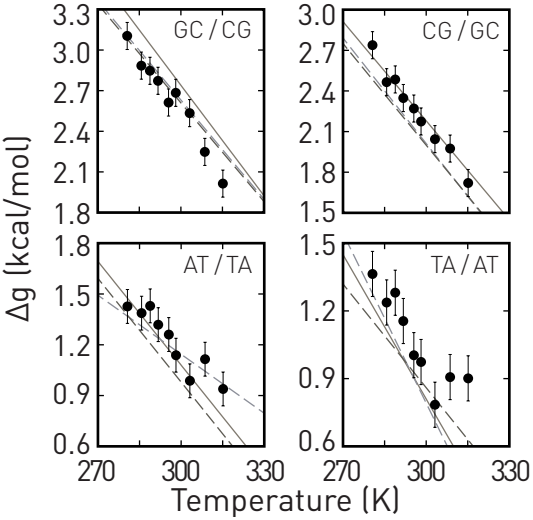

Figure 8.9.: **Folding free energy.** Experimental folding free energy (symbols) for the four anti-symmetric NN DNA motifs derived from the FECs. The solid lines are the expected behavior from the oligonucleotide database [127, 128], while the dashed lines are the expected behavior using the data from [116, 129].

At first sight, in Fig. 8.9, we can see that the average energy per AT bp (i.e., average value considering the AT and TA motifs) is lower than the average energy per GC bp (i.e., mean value considering the GC and CG motifs). These results agree with the fact that GC bps have three hydrogen bonds, while an AT bp has only two. As the previous motifs, we have adjusted the temperature behavior with a linear function to derive the entropy and enthalpy difference (see Tabs. 8.1 and 8.2).

Finally, we have compared the energy per GC and AT bp derived from unzipping experiments with those derived from hopping experiments using short DNA hairpins.

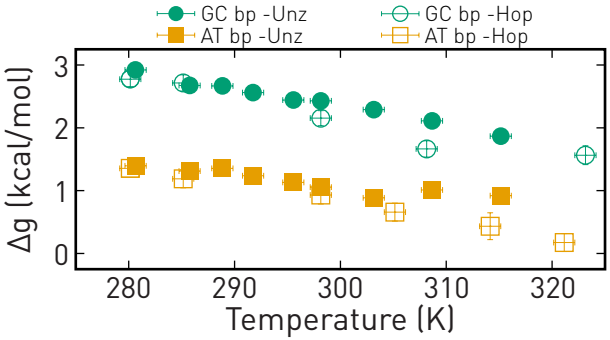

Figure 8.10.: **Folding free energy per GC and AT bp.** Comparison between the results derived from unzipping (solid symbols) and hopping (empty symbols) experiments for a GC (green) and AT (yellow) base pairs.

We have represented the energy per GC and AT base pair as a function of temperature in Fig. 8.10. Here, the solid symbols represent the values derived from the unzipping experiments, whereas the empty symbols correspond to those derived in Chap. 7. As can be seen, both results agree pretty well. The enthalpy and entropy difference per GC and AT bp are summarized in Tab. 8.3.

	Hopping		Unzipping	
	Δh_0 (kcal/mol)	Δs_0 (cal/mol·K)	Δh_0 (kcal/mol)	Δs_0 (cal/mol·K)
GC bp	12 ± 1	31 ± 5	11.1 ± 0.6	30 ± 2
AT bp	9.2 ± 0.4	28 ± 1	5.7 ± 0.7	16 ± 2

Table 8.3.: Derived enthalpy and entropy per bp

8.3 CONCLUSION

Here, we have extended the investigations begun in Chap. 7 by investigating the unzipping of a long DNA hairpin at different temperature conditions keeping fixed the ionic strength. We remark that the ionic strength has been kept fixed because previous unzipping experiments have tuned the salt concentration to simulate different temperatures (the melting temperature varies with the ionic strength).

We have carried out unzipping experiments at slow pulling rates ($r \sim 3$ pN/s) to derive the folding free energy per bp considering the nearest-neighbor (NN) model. In this model, the formation energy of a bp of DNA is affected by the first neighbor. Therefore, we have sixteen combinations (see Fig. 8.1) of bp motifs. Nevertheless, due to some symmetries, these sixteen folding energies have only ten different values. We have derived the temperature-dependent folding free energy of such NN bp motifs by adjusting the equilibrium force-extension curve (Eq. (8.4b)) to the experimental FECs using a Monte Carlo based algorithm.

The measured folding energies are shown in two different figures. On the one hand, Fig. 8.8 shows the energy of the symmetric NN bp motifs as a function of temperature. Notice that the temperature strongly affects the measured folding free energy per motif, Δg. According to the Gibbs free energy definition, we have derived the entropy and enthalpy difference of such DNA motifs. The recorded values are tabulated in Tabs. 8.1 and 8.2. Moreover, in these tables, we show the values used in the oligonucleotide database (second column) and the values derived from unzipping experiments varying the salt concentration (third and fourth columns). We found a good agreement between our experimental values and the published in the literature [116, 127–129].

On the other hand, the motifs without symmetry, i.e., GC/CG, CG/GC, AT/TA, and TA/AT, are shown in Fig. 8.9. As for the other motifs, we have derived its enthalpy and entropy difference by adjusting their temperature behavior with a linear trend (Tabs. 8.1 and 8.2).

Finally, we have compared the average temperature-dependent energy per GC and AT bp derived from unzipping experiments with the results obtained from hopping experiments using short DNA hairpins (Chap. 7). The folding free energy per bp for the GC and AT bps is shown in Fig. 8.10, and the derived enthalpy and entropy differences are compared in Tab. 8.3.

Part V

COLLABORATIONS

9

EXTENDED CONTINUOUS EFFECTIVE BARRIER APPROACH

MOTIVATION

Some nucleic acids and proteins require intermediate or partially folded configurations to perform their biological function. Therefore, a quantitative characterization of the dynamical formation of intermediates is a critical step towards the elucidation of many molecular processes. Accordingly, it is of high interest to develop accurate tools to investigate the thermodynamic and kinetic properties of partially folded domains occurring in proteins and nucleic acid hairpins.

9.1 INTRODUCTION AND HISTORICAL CONTEXT

Intermediate or partially folded configurations appear in many biological processes. For example, RNA riboswitches are regulatory molecules that induce or repress gene transcription depending on their conformation [136]; RNA thermometers act like lockers whose ribosomal binding site becomes accessible only at high temperatures when they partially unfold [137, 138]; and proteins fold into the native structure by forming intermediate folding units (foldons) [139, 140]. Single-molecule experiments (SMEs) provide an ideal platform to experimentally investigate

these intermediates since they allow to sample transient molecular states with high temporal ($\sim$ ms) and spatial resolution ($\sim$ nm). In particular, atomic force microscopy [141, 142], magnetic [143, 144] and optical tweezers [129, 145, 146], allow us to pull on individual molecules and to monitor the unfolding and folding reactions from the recorded changes in force or extension. The molecular extension is commonly adopted as the reaction coordinate in these experiments [147, 148].

Kinetic folding studies are often performed in equilibrium conditions, for instance, in hopping experiments [65, 71, 72, 74, 149, 150]. There, the control parameter (in our case the trap position) is kept fixed as the molecule executes thermally-driven transitions between different molecular states. In such experiments, the unfolding and folding kinetics are derived from the average lifetime of each state [24, 113]. However, equilibrium experiments are strongly limited by the height of the kinetic barrier B mediating transitions between contiguous states along the molecular free energy landscape (mFEL) (Fig. 9.1). A too high kinetic barrier

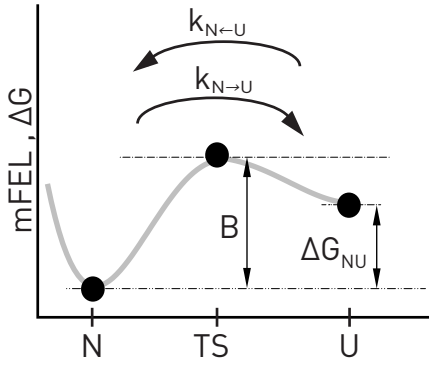

Figure 9.1.: **Molecular free energy landscape.** Illustrative free energy landscape with two states, i.e., Native (N), and Unfolded (U) separated by a Transition States TS placed at the top of the kinetic barrier B.

($B \gg k_B T$, being k_B the Boltzmann constant and T the temperature) prevents molecular transitions over measurable timescales, leading to inefficient sampling of the configurational space. Instead, non-equilibrium

experiments facilitate transitions over large kinetic barriers, providing an alternative and efficient way to sample the mFEL.

Two widely used phenomenological approaches to extract equilibrium information from pulling experiments are the Bell-Evans (BE) [85, 86] and the kinetic diffusive (KD) models. The BE model describes mechanically induced folding/unfolding transitions as thermally activated processes over a transition state energy barrier. The BE model assumes that, for a fixed transition state position, the height of the kinetic barrier decreases linearly with the applied force, $B = B_0 - fx^\dagger$ (being $x^\dagger$ the distance from the folded to the transition state). This assumption is relaxed in the KD model, which considers the folding reaction as a diffusive process in a one-dimensional force-dependent mFEL. While the BE model only considers the height and position of the transition state, the full description of the mFEL in the KD model requires the knowledge of all the partially folded states. The advantage of the KD as compared to the BE model is its higher predictive power. The same experimental data can be readily employed to extract additional information about the mFEL without the need to adopt the assumptions of the BE model.

9.1.1 *CEBA: A brief remember*

A systematic method to extract information about the mFEL from pulling or hopping experiments in the context of the KD model is the Continuous Effective Barrier Approach (CEBA). CEBA has been used to study RNA hairpins [151], it has been later applied to extract the elastic properties of short RNA hairpins at different ionic conditions [152], the thermodynamic and kinetic properties of protein Barnase [87], and DNA hairpins with different mechanical fragilities [130].

In CEBA, the force-dependent effective barrier between the native (N) and the unfolded state (U), $B_{NU}(f)$, is derived by imposing detailed bal-

ance between the unfolding ($k_{N\to U}(f)$) and folding ($k_{N\leftarrow U}(f)$) kinetic rates:

$$k_{N\to U}(f) = k_0 \exp\left(-\frac{B_{NU}(f)}{k_B T}\right) \tag{9.1a}$$

$$k_{N\leftarrow U}(f) = k_{N\to U}(f) \exp\left(\frac{\Delta G_{NU}(f)}{k_B T}\right) . \tag{9.1b}$$

In equations (9.1a) and (9.1b), k_0 is the attempt rate, and $\Delta G_{NU}(f)$ is the folding free energy at force f. $\Delta G_{NU}(f)$ is defined as $\Delta G^0_{NU} - \int_0^f (x_U(f') - x_N(f'))df'$, with ΔG^0_{NU} the folding free energy difference between N and U at zero force, and $-\int_0^f x_{U(N)}(f')df'$ the free-energy gain to stretch the molecule in state $U(N)$ at force f. Commonly, the elastic response of U and N are modeled using the Worm-Like Chain and Freely-Jointed Chain models [111], respectively. In CEBA we take advantage of the fact that $B_{NU}(f)$ appears in Eqs. (9.1a) and (9.1b) to conveniently rewrite them as,

$$\frac{B_{NU}(f)}{k_B T} = \log k_0 - \log k_{N\to U}(f) \tag{9.2a}$$

$$\frac{B_{NU}(f)}{k_B T} = \log k_0 - \log k_{N\leftarrow U}(f) + \frac{\Delta G_{NU}(f)}{k_B T} \tag{9.2b}$$

Therefore, by assuming the continuity of $B_{NU}(f)$, the difference between $-\log k_{N\to U}(f)$ and $-\log k_{N\leftarrow U}(f) - (1/k_B T)\int_0^f (x_U(f') - x_N(f'))df'$ equals ΔG^0_{NU}. This permits us to derive the folding free energy ΔG^0_{NU} if the elastic response $(x_U(f) - x_N(f))$ is known.

Moreover, by comparing the experimental profile of $B_{NU}(f) - \log k_0$ with the theoretical prediction of the $B_{NU}(f)$ obtained within the KD model one can extract k_0 [130]. The theoretical prediction of B_{NU} for DNA hairpins is calculated as [153]:

$$\frac{B_{NU}(f)}{k_B T} = \log\left(\sum_{m=0}^{M}\sum_{m'=0}^{m} e^{\left(\frac{\Delta G_m(f) - \Delta G_{m'}(f)}{k_B T}\right)}\right) \tag{9.3}$$

where the double sum runs over all hairpin configurations, labeled by m and m', and M being the total number of base pairs (bp).

9.1.2 *Extended-CEBA for intermediate states*

Here, we extend CEBA (eCEBA) to investigate intermediate kinetic states from non-equilibrium experiments and derive the force dependence of the effective kinetic barrier. At the end of this chapter, we also investigate the temperature dependence of the kinetic barrier. For the first purpose, we consider DNA hairpins with one, two, and three intermediates pulled with optical tweezers as helpful models to experimentally investigate intermediate states. The existing knowledge about DNA thermodynamics [127, 128] allows us to accurately predict the force-dependent kinetic barriers of arbitrary DNA hairpin sequences, facilitating the comparison between theory and experiments. The chosen examples span a wide range of situations often encountered in macro-molecular folding. Finally, for the temperature dependence of the kinetic barrier, we have considered two short (20bp) DNA hairpins. The first one folds in a two-state manner. In contrast, the second presents an intermediate state during the folding/unfolding process.

9.2 FORCE-DEPENDENT KINETIC BARRIERS

To investigate the mFEL of molecules involving intermediates we have considered three DNA hairpins (sequences shown in Fig. 9.2). The first hairpin, named *HI*1, has an inner-loop to induce a single intermediate state during folding/unfolding. The second hairpin (*HI*2) consists of a two-hairpin structure with a doubly degenerate intermediate state. The last hairpin (*HI*3) is a three-way junction with three intermediate states: the unfolding until the junction plus the opening of each of the two arms.

The mFEL for DNA hairpins with an intermediate state is depicted in Fig. 9.3. Each black dot denotes a state: native N, intermediate I the unfolded U, separated by a kinetic barrier, B_{NI} and B_{IU} (see Fig. 9.3).

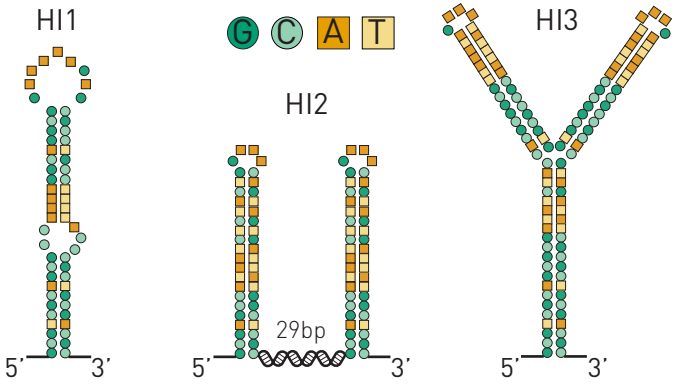

Figure 9.2.: **DNA sequences.** Sequence of the studied DNA hairpins: HI1 (left), HI2 (center), and HI3 (right).

Therefore, CEBA can be extended to include the intermediate state and determine the force-dependent effective barriers $B_{NI}(f)$ and $B_{IU}(f)$ from the kinetic rates by generalizing Eqs. (9.2a), (9.2b):

$$\frac{B_{ij}(f)}{k_BT} = \log k_0^{ij} - \log k_{i\to j}(f) \tag{9.4a}$$

$$\frac{B_{ij}(f)}{k_BT} = \log k_0^{ij} - \log k_{i\leftarrow j}(f) + \frac{\Delta G_{ij}}{k_BT} \ . \tag{9.4b}$$

with $i,j \in \{N,I,U\}$, $k_{i\to j}$ and $k_{i\leftarrow j}$ being the unfolding and folding kinetic rates between i and j, and k_0^{ij} the attempt rate. ΔG_{ij} equals $\Delta G_{ij}^0 - \int_0^f (x_j(f') - x_i(f'))df'$, where $-\int_0^f x_k(f')df'$ is the free-energy gain to stretch state k up to force f, and ΔG_{ij}^0 is the folding free energy difference between states i and j at zero force.

By imposing continuity between the two expressions for $B_{ij}(f)/k_BT - \log k_0^{ij}$ in Eqs. (9.4a) and (9.4b), we derive the free-energy difference between states i and j at zero force ΔG_{ij}^0. Finally, we derive the attempt

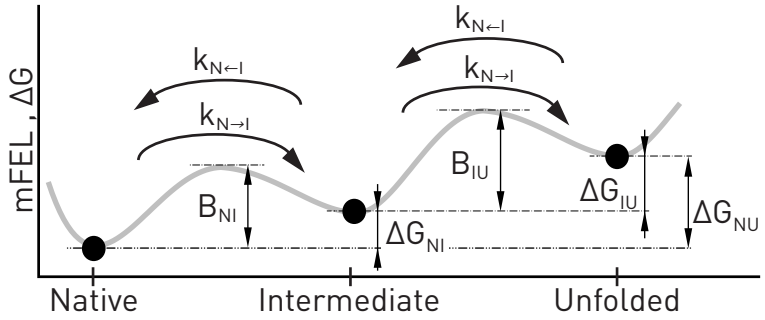

Figure 9.3.: **Molecular free energy landscape.** Illustrative free energy landscape with three states, i.e., Native (N), Intermediate (I), and Unfolded (U) separated by two kinetic barriers B_{NI} and B_{IU}, respectively.

rates k_0^{ij} by matching the experimental results for $B_{ij}(f)$ with the theoretical Kramers prediction, calculated as [153]:

$$\frac{B_{NI}(f)}{k_B T} = \log\left(\sum_{m=0}^{M_I}\sum_{m'=0}^{m} e^{\left(\frac{\Delta G_m(f)-\Delta G_{m'}(f)}{k_B T}\right)}\right) \tag{9.5a}$$

$$\frac{B_{IU}(f)}{k_B T} = \log\left(\sum_{m=M_I+1}^{M}\sum_{m'=M_I+1}^{m} e^{\left(\frac{\Delta G_m(f)-\Delta G_{m'}(f)}{k_B T}\right)}\right) \tag{9.5b}$$

where the double sum runs over all hairpin configurations, labeled by m and m', being M_I the number of unzipped base-pairs (bp) at I and M the total number of bp.

9.2.1 *HI1 hairpin*

First, we have investigated HI1. The theoretical prediction of the mFEL based on the nearest neighbor model [127, 128] calculated at 15pN is shown in Fig. 9.4 together with the expected unfolding/folding pathways.

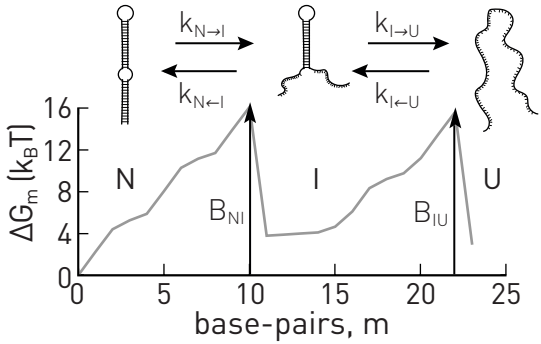

Figure 9.4.: **Molecular free energy landscape of HI1.** mFEL derived using the energy per DNA motif based on the nearest neighbor model [127, 128] at 15pN.

Partial folding and unfolding between states N, I, and U can be observed as sudden drops and rises of force, respectively, in hopping (equilibrium) and pulling (non-equilibrium) experiments (Fig. 9.5). In hopping experiments, the molecule is held at different fixed trap positions (distance) where each observed force level corresponds to a different state (Fig. 9.5a). In pulling experiments, the optical trap is moved back and forth at constant speed to mechanically unfold and fold the hairpin. The different force branches observed in the force-distance curves (FDCs) arise from the elastic response of the hairpin in each state (Fig. 9.5b).

In equilibrium hopping experiments, the kinetic rate $k_{i\to j}$ ($k_{i\leftarrow j}$) equals the inverse of the average lifetime of the state i (j) before jumping to state j (i). The force-dependent rates $k_{N\to I}$, $k_{I\to U}$, $k_{N\leftarrow I}$, and $k_{I\leftarrow U}$ are shown as empty symbols in Fig. 9.6. In non-equilibrium pulling experiments, we determine them from the survival probabilities of each state along the unfolding and folding FDCs.

Here, we provide the mathematical details to determine the survival probability of N, I, and U from pulling experiments. This methodology is general and will be used in the other hairpins. Firstly, we define a threshold force, f_{th}, and measure the first rupture/formation event

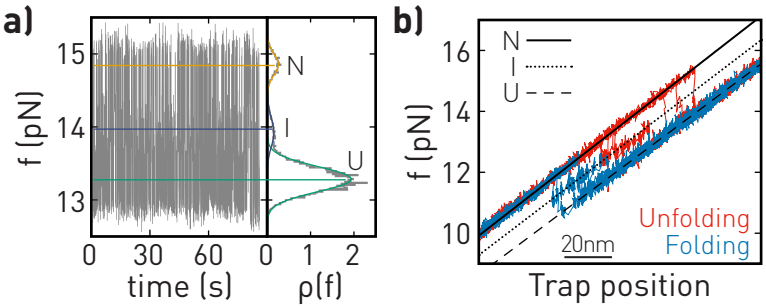

Figure 9.5.: **Hopping and pulling experiments for HI1.** (a) Left: force versus time recorded in hopping experiments. Right: histogram of the force signal showing the three states (N, I, and U). (b) Unfolding (red) and folding (blue) FDCs. The black solid, dotted and dashed lines correspond to the N-, I-, and U-force branches.

taking place at a force above (below) f_{th} for each unfolding (folding) trajectory. Subsequently, we classify the force events as: $f^i_{\rightarrow}$ and $f^i_{\leftarrow}$, where $i = N,\ I$ or U indicates the molecular state at f_{th} and the arrow indicates the direction of the FDCs: unfolding ($\rightarrow$) or folding ($\leftarrow$). For the survival probabilities of the different states, $f^I_{\rightarrow}$ and $f^I_{\leftarrow}$ comprise both rupture and formation events indistinguishably, while $f^N_{\rightarrow}$ and $f^N_{\leftarrow}$ only contain rupture events and $f^U_{\rightarrow}$ and $f^U_{\leftarrow}$ only contain formation events. From the $f^i_{\rightarrow}$ and $f^i_{\leftarrow}$, we calculate the force-dependent survival probabilities conditioned to f_{th} along unfolding ($P^N_{\rightarrow}(f|f_{th}),\ P^I_{\rightarrow}(f|f_{th}),\ P^U_{\rightarrow}(f|f_{th})$) and folding ($P^N_{\leftarrow}(f|f_{th}),\ P^I_{\leftarrow}(f|f_{th}),\ P^U_{\leftarrow}(f|f_{th})$):

$$P^i_{\rightarrow}(f|f_{th}) = 1 - \frac{n(f_{th} < f^i_{\rightarrow} < f)}{n^i_{\rightarrow}} \tag{9.6a}$$

$$P^i_{\leftarrow}(f|f_{th}) = 1 - \frac{n(f_{th} > f^i_{\leftarrow} > f)}{n^i_{\leftarrow}} \tag{9.6b}$$

In equation (9.6a) (Eq. (9.6b)), $n(f_{th} < f^i_{\rightarrow} < f)$ ($n(f_{th} > f^i_{\leftarrow} > f)$) denotes the number of events during unfolding (folding) leaving state i for the first time between f_{th} and f, $n^i_{\rightarrow}$ ($n^i_{\leftarrow}$) is the total number of events

leaving state i above (below) f_{th}. Both quantities are calculated considering the whole set of trajectories. Hence, by repeating this analysis for different values of f_{th}, we reconstruct $P^i_{\rightarrow}(f|f_{th})$ and $P^i_{\leftarrow}(f|f_{th})$ for different force ranges. The survival probabilities satisfy the following master equations:

$$\frac{\partial P^N(f|f_{th})}{\partial f} = -\frac{k_{N\rightarrow I}(f)}{r} P^N(f|f_{th}) \tag{9.7a}$$

$$\frac{\partial P^I(f|f_{th})}{\partial f} = -\frac{k_{N\leftarrow I}(f) + k_{I\rightarrow U}(f)}{r} P^I(f|f_{th}) \tag{9.7b}$$

$$\frac{\partial P^U(f|f_{th})}{\partial f} = -\frac{k_{I\leftarrow U}(f)}{r} P^U(f|f_{th})\,. \tag{9.7c}$$

with boundary conditions $P^{N,I,U}(f_{th}|f_{th}) = 1$. In equations (9.7a), (9.7b), and (9.7c), $k_{N\rightarrow I}(f)$, $k_{N\leftarrow I}(f)$ and $k_{I\leftarrow U}(f)$ are the kinetic rates connecting the molecular states, and f increases/decreases linearly with time at a constant loading rate $r = |df/dt|$.

For a Markovian system Eqs. (9.6a-9.7c) give estimates of $k_{N\rightleftarrows I}$, $k_{I\rightleftarrows U}$ that are independent of f_{th}. Notably by merging results obtained at different f_{th} we enlarge the statistics improving estimates of the kinetic rates.

Therefore, once the survival probabilities are experimentally calculated it is straightforward to determine $k_{N\rightarrow I}(f)$, the sum $k_{N\leftarrow I}(f) + k_{I\rightarrow U}(f)$, and $k_{I\leftarrow U}(f)$. To decouple $k_{N\leftarrow I}(f)$ from $k_{I\rightarrow U}(f)$ we use the following relation:

$$\frac{k_{I\rightarrow U}(f)}{k_{N\leftarrow I}(f)} = \frac{\phi_{I\rightarrow U}(f)}{\phi_{N\leftarrow I}(f)} = \frac{\phi_{I\rightarrow U}(f)}{1-\phi_{I\rightarrow U}(f)} \tag{9.8}$$

where $\phi_{I\rightarrow U}(f)$ and $\phi_{N\leftarrow I}(f)$ are the fraction of transitions leaving I towards U and I towards N, respectively, at force f. Then $\phi_{I\rightarrow U}(f) + \phi_{N\leftarrow I}(f) = 1$. These fractions are experimentally measured on a force window $df = 0.1$pN.

For the particular case of HI1, we have found a good agreement between kinetic rates recovered from equilibrium (empty symbols) and

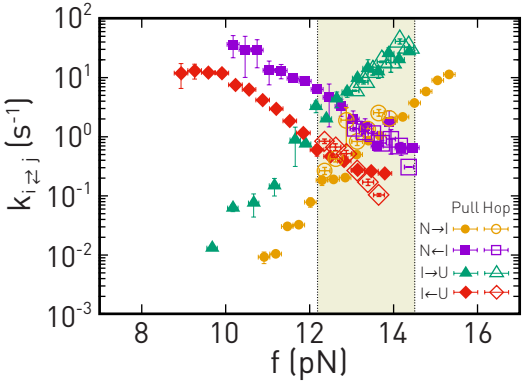

Figure 9.6.: **Unfolding/folding kinetic rates of HI1.** $k_{N\to I}$ (yellow), $k_{N\leftarrow I}$ (purple), $k_{I\to U}$ (green), and $k_{I\leftarrow U}$ (red) measured from hopping (empty symbols) and pulling (solid symbols) experiments.

pulling experiments (solid symbols, Fig. 9.6). Notably, the force range where transitions are observed in equilibrium (highlighted in light yellow) is narrower compared to that from pulling experiments. This shows that non-equilibrium experiments provide a wider force range to investigate kinetic rates.

The resulting barriers measured using Eqs. (9.4a) and (9.4b) are shown in Fig. 9.7 (solid symbol: pulling; empty symbol: hopping). In addition, in Fig. 9.7, we show the theoretical prediction of B_{NI} (solid black line) and B_{IU} (black dashed line). Notice that $\log(k_{\to})$, $\log(k_{\leftarrow})$, B_{NI} and B_{IU} are well approximated with a linear trend in the explored force range, validating the assumptions of the BE model, i.e., constant $x^{\dagger}$ (linearity of $\log(k_{\to})$ and $\log(k_{\leftarrow})$) and a linear relation between the barrier and the applied force (linearity of B_{NI} and B_{IU}). The measured $x^{\dagger}$ considering the BE model equals 6.8 ± 0.2nm. and 6.9 ± 0.3nm for the $N \to I$ and $I \to U$ transitions, respectively. The extracted values for ΔG^0_{ij} and k^{ij}_0 are: $k^{NI}_0 (5 \pm 1) \times 10^7$ s^{-1}, $\Delta G^0_{NI} = 30 \pm 2\ k_B T$, $k^{IU}_0 = (7 \pm 1) \times 10^6$ s^{-1} and $\Delta G^0_{IU} = 27 \pm 3\ k_B T$. These results are compared with theoretical predictions in Tab. 9.1.

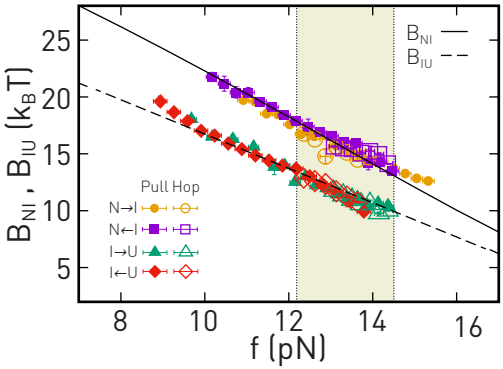

Figure 9.7.: **Effective kinetic barrier of HI1.** B_{NI} (black solid line derived using Eq. (9.5a)) derived from the $k_{N\to I}$ (yellow) and $k_{N\leftarrow I}$ (purple) measured from hopping (empty symbols) and pulling (solid symbols) experiments. B_{IU} (black dashed line derived using Eq. (9.5b)) derived from the $k_{I\to U}$ (green) and $k_{I\leftarrow U}$ (red) measured from hopping (empty symbols) and pulling (solid symbols) experiments.

9.2.2 *HI2 hairpin*

Second, we have investigated the HI2 hairpin, which contains two identical DNA hairpins serially connected and separated by a short (29bp) dsDNA segment (Fig. 9.2). The native state (N) of the hairpin is defined when both hairpins are folded. Hence, N can unfold via two different pathways, each characterized by an intermediate corresponding to the unfolding of one of the two hairpins (Fig. 9.8). However, as both hairpins are identical, they cannot be experimentally distinguished. Therefore, we define a global intermediate I comprising the two intermediates.

The theoretical mFEL for HI2 hairpin considers several configurations compatible with a given number of m unfolded base pairs. Note the total number of open base pairs can distribute between the two hairpins. That is, if m_1 and m_2 are the number of open base pairs in each of the two hairpins, compatible configurations satisfy $m_1 + m_2 = m$. The mFEL for

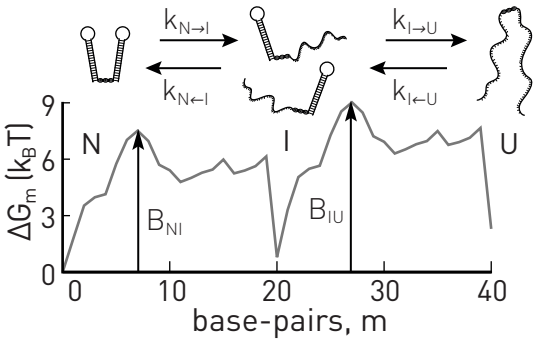

Figure 9.8.: **Molecular free energy landscape of HI2.** mFEL derived using the energy per DNA motif based on the nearest neighbor model [127, 128] at 15pN.

HI2 equals the Boltzmann average taken over all possible configurations at a given m:

$$\frac{\Delta G_m}{k_BT} = -\log\left\{\sum_{m_1=0}^{m}\left[\exp\left(-\frac{\Delta G^{(1)}_{m_1}}{k_BT}\right)+\exp\left(-\frac{\Delta G^{(2)}_{m-m_1}}{k_BT}\right)\right]\right\}. \quad (9.9)$$

The mFEL measured at 15 pN reveals a single intermediate at $m = 20$ in the mFEL (Fig. 9.8), which mainly corresponds to the case where one hairpin is folded, and the other is unfolded.

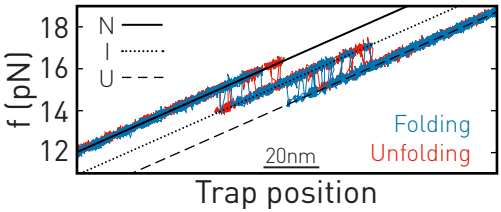

Figure 9.9.: **Pulling experiments using HI2.** Unfolding (red) and folding (blue) FDCs highlighting the N (solid black line), I (dotted black line), and U (dashed black line) force branches.

In Figure 9.9, we show a few unfolding (red) and folding (blue) FDCs. Like HI1, there are three force branches for states N, I, and U. We

derived the kinetic rates following the methodology explained for the previous hairpin, HI1. The resulting force-dependent kinetic rates are shown in Fig. 9.10a. Notice that $\log(k_{\rightarrow})$ and $\log(k_{\leftarrow})$ are approximately linear in the explored force range, i.e., the BE model could be used to derive the $x^{\dagger}$ values. The measured $x^{\dagger}$ for the $N \rightarrow I$ and $I \rightarrow U$ transitions are 10.2 ± 0.5nm and 12.6 ± 0.5nm, respectively. The effective

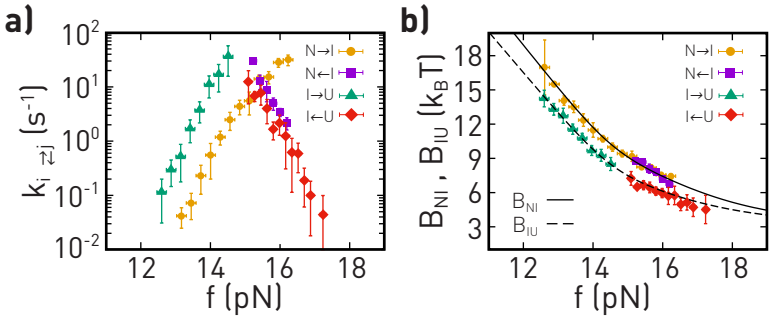

Figure 9.10.: **Results for HI2.** (a) Derived unfolding and folding kinetic rates and (b) effective kinetic barriers of HI2.

barriers B_{NI} and B_{IU} mediating transitions between the three states are shown in Fig. 9.10b. Notice that although the kinetic rates exhibit a linear trend in the log-normal plot, the derived barrier shows a curvature in the explored range. The position of the TS can be determined from the kinetic barrier as $x^{\dagger} = -\partial B/\partial f$. The measured TS distances close to the coexistence force f_c where $k_{\rightarrow}(f_c) = k_{\leftarrow}(f_c)$ are 9.0 ± 0.4nm and 8.5 ± 0.3nm for the $N \rightarrow I$ and $I \rightarrow U$ transitions, respectively. By comparing these values with the ones derived from the BE model, we see that the BE model overestimates $x^{\dagger}$. Moreover, although both hairpins have the same sequence, the barrier that mediates transitions between N and I is different from that between I and U. We have used CEBA to determine the folding free energies and kinetic attempt rates: $k_0^{NI}(5 \pm 1) \times 10^5$ s^{-1}, $\Delta G_{NI}^0 = 54 \pm 2\ k_BT$, $k_0^{IU} = (2 \pm 1) \times 10^6$ s^{-1} and $\Delta G_{IU}^0 = 51 \pm 1\ k_BT$. Finally, the results of ΔG_{ij}^0 and k_0^{ij} are compared with the theoretical predictions in Tab. 9.1 (middle).

9.2.3 *HI3 hairpin*

We have studied a DNA three-way junction (HI3) with three different intermediates (Fig.9.11). The first intermediate state, which we denote as I_1, corresponds to the opening of the first 20bp. Subsequently, as it was the case for molecule HI2, HI3 can take two different pathways to unfold from I_1: $I_1 \to I_2 \to U$ or $I_1 \to I_3 \to U$ depending on which arm is opened. Nevertheless, we cannot distinguish between I_2 and I_3 from the FDCs (Fig.9.12). Therefore, we studied the unfolding and folding

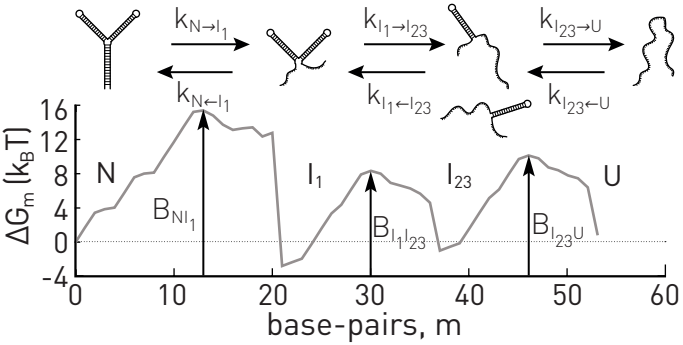

Figure 9.11.: **Molecular free energy landscape of HI3.** mFEL derived at 15pN.

pathway as $N \leftrightarrows I_1 \leftrightarrows I_{23} \leftrightarrows U$. Here, I_{23} comprises I_2 and I_3 where one of the arms is opened. Again, to compute the theoretical mFEL, we assume that several configurations are compatible with a given m after I_1 is fully unfolded (see Eq. (9.9)). Like for HI2, the m unzipped base pairs distribute between the two arms of the HI3 hairpin. The prediction of the mFEL at 15 pN is shown in Fig.9.11 and reveals the two intermediate I_1 and I_{23} states and the barriers B_{NI_1}, $B_{I_1I_{23}}$, and $B_{I_{23}U}$.

In Figure 9.12, we show a few unfolding (red) and folding (blue) FDCs for HI3. For this hairpin, there are four force branches for states N, I_1, I_{23}, and U. In Figure 9.12, the force branch associated to N is represented by the solid black line, the force branch associated to I_1 is represented by

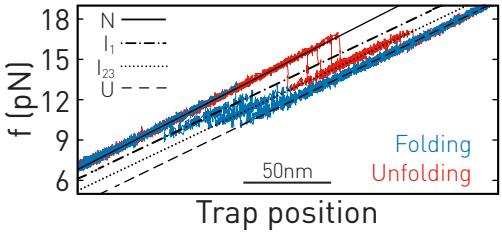

Figure 9.12.: **Pulling experiments for HI3.** Unfolding (red) and folding (blue) FDCs highlighting the N (solid black line), I_1 (dotted-dashed black line), I_{23} (dotted line), and U (dashed black line) force branches.

the dotted-dashed line, the force branch associated to I_{23} corresponds to the dotted line, and the dashed line corresponds to the elastic response of the system when the hairpin is in U. These lines are used to determine the first unfolding/folding event along the FDCs necessary to determine

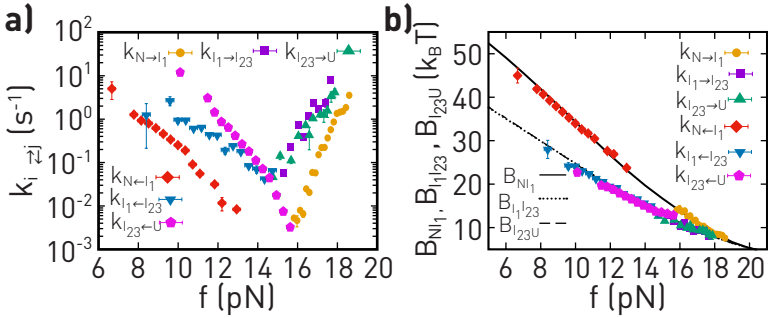

Figure 9.13.: **Results of HI3.** (a) Derived unfolding and folding kinetic rates and (b) effective kinetic barriers for HI3.

the unfolding and folding kinetic rates. The kinetic rates have been derived following the methodology explained for at section 9.2.1.

In Figure 9.13a, we show the six kinetic rates that describe the unfolding and folding pathway of HI3. Notice that while the unfolding rates are linear in a log-normal plot, the folding rates show a subtle concavity, meaning that the BE model would overestimate these rates at low forces.

The derived $x^{\dagger}$ using the BE model equals 10.4 ± 0.4nm, 7.2 ± 0.9nm, and 5.5 ± 0.4nm for the $N \rightarrow I_1$, $I_1 \rightarrow I_{23}$, and $I_{23} \rightarrow U$ transitions, respectively. From these kinetic rates, we have derived the effective barrier to determine the folding free energy and attempt rates between N and I_1, between I_1 and I_{23}, and between I_{23} and U (Tab. 9.1 (bottom)).

In Figure 9.13b, we show the barriers $B_{NI_1}(f)$, $B_{I_1 I_{23}}(f)$, and $B_{I_{23}U}(f)$ obtained from the experimental data together with the Kramers theoretical predictions. According to the Kramers solution (lines in Fig. 9.13b), the kinetic barriers are approximately linear in force. This result is verified by the unfolding kinetic rates that are well fitted with the BE model and the experimentally derived kinetic barrier from the folding kinetic rates. The TS distance measured from Fig. 9.13b ($x^{\dagger} = -\partial B/\partial f$) are 14.0 ± 0.2nm, 8.9 ± 0.2nm, and 7.6 ± 0.2nm for the $N \rightarrow I_1$, $I_1 \rightarrow I_{23}$, and $I_{23} \rightarrow U$ transitions, respectively. By comparing these values with the ones derived from the BE model, we can see that the BE model underestimate these molecular distances. Finally, we emphasize that the barrier mediating transitions between I_{23} and U is similar to those between I_1 and I_{23}. The folding energy and kinetic attempt that match Eqs. (9.4a) and (9.4b) are $k_0^{NI_1} = (6 \pm 1) \times 10^5$ s^{-1}, $\Delta G^0_{NI_1} = 57 \pm 4\ k_B T$, $k_0^{I_1 I_{23}} = (2 \pm 1) \times 10^6$ s^{-1}, $\Delta G_{I_1 I_{23}} = 38 \pm 3\ k_B T$, $k_0^{I_{23}U} = (9 \pm 2) \times 10^5$ s^{-1}, and $\Delta G_{I_{23}U} = 39 \pm 2\ k_B T$.

In Table 9.1 are summarized the kinetic attempts and free-energy differences of the three studied hairpins.

9.3 TEMPERATURE-DEPENDENT KINETIC BARRIERS

In the second part of this chapter, we have investigated the temperature dependence of the effective barrier mediating transitions using two short (20bp) DNA hairpin (see sequences in Fig. 9.14-top). First, we have used a DNA hairpin, *H2s*, that folds and unfolds in a two-state manner. Second, we have employed a DNA hairpin that folds and unfolds via a short lifetime intermediate state. We have named this hairpin as *H3s*.

		k_0^{ij} (s^{-1})	$\Delta G^0_{i,j}$ (k_BT)	
	i, j		Exp.	Prediction
HI1	N, I	$(5 \pm 1) \times 10^7$	30 ± 2	30 ± 1
	I, U	$(7 \pm 1) \times 10^6$	27 ± 3	28 ± 1
	N, I	$(5 \pm 1) \times 10^7$	31 ± 3	30 ± 1
	I, U	$(6 \pm 1) \times 10^6$	28 ± 4	28 ± 1
HI2	N, I	$(5 \pm 1) \times 10^5$	54 ± 2	52 ± 2
	I, U	$(2 \pm 1) \times 10^6$	51 ± 1	55 ± 2
HI3	N, I_1	$(6 \pm 1) \times 10^5$	57 ± 4	52 ± 2
	I_1, I_{23}	$(2 \pm 1) \times 10^6$	38 ± 3	41 ± 2
	I_{23}, U	$(9 \pm 2) \times 10^5$	39 ± 2	40 ± 2

Table 9.1.: Folding free energies and kinetic attempt rates for the three studied hairpins. The results of molecule HI1 in top (bottom) rows correspond to pulling (hopping) experiments.

9.3.1 *Unfolding and folding kinetic rates*

We have carried out equilibrium hopping experiments at different tempearatures in the range $6 - 46^oC$ to measure kinetic rates between the different states (Fig. 9.14-bottom). As previously explained in section 9.2.1, the force-dependent unfolding and folding kinetic rates are measured from the average residence time of each state determined from the force-time traces measured at different trap positions,

$$k_{i\to j} = \frac{1}{\langle \tau_{i\to j} \rangle} \tag{9.10}$$

where $k_{i\to j}$ is the kinetic rate to go from state i to state j, and $\langle \tau_{i\to j} \rangle$ is the lifetime at state i conditioned to jumping to state j. The derived unfolding, $k_{i\to j}$, and folding, $k_{i\leftarrow j}$, kinetic rates of H2s and H3s are shown in Fig. 9.15.

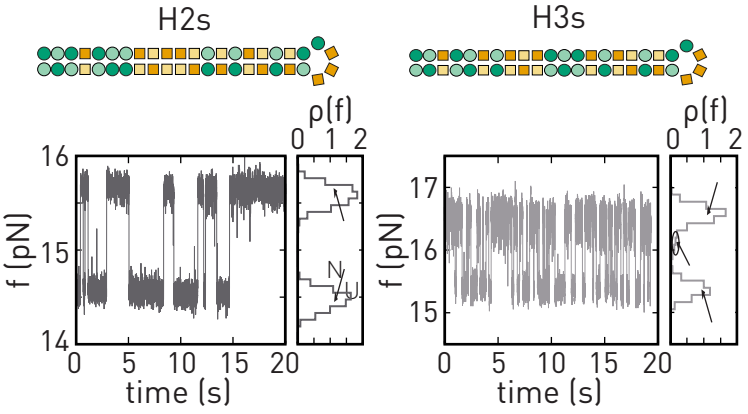

Figure 9.14.: **Hopping experiments with short DNA hairpins.** Top: Sequence of the two short DNa hairpins. Bottom: Force-time trace measured at 25ºC for the H2 and H3s hairpins. The H2s hairpin folds as a two-state manner, while the H3s hairpin folds and unfolds via a short lifetime intermediate state.

The unfolding $k_{N\to U}$ and folding $k_{N\leftarrow U}$ kinetic rates of H2s are shown in Fig. 9.15a. Notice that these rates are the same ones presented in Chap. 7. As we have seen before, these rates are well adjusted with the BE model in the explored temperature range. The unfolding rates $k_{N\to I}$ and $k_{I\to U}$ and folding rates $k_{N\leftarrow I}$ and $k_{I\leftarrow U}$ of H3s at 7, 25 and 45ºC are shown in Fig. 9.15b. The transitions $N \leftrightarrows I$ are represented by the dark colors, while the transitions $I \leftrightarrows U$ are shown with light colors. At first sight, we see that the coexistence force between N and U is higher than the one between I and U. Here, the coexistence force is determined as the force where $k_{i\to j}$ equals $k_{i\leftarrow j}$.

To determine the kinetic barriers mediating the transitions of H2s and H3s, we have used eCEBA in Eqs. (9.2a) and (9.2b) for H2s, and Eqs. (9.4a) and (9.4b) for H3s. The obtained force- and temperature-dependent kinetic barriers are shown in Fig. 9.16. In both cases, the kinetic barriers become more flat as we increase the temperature, keeping the linearity in the explored force range. This result suggests that

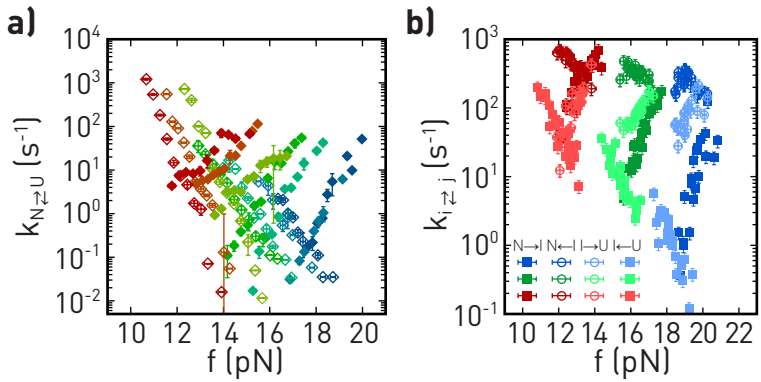

Figure 9.15.: **Unfolding and folding kinetic rates.** (a) Unfolding (solid symbols) and folding (empty symbols) kinetic rates of H2s hairpin measured at 6, 8, 17, 25, 32, 41 and 46°C. Dark blue (red) color symbols correspond to 6°C (46°C). The temperature in-between are plotted following a degradation from blue to red. (b) Unfolding and folding kinetic rates (see legend to differentiate transitions from N or U) measured at 7 (blue), 25 (green), and 45°C (red).

the transition state position in the BE model, is approximately constant with force and different at each temperature. Moreover, the barrier B_{NI} is higher than B_{IU} for H3s (Fig. 9.16b).

Moreover, in Fig. 9.16, we present the derived effective barriers together with the theoretical barriers based on the Kramers solution (solid lines). Mention that the agreement between the Kramers prediction and the experimental results is good in the explored temperature and force ranges.

We show the derived folding free energy ΔG^0_{NU} for the H2s and H3s hairpins in Fig. 9.17a. According to the Gibbs energy definition, $G = H - TS$ where H and S are the enthalpy and entropy, the folding free energy is well adjusted by a linear trend in the assumption of zero or negligible heat capacity change across unfolding/folding. The measured enthalpy and entropy differences are: $\Delta H^0_{NU} = 143 \pm 3$ kcal/mol and

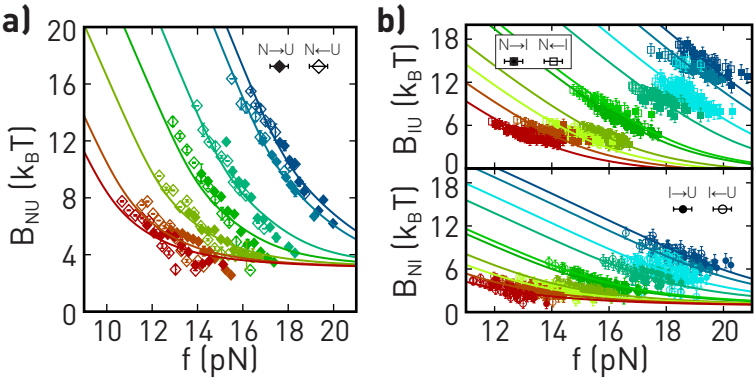

Figure 9.16.: **Temperature dependence of kinetic barriers.** (a) Hairpin H2s between N and U. (b) Hairpin H3s (top: $N-I$, bottom: $I-U$). In both panels, dark blue (red) color symbols correspond to $6-7^oC$ ($46-45^oC$). The temperatures in-between are plotted following a degradation from blue to red (8, 17, 25, 32, 45). The solid lines are the theoretical barriers based on the Kramers solution.

$\Delta S^0_{NU} = 368 \pm 15$ cal/mol·K for H2s and $\Delta H^0_{NU} = 149 \pm 9$ kcal/mol and $\Delta S^0_{NU} = 388 \pm 30$ cal/mol·K for H3s. The values of entropy and enthalpy for both hairpins agree with the estimations based on the nearest neighbor model [127, 128] (Tab. 9.2). Finally, in Fig. 9.17b, we show the free energies differences of H3s, ΔG^0_{NI} and ΔG^0_{IU}, as a function of temperature. As before, we have adjusted the experimental value to a linear function to derive the entropy and enthalpy differences between N and I, and between I and U. The measured entropy and enthalpy are summarized in Tab. 9.2.

9.4 CONCLUSIONS

Here, we have extended CEBA in two directions: First, we have investigated the folding/unfolding transitions involving intermediates states. Second, we have studied the temperature dependence of the kinetic bar-

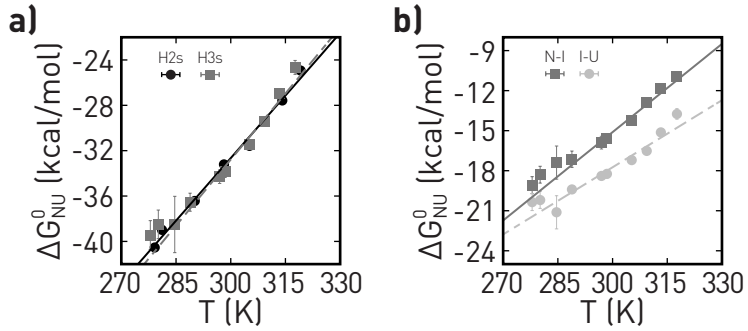

Figure 9.17.: **Temperature-dependent folding free energy.** (a) Folding free energy ΔG^0_{NU} of H2s (black circles) and H3s (gray squares). (b) Partial folding free energy ΔG^0_{NI} (dark gray squares) and ΔG^0_{IU} (light gray circles) of H3s. Solid and dashed lines in panels (a) and (b) are linear fits to the experimental values.

rier. For the first purpose, we considered three DNA hairpins (HI1, HI2, and HI3) with three different kinds of intermediates. These are a hairpin with an inner-loop and a single intermediate (HI1); a two-hairpin structure with a doubly degenerate intermediate (HI2); and a three-way junction with three intermediates (HI3).

To derive the force-dependent unfolding $k_{\rightarrow}$ and folding $k_{\leftarrow}$ kinetic rates of HI1, HI2, and HI3, we have carried out non-equilibrium pulling experiments. To do so, we have introduced a methodology to extract the survival probability of each state, i.e., native N, intermediate I, and unfolded U for arbitrary molecules. Here, we have demonstrated that eCEBA is most convenient approach to investigate molecules involving intermediates and high kinetic barriers where hopping experiments are unfeasible. We have shown that pulling experiments allow us to extract $k_{\rightarrow}$ and $k_{\leftarrow}$ in a broader force range, facilitating the observation of rare transitions.

We have compared the experimentally measured kinetic barriers with the Kramers predictions based on the nearest-neighbor model to validate the methodology. The derived folding free energies of each molecule are

		$\Delta H^0_{i,j}$ (kcal/mol)		$\Delta S^0_{i,j}$ (cal/mol·K)	
	i, j	Exp.	Prediction	Exp.	Prediction
H2s	N, U	143 ± 3	156 ± 15	368 ± 15	419 ± 42
	N, U	149 ± 9	161 ± 16	388 ± 30	427 ± 43
H3s	N, I	82 ± 5	84 ± 8	220 ± 15	223 ± 22
	I, U	68 ± 4	77 ± 8	167 ± 14	204 ± 20

Table 9.2.: Enthalpy and entropy differences derived using CEBA.

compared with the theoretical prediction and summarized in Tab. 9.1. Additionally, the good agreement observed between the experimentally determined force-dependent kinetic barriers and those predicted by the KD model allowed us to estimate values for the attempt rates for native, intermediate and unfolded states. These values are essential for molecular dynamic simulations, where timescales need to be appropriately set.

Second, we have investigated the temperature dependence of the kinetic barriers of hairpins H2s and H3s. Both hairpins are formed by 20bp with 10 and 11 GC bp, respectively. As both are short hairpins, we have carried out hopping experiments to derive their folding and unfolding kinetic rates. On the one hand, H2s is a two-state folder, i.e., it has one barrier mediating transitions between N and U. On the other hand, H3s has an intermediate state of short lifetime mediating transitions between N and U. Therefore, H3s has two unfolding ($k_{N\to I}$ and $k_{I\to U}$) and folding ($k_{N\leftarrow I}$ and $k_{I\leftarrow U}$) kinetic rates. Using eCEBA, we have derived the partial folding free energies ΔG^0_{NI} and ΔG^0_{IU} of H3s, and the folding free energy ΔG^0_{NU} of H2s and H3s.

Assuming a linear temperature dependence of ΔG^0_{ij} ($\Delta G^0_{ij} = \Delta H^0_{ij} - T\Delta S^0_{ij}$), we have derived the folding entropy and enthalpy difference between states i and j at the melting temperature T_m ($T_m = \Delta H^0_{ij}/\Delta S^0_{ij}$). Notice that in this analysis we assume temperature independent values for ΔH^0_{ij} and ΔS^0_{ij}. In other words, we are imposing $\Delta C^{ij}_p = 0$. We have

compared the experimental values of ΔH^0_{ij} and ΔS^0_{ij} with the predictions based on the nearest-neighbour nucleic acids database [127, 128] and previous experiments carried out using unzipping experiments varying the melting temperature [116, 129]. We find good agreement (Tab. 9.2) with our results and the predictions based on the reported values.

To determine the temperature-dependent entropy and enthalpy is necessary to determine first the attempt rate k_a. Here, we have derived the value of k_a by matching the experimental values of $\log(k_{\rightarrow})$ with the theoretically predicted kinetic barrier based on the Kramers solution using the energy values of the nearest-neighbour model from the existing databases [116, 127–129]. Energy parameters in these databases assume that the enthalpy and entropy are constant with the temperature, i.e., $\Delta C_p = 0$. Therefore, to fit the experimental kinetic rates, all temperature dependence goes into k_a, which is not realistic. In fact, the temperature dependent kinetic rates can be used to determine ΔC_p by using a temperature-independent attempt rate. This analysis is left for the future.

10

NETROPSIN BINDING STUDY

MOTIVATION

Single-molecule techniques are powerful tools to investigate the thermodynamics of intra and inter molecular interactions. Here, we want to go a step further, implementing a new approach to investigate drugs that specifically and non-specifically bind to dsDNA molecules.

10.1 INTRODUCTION AND HISTORICAL CONTEXT

Molecular binding to DNA is crucial in chemical therapies [154, 155] and many biological processes, e.g., DNA polymerases in DNA replication and repair. The most usual binding interactions are intercalation and major/minor groove binding. Intercalators are the most common small aromatic molecules that link to DNA, modifying the double-helix structure [156, 157]. On the other hand, major/minor groove binders are positively charged molecules that attach to DNA by hydrogen bridges, electrostatic or Van der Walls bonds in AT-rich DNA's regions [154, 158]. These interactions with DNA have been widely studied by bulk techniques, e.g., DNA footprinting [159, 160], NMR spectroscopy [161, 162], and X-ray cristallography [155, 163]. However, since the invention of single-molecule techniques, new approaches have emerged.

Nowadays, single-molecule experiments (SMEs) are used to characterize binding interactions by mechanically unzipping molecules with proteins or intercalators attached [110, 157]. In particular, in [110], an extension of the Crooks-fluctuation theorem to derive binding energies was presented. In this method, unfolding trajectories are classified into native unbound, native bound and unfolded. For ligands that weakly bind to DNA, it is difficult to discriminate between native bound and unbound trajectories, making difficult the applicability of the fluctuation theorem. On the other hand, it is possible to get information of binding energies from equilibrium hopping experiments as it is shown in [164]. In these kinds of experiments, binding events are monitored in the folding/unfolding time scales. Therefore, many molecules that link DNA cannot be studied using hopping experiments due to their high binding rates (k^{on} and k^{off}).

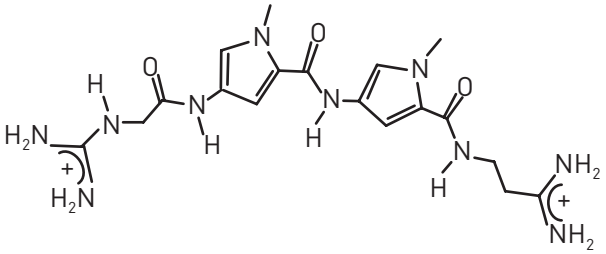

Figure 10.1.: **Molecular structure of netropsin.**

Here, we investigate the binding interaction of the minor groove binder netropsin from non-equilibrium pulling experiments using DNA hairpins. Netropsin (see Fig. 10.1) is a small drug with antiviral, antitumor, and antibiotic properties which is highly toxic for clinical uses. Netropsin binds to dsDNA, interfering with its replication and transcription, but it does not bind to ssDNA [158, 165, 166]. Bulk studies based on NMR, X-ray crystallography, circular dichronism, and calorimetric assays pointed out that the binding of netropsin to the minor groove of DNA is directed by the formation of hydrogen bonds and Van der Waals interactions along AT domains. However, some studies claim that

netropsin binds dsDNA with a preference for the tetranucleotide motif 5′-AATT-3′ [165, 166]. Moreover, the reported binding rates of netropsin, k^{on} and k^{off}, are in the order of 10^7 s^{-1} M^{-1} and 10^{-2} s^{-1}, respectively [167–169]. Therefore, netropsin is an excellent candidate to study binding to DNA via two modes: the specific (5′-AATT-3′) and the non-specific (AT-rich domain) binding modes.

To study the two binding modes, we have used three different short ($10 - 20$bp) DNA hairpins (sequences in Fig. 10.2). The first hairpin, Hns, has been synthesized with an AT-rich zone in the middle of the stem. The second hairpin, Hs, contains a 5′-AATT-3′ motif in the middle of the stem. The last studied hairpin, poly(GC), contains only GC base pairs in the stem. These sequences are specifically chosen to discriminate non specific binding of netropsin to an AT-rich domain in Hns, specific binding to the AATT motif in Hs, and no binding in GC-rich regions.

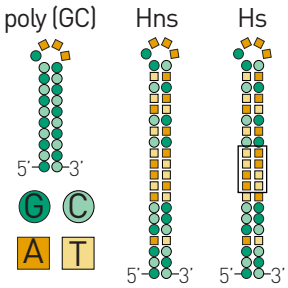

Figure 10.2.: **Sequences of the short DNA hairpins.** The poly(GC) hairpin only contains GC bp in the stem. The stem of hairpins Hns and Hs are identical except by the specific 5′-AATT-3′ motif (framed bps).

10.2 RESULTS

10.2.1 *Short hairpins*

We have carried out pulling experiments at different netropsin concentrations (from 0 to 10nM) using short (less than 50bp) DNA hairpins to derive the binding and unbinding kinetic rates k^{on} and k^{off} of netropsin. In such experiments, the central channel of the microfluidic chamber is filled with a buffer solution containing the desired netropsin concentration.

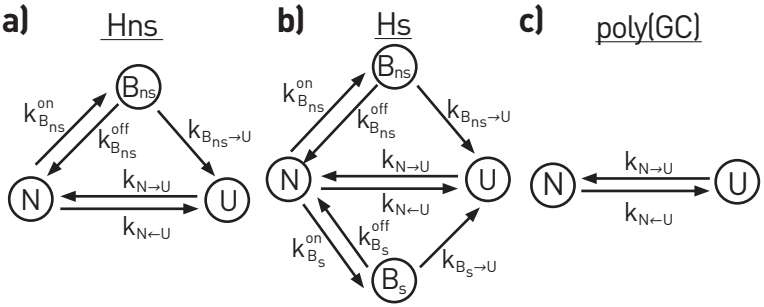

Figure 10.3.: **Unfolding/folding pathways.** Pathways in the presence of netropsin for Hns (a), Hs (b), and poly(GC) (c). $N \equiv$native; $U \equiv$unfolded; $B_{ns} \equiv$ non-specifically bound; $B_s \equiv$ specifically bound.

In the presence of netropsin, Hns can unfold from the native state, N, or from a non-specific bound state, B_{ns} (Fig. 10.3a); Hs can unfold from N or from two bound states: the non-specifically bound state, B_{ns}, or the specifically bound (to the AATT motif) state, B_s (Fig. 10.3b); in the absence of binding, the poly(GC) sequence is expected to unfold from N with or without netropsin (Fig. 10.3c). In pulling experiments, hairpins are mechanically pulled from an initial low force value (typically $3-5$pN), where the hairpins are folded, to a maximum force (typically $\sim$ 25pN) where the hairpins are unfolded (U) (see Fig. 10.4). From the

unfolding Force-Distance Curves (FDCs) (dark color curves in Fig. 10.4), the unfolding events are recorded as sudden force rips. In contrast, from the folding FDCs (light color curves in Fig. 10.4), the folding events are recorded as sudden force jumps. As expected, the poly(GC) at 10nM and 0nM netropsin concentration do not show different FDCs, indicating that netropsin does not bind to GC-rich regions.

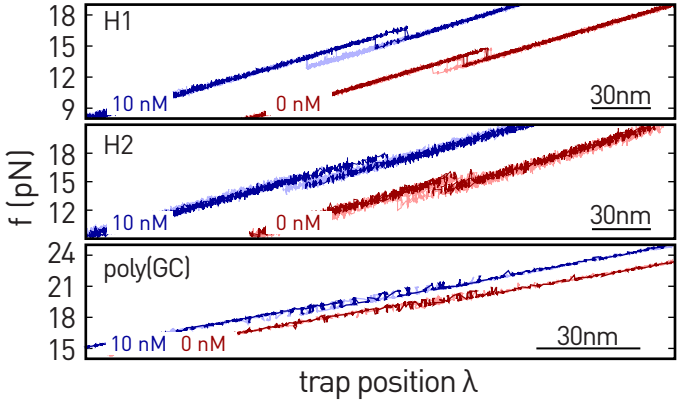

Figure 10.4.: **Force-Distance curves.** Unfolding (dark color) and folding (light color) FDCs measured at 0nM (red) and 10nM (blue) for Hns (top), Hs (middle), and poly(GC) (bottom).

In Figure 10.5 are shown the first unfolding and folding force distributions for all the studied netropsin concentrations. The first unfolding and folding forces are the first force where the molecule changes state in the unfolding ($N \rightarrow U$) and folding ($N \leftarrow U$) trajectories, respectively. Notice that the folding force distributions are not affected by the presence of netropsin, suggesting that the ligand does not interact with the (unfolded) ssDNA. In contrast, the unfolding force distributions shift to higher forces when netropsin concentration increases. Nevertheless, as we see in Fig. 10.5-bottom, the unfolding force distributions for Hns or Hs do not show two or more well-defined populations, corresponding to the unfolding and unbinding events. Therefore, we cannot separate the events where the hairpins are in N, B_{ns}, or B_s. This fact makes the clas-

sification of these events difficult, compared to other ligands that show distinct binding modes [110, 157].

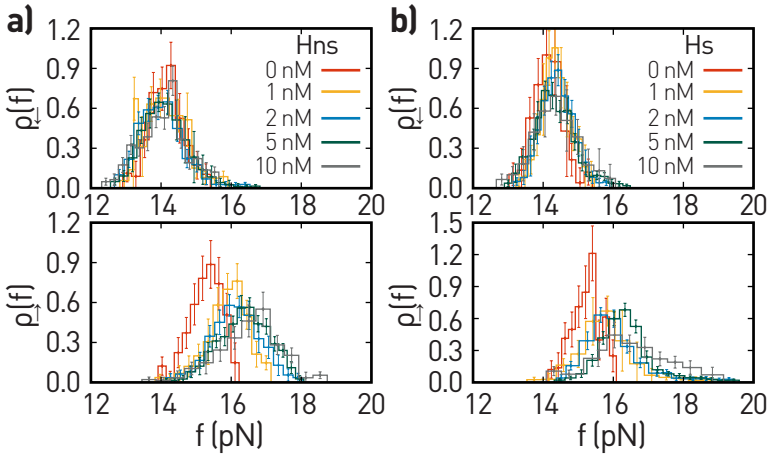

Figure 10.5.: **Unfolding and folding force distributions.** First unfolding (bottom) and folding (top) force distributions at 0 (red), 1 (yellow), 2 (blue), 5 (green), and 10nM (gray) netropsin concentrations for Hns (a) and Hs (b).

In consequence, to study the unfolding trajectories, we have defined a folded state F containing N and B_{ns} for Hns ($F = N \cup B_{ns}$) and N, B_{ns}, and B_s for Hs ($F = N \cup B_{ns} \cup B_s$).

We have derived the unfolding and folding kinetic rates from the force-dependent survival probabilities of the folded ($P_F(f)$) and unfolded ($P_U(f)$) states along the unfolding and folding trajectories, respectively. The survival probabilities for pulling experiments are estimated as,

$$P_F(f) = 1 - \frac{n(f^{\dagger} < f)}{n_{traj}} \tag{10.1a}$$

$$P_U(f) = 1 - \frac{n(f^{*} > f)}{n_{traj}}\,, \tag{10.1b}$$

where $f^{\dagger}$ (f^{*}) in Eq. (10.1a) (Eq. (10.1b)) denotes the first unfolding (folding) force along the unfolding (folding) FDCs. $n(f^{\dagger} < f)$ ($n(f^{*} > f)$) is the number of first unfolding (folding) forces below (above) the force f, and n_{traj} is the total number of trajectories. The measured survival probabilities for poly(GC) are shown in Fig. 10.6a.

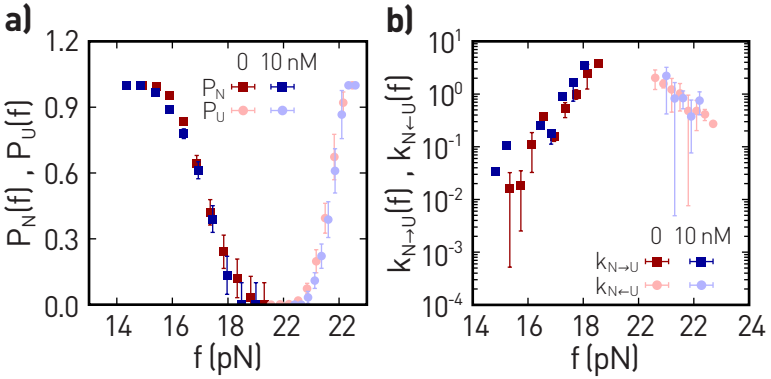

Figure 10.6.: **Survival probabilities and kinetic rates of poly(GC).** (a) Survival probability of N (dark color) and U (light color) measured at 0nM (red) and 10nM (blue). (b) Derived kinetic rates using the survival probabilities of panel (a) and Eqs. (10.2a) and (10.2b).

As mentioned above, netropsin should not bind to poly(GC). Hence, state F only includes N for poly(GC). Figure 10.6 shows the survival probabilities and unfolding and folding kinetic rates of poly(GC). The unfolding and folding kinetic rates are derived using Eqs. (10.2a) and (10.2b), respectively. The presented results demonstrate that netropsin does not bind to ssDNA molecules and dsDNA GC bp regions.

Next, using the derived $P_F(f)$ and $P_U(f)$, we have calculated the force-dependent kinetic rates unfolding, $k_{F\to U}$, and folding, $k_{N\leftarrow U}$, by considering the master equations for unfolding and folding a first order Markov process,

$$k_{F\to U}(f) = r\frac{\rho_{to}(f)}{P_F(f)} \tag{10.2a}$$

$$k_{N\leftarrow U}(f) = r\frac{\rho_{\leftarrow}(f)}{P_U(f)} \ . \tag{10.2b}$$

In equations (10.2a) and (10.2b), r is the loading rate ($r \sim 4-5$ pN/s in the present experiments), and $\rho_{\rightarrow}(f)$ ($\rho_{\leftarrow}(f)$) is the unfolding (folding) first rupture force distribution shown in Fig. 10.5-bottom (top). Notice that the unfolding kinetic rates always start at the folded F macro-state, which comprises states N, B_{ns} and B_s: $F : N \cup B_{ns} \cup B_s$. In contrast, the folding kinetic rate always starts at U and end at N because, as shown in Fig. 10.6, netropsin does not bind to ssDNA. The unfolding and folding kinetic rates of Hns and Hs are shown in Fig. 10.7.

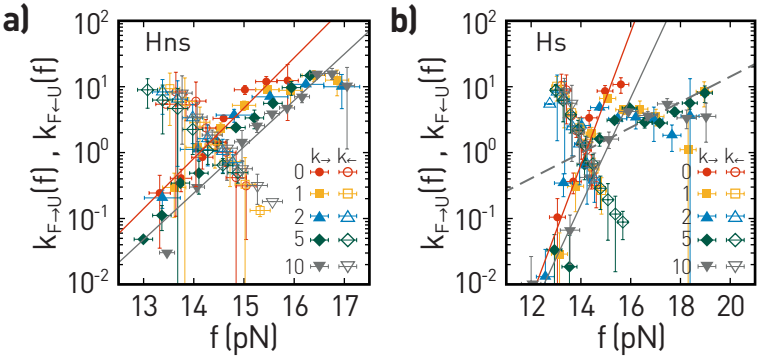

Figure 10.7.: **Kinetic rates of Hns and Hs.** Unfolding (solid symbols) and folding (empty symbols) kinetic rates measured at 0 (red), 1 (yellow), 2 (blue), 5 (green), and 10nM (gray) netropsin concentrations. The solid and dashed lines in panels (a) and (b) are fits to the Bell-Evans kinetic model.

At first sight, we discern three featyres in Fig. 10.7. First, the folding kinetic rates of Hns and Hs are not affected by netropsin, like we have seen for poly(GC), i.e., netropsin does not interact with ssDNA, and therefore with U. Second, the unfolding kinetic rates of Hns at 0 and 10nM grow exponentially with force in the explored force range as predicted by the Bell-Evans model. In contrast, the kinetic rates at the other concentrations are not properly fitted with the expected exponen-

tial behavior. This is indicative of a mixture of two typos of folded states: $F = N \cup B_{ns}$. Third, for Hs a complex force dependence is observed for $k_{F\to U}$ with netropsin. In particular, the slope of $\log(k_{F\to U})$ versus f decreases above $\sim$ 16pN indicating two binding modes for netropsin. To infer the binding kinetic rates for both binding modes, we solved the master equations governing the transitions $F \to U$ observed in the pulling experiments.

Hns hairpin

The survival probability $P_F(f)$ for Hns is defined as the sum of the survival probabilities of N and B_{ns}, i.e., $P_F(f) = P_N(f) + P_{B_{ns}}(f)$. Where $P_N(f)$ and $P_{B_{ns}}(f)$ satisfy the following master equations:

$$r\frac{dP_N(f)}{df} = -(k_{N\to U}(f) + k^b_{ns})P_N(f) + k^{off}_{ns}P_{B_{ns}}(f) \tag{10.3a}$$

$$r\frac{dP_{B_{ns}}(f)}{df} = k^b_{ns}P_N(f) - (k_{B_{ns}\to U}(f) + k^{off}_{ns})P_{B_{ns}}(f)\ . \tag{10.3b}$$

In equations (10.3a) and (10.3b), $k^b_{ns} = c \cdot k^{on}_{ns}$ where k^{on}_{ns} is the association rate and c the netropsin concentration. k^{off}_{ns} is the dissociation rate, which is independent of the concentration. The sub-index ns denotes the non-specific binding mode. Finally, $k_{N\to U}$ is the naked-DNA unfolding kinetic rate, and $k_{B_{ns}\to U}$ is the unfolding kinetic rate for the non-specific bound state. $k_{N\to U}$ and $k_{B_{ns}\to U}$ have been modeled using the Bell-Evans kinetic model [85,86],

$$k_{N\to U}(f) = k_0 \exp\left[\frac{fx^\dagger}{k_BT}\right] \tag{10.4a}$$

$$k_{B_{ns}\to U}(f) = k^{ns}_0 \exp\left[\frac{fx^\dagger_{ns}}{k_BT}\right] \tag{10.4b}$$

where $x^\dagger$ and $x^\dagger_{ns}$ are the transition state position relative to N and B_{ns}, respectively, k_0 (k^{ns}_0) is the unfolding kinetic rate at zero force for the N

$(B_{ns}) \rightarrow U$ process, k_B is the Boltzmann constant, and T is the temperature.

First, we have characterized $x^{\dagger}$ and k_0 by fitting the experimental values of $k_{N\rightarrow U}$ obtained at 0nM to Eq. (10.4a) (solid red line line in Fig. 10.7a). Then, we determined $x^{\dagger}_{ns}$ and k_0^{ns} by fitting Eq. (10.4b) to the values of $k_{F\rightarrow U}$ measured at 10nM (solid grey line line in Fig. 10.7a). As mentioned before, the force behavior of $k_{F\rightarrow U}(f)$ at 10nM follows the exponential trend of the Bell-Evans model while intermediate concentrations not. This fact suggests that at 10nM, netropsin is nearly saturated for this short DNA hairpins, i.e., all the unfolding trajectories at 10nM start at B_{ns}. The measured parameters that best fit the experimental data are $x^{\dagger} = 8.8 \pm 1.1$ nm, $k_0 = (2.2 \pm 0.9) \times 10^{-14}$ s^{-1}, $x^{\dagger}_{ns} = 8.5 \pm 1.0$nm, and $k_0^{ns} = (1.7 \pm 0.8) \times 10^{-14}$ s^{-1}, respectively.

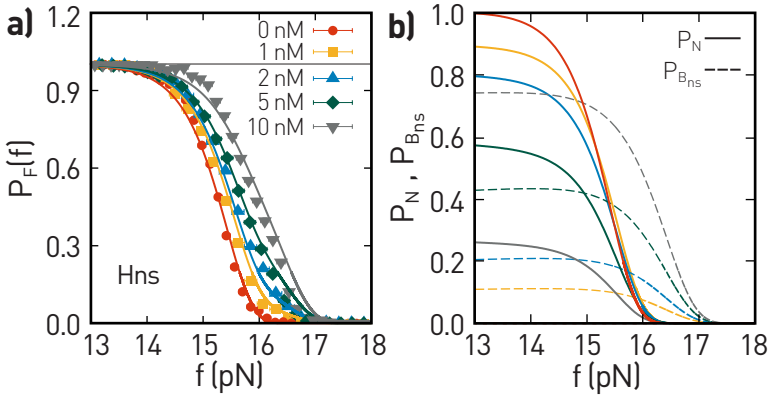

Figure 10.8.: **Unfolding survival probabilities of Hns.** (a) Folded survival probability along unfolding (solid symbols) measured at 0 (red), 1 (yellow), 2 (blue), 5 (green), and 10 nM (gray) netropsin concentrations. The solid lines are fits to Eqs. (10.3a) and (10.3b). (b) Derived P_N (solid lines) and $P_{B_{ns}}$ (dashed lines).

Using the derived values of $x^{\dagger}$, $x^{\dagger}_{ns}$, k_0, and k_0^{ns}, we have solved and fitted Eqs. (10.3a) and (10.3b) simultaneously at all studied concentrations to derive the values of k_{ns}^{off} and k_{ns}^{on}. We have numerically integrated Eqs.

(10.3a) and (10.3b) from an initial force equal to $-\infty$ up to a maximum force of 20pN using the Euler method procedure for solving differential equations. Nevertheless, to use this approach, it is necessary to define an initial condition. From the experiments, we have concluded that hairpins always fold into N during the refolding process, and netropsin binds only afterward. Hence, we can assume that at the most probable folding force ($\sim$ 14.5pN), the molecule is in N. Notice that $P_N = 1$ at 14.5pN would be only valid in the folding process. Therefore, to integrate Eqs. (10.3a) and (10.3b), we have considered that the unfolding process starts at $f_{min} = -14.5$pN and the force is ramped at the loading rate r until f_{max}. Between f_{min} and the unfolding event there is a lapse of time during the while netropsin can bind to N. Therefore, the initial condition to numerically integrate the master equations are $P_N(-14.5) = 1$ and $P_{B_{ns}}(-14.5) = 0$.

The measured $P_F(f)$ for Hns are shown in Fig. 10.8a together with the derived values solving Eqs. (10.3a) and (10.3b). In figure 10.8b, we show the derived values of $P_N(f)$ and $P_{B_{ns}}(f)$ as solid and dashed lines, respectively. As can be seen, for a given force value, when the netropsin concentration increases, the probability to be in B_{ns} increases too. In particular, at 10nM, we can see that $P_{B_{ns}}$ at the most probable unfolding force is close to $\sim$ 75%, supporting our hypothesis that netropsin is close to being saturated at this c. The values of k_{ns}^{off} and k_{ns}^{on} that best fit the experimental data are summarized in Tab. 10.1.

Hs hairpin

Second, we have characterized the specific AATT binding mode using the Hs hairpin. As stated before, Hs can unfold from N, B_{ns} or B_s states (scheme Fig. 10.3b), here the subindex ns denotes the non-specific interaction and s the specific one. The folded sate F comprises N, B_{ns}, and B_s. The survival probability $P_F(f)$ is given by $P_F(f) = P_N(f) + P_{B_{ns}}(f) +$

$P_{B_s}(f)$, where $P_N(f)$, $P_{B_{ns}}(f)$, $P_{B_s}(f)$ satisfy the following master equations:

$$r\frac{dP_N(f)}{df} = -(k_{N\to U}(f) + k^b_{ns} + k^b_s)P_N(f) + k^{off}_{ns}P_{B_{ns}}(f) + k^{off}_s P_{B_s}(f) \tag{10.5a}$$

$$r\frac{dP_{B_{ns}}(f)}{df} = k^b_{ns}P_N(f) - (k_{B_{ns}\to U}(f) + k^{off}_{ns})P_{B_{ns}}(f) \tag{10.5b}$$

$$r\frac{dP_{B_s}(f)}{df} = k^b_s P_N(f) - (k_{B_s\to U}(f) + k^{off}_s)P_{B_s}(f) \tag{10.5c}$$

where k^b_{ns} and k^b_s are the binding rates, defined as the product of the netropsin concentration and the association rates k^{on}_{ns} and k^{on}_s, respectively. k^{off}_{ns} and k^{off}_s are the dissociation rates. $k_{N\to U}$ is the naked-DNA unfolding kinetic rate, $k_{B_{ns}\to U}$ is the unfolding kinetic rate from the non-specifically bound state, and $k_{B_s\to U}$ is the unfolding kinetic rate from the specifically bound state.

As for the case of Hns, the kinetic rates $k_{N\to U}$, $k_{B_{ns}\to U}$, and $k_{B_s\to U}$ have been investigated using the Bell-Evans model (Eqs. (10.4a) and (10.4b)) and the following are,

$$k_{N\to U}(f) = k_0 \exp\left[\frac{fx^\dagger}{k_B T}\right] \tag{10.6a}$$

$$k_{B_{ns}\to U}(f) = k^{ns}_0 \exp\left[\frac{fx^\dagger_{ns}}{k_B T}\right] \tag{10.6b}$$

$$k_{B_s\to U}(f) = k^s_0 \exp\left[\frac{fx^\dagger_s}{k_B T}\right] \quad . \tag{10.6c}$$

In equations (10.6b) and (10.6c), the sub-indices *ns* and *s* denote the non-specific and specific binding mode. The parameters $x^\dagger$ and k_0 ($x^\dagger = 9.9 \pm 1.0$ nm, $k_0 = (1.0 \pm 0.7) \times 10^{-15}$ s^{-1}), have been derived by fitting Eq. (10.6a) to the measured kinetic rates at 0nM (red solid line in Fig. 10.7b). Notice that the measured $k_{F\to U}$ at 10nM shown in Fig. 10.7b exhibits two linear trends in the log-normal plot. The first trend, between $12 - 16$ pN, has been fitted to Eq. (10.6b) to determine the parameters

$x^{\dagger}_{ns}$ and $k_{0_{ns}}$. The second trend, between $16-18$pN, has been fitted to (10.6c) to determine $x^{\dagger}_{s}$ and k_{0_s}. Both fits are shown as solid and dashed gray lines in Fig. 10.7b, respectively. The numerical values that best fits the experimental values are: $x^{\dagger}_{ns} = 7.7 \pm 0.9$nm, $k^{ns}_0 = (7.1 \pm 1.4) \times 10^{-13}$ s^{-1}, $x^{\dagger}_{s} = 3.1 \pm 0.5$nm, and $k^{s}_0 = (2.5 \pm 0.9) \times 10^{-5}$ s^{-1}.

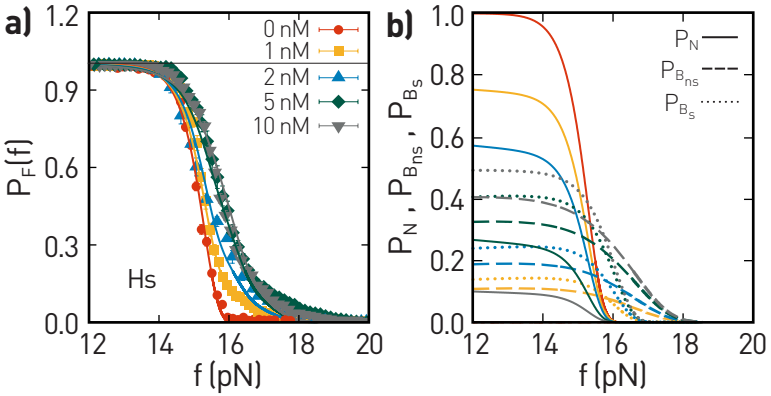

Figure 10.9.: **Unfolding survival probabilities of Hs.** (a) Folded survival probability along unfolding (solid symbols) measured at 0 (red), 1 (yellow), 2 (blue), 5 (green), and 10 nM (gray) netropsin concentrations. The solid are the solved adjustments to Eqs. (10.3a) and (10.3b). (b) Derived P_N (solid lines), $P_{B_{ns}}$ (dashed lines), and P_{B_s} (dotted lines).

Using the present kinetic parameters and the binding kinetic rates of the non-specific bound state derived from Hns, we have fitted and integrated Eqs. (10.5a), (10.5b), and (10.5c) simultaneously using all the experimental data obtained for hairpin Hs. As mentioned previously, DNA hairpins always fold into N during the folding process, and netropsin binds afterward. To numerically integrate the master Eqs. (10.5a), (10.5b), and (10.5c), we have used the Euler method. The initial condition is defined as we did for hairpin Hns. At the most probable folding force ($\sim$ 14.5pN), the hairpin is N with probability equal 1. To integrate the master equations, we have defined $f_{min} = -14.5$pN as the starting initial condition for the unfolding process. Therefore,

$P_N(-14.5) = 1$, $P_{B_{ns}}(-14.5) = 0$, and $P_{B_s}(-14.5) = 0$. The measured force-dependent $P_F(f)$ values are shown in Fig. 10.9a together with the derived values integrating the master equations.

Comparing figures 10.8a and 10.9a, we can see that the $P_F(f)$ values for Hs have a long tail at high values of force, above 16 pN. Notice that this tail, which corresponds to the force range where we have assumed that the specific binding takes place, is approximately independent of the netropsin concentration. Moreover, the survival probability of the bound state ($P_{B_{ns}} + P_{B_s}$) at 10nM is around $\sim$ 90% between 10 and 15 pN. This result further supports the fact that at 10nM binding is nearly saturated. The binding kinetic rates for both modes are summarized in Tab. 10.1. The error bars in Tab. 10.1 correspond to the minimum

	k^{on} (s^{-1} M^{-1})	k^{off} (s^{-1})
B_{ns}	$(2.3 \pm 0.4) \times 10^7$	$(5.2 \pm 0.1) \times 10^{-2}$
B_s	$(8.1 \pm 0.9) \times 10^7$	$(8.4 \pm 0.1) \times 10^{-2}$

Table 10.1.: Derived k^{on} and k^{off} for the two binding modes.

step size of the iterative method that fits and integrates the master Eqs. (10.3a) and (10.3b) for Hns and Eqs. (10.5a), (10.5b), and (10.5c) for Hs. The presented values in Tab. 10.1 are derived by a χ^2 method by defining the relative squared difference between the theoretical curve and the experimental data by minimizing over all data points and netropsin concentrations.

10.3 CONCLUSIONS

We have investigated the binding properties of the minor groove binder netropsin using single-molecule experiments. In particular, netropsin is a paradigmatic drug with a high association, k^{on}, and dissociation, k^{off}, rates that binds to dsDNA with a weak and strong binding modes.

The weak interaction comes from the link to AT-rich DNA's regions. In contrast, the strong interaction results from binding to specific motifs 5'-AATT-3'. Here, we have carried out pulling experiments using three short ($10-20$bp) DNA hairpins (poly(GC), Hns, and Hs) to derive k^{on} and k^{off} for the two binding modes.

From the FDCs for poly(GC) hairpin, we have derived the unfolding and folding kinetic rates at 0 and 10nM netropsin concentrations. The unfolding kinetic rates measured at 0 and 10nM do not present differences, suggesting that netropsin does not bind to GC-rich sequences. In addition, the folding kinetic rates for poly(GC) do not change when adding netropsin, proving that this ligand does not bind to ssDNA molecules. Regarding the Hns hairpin, designed with an AT-rich region at the middle of the stem, we have characterized the non-specific interaction between netropsin and dsDNA. Moreover, Hs was designed with the specific 5'-AATT-3' motif at the middle of the hairpin to study the specific interaction. From the FDCs, we have determined the unfolding and folding kinetic rates and the survival probabilities of the unfolded U and folded F states. F comprises the native N, the specifically bound B_s and the non-specifically bound B_{ns} states, i.e. $F = N \cup B_{ns} \cup B_s$. To infer the association (k^{on}) and dissociation (k^{off}) rates, we have numerically solved the master equations for the survival probabilities of N, B_{ns}, and B_s fitting them to the experimental data for all the studied netropsin concentrations. The obtained k_{ns}^{off}, k_{ns}^{on}, k_s^{off}, and k_s^{on} are summarized in Tab. 10.1. Using the rate constants obtained for both hairpins, we have calculated the dissociation constant, $K_d = k^{off}/k^{on}$, and the binding Gibbs free-energy difference, $\Delta G^0 = k_B T \ln(K_d/c)$. We have estimated a $\Delta G_{ns}^0 = -11.9 \pm 0.4$ kcal/mol for the non-specific binding mode, and a $\Delta G_s^0 = -12.7 \pm 0.4$ kcal/mol for the the specific one at 1M. The value for the non-specific binding mode agrees with the reported one [170]. Notice that the binding energy for the specific and non-specific interaction are equal. In contrast the position of the transition state $x^\dagger$ is different depending on the kind of interaction, $x_{ns}^\dagger \sim 8$nm and $x_s^\dagger \sim 3$nm.

Part VI

CONCLUSIONS AND FUTURE PERSPECTIVES

11

CONCLUSIONS

Since the invention of the laser (acronym of Light Amplification by Stimulated Emission of Radiation) in the 1960s, biology and physics have made transcendent steps toward understanding how life works at the microscopic level (from nucleic acids to proteins and beyond). From the biological point of view, the accurate knowledge of how a polypeptide chain folds into a functional protein structure is crucial to understand some pathologies, e.g., Parkinson's and diabetes mellitus type II. From the physics point of view, the microscopic world is governed by energies of magnitude comparable to thermal energies. Hence, these systems are an ideal playground for developing new mathematical and theoretical tools, e.g., in stochastic thermodynamics.

A relevant tools relevant invention to applications to laser was discovered by Arthur Ashkin in 1970. For his invention, A. Ashkin was awarded the 2018 Nobel Prize in Physics. Optical tweezers permit us to exert forces on individual molecules, such as nucleic acids and proteins. Single-molecule experiments (SMEs) emerged in the 90's and are nowadays commonplace in biophysical laboratories worldwide. SMEs exert forces in the piconewton range (10^{-12} N), which allows us to manipulate molecular constructs with a spatial resolution of nanometers (10^{-9} m). These characteristics enable us to exert work in the $k_B T$ range (10^{-21} J). This energy is on the same order of magnitude as the kinetic energy of the water molecules. Besides, the bonds that stabilize biomolecular structures, such as DNA or proteins, are in the same range. SMEs techniques

are ideal for investigating the folding thermodynamics of nucleic acids and proteins and developing novel theories in the stochastic thermodynamic field. Throughout this thesis, we have made original advances in information theory (Part II) and in biophysics (Parts III, IV, V).

In Chap. 3, we have investigated the touchstone of the thermodynamics information field: the information-to-energy conversion. In addition, we have introduced the paradigm of information-to-measurement conversion in non-equilibrium systems by combining theory and single-molecule pulling experiments to present the first experimental estimation of the efficiency of information-to-measurement conversion based on improved free-energy predictions with feedback. To do so, we have introduced a novel feedback fluctuation theorem for multiple measurements in stochastic systems. We derive a crucial quantity, thermodynamic information Y, which quantifies the amount of reduced work dissipation and the efficiency of information-to-energy (η_I) and information-to-measurement (η_M) conversion. Although feedback reduces dissipation, we have shown that a simple feedback protocol cannot be used for improved free energy determination due to the unexpected inefficiency of information-to-measurement conversion ($\eta_M \ll \eta_I$). We have introduced a novel protocol in the field, which we named feedback strategy. A feedback strategy is a combined series of mixed feedback protocols, which choice depends on the measurement outcome. Our study underlines temporal correlations' role in developing feedback strategies for efficient information-to-energy conversion in small systems. The derivation of a fluctuation theorem for feedback strategies is left for future work.

In Chaps. 4 and 5, we have investigated the molecular free energy landscape (mFEL) of a paradigmatic and well-studied two-state folded protein named Barnase. We have used a temperature controller to mechanically unzip/rezip the protein in a temperature range of $7 - 37^{o}$C. The measured force-distance curves (FDCs) exhibit strong irreversibility in mechanical folding. We derived the temperature-dependent free energy, entropy, and enthalpy difference between the native and unfolded

states by combining fluctuation theorems and the well-known Bell-Evans kinetic model. From their temperature dependence, we derived the total heat capacity change, $\sim$ 1000cal/mol·K. Based on the Bell-Evans kinetic model, we derived the energy, entropy, and enthalpy change across the transition state along the folding process. We find that entropy and enthalpy differences between the transition state and the native state ($\Delta S^\dagger$ and $\Delta H^\dagger$) are much larger than those relative to the unfolded state (ΔS^* and ΔH^*): $\Delta S^\dagger \gg \Delta S^*$ and $\Delta H^\dagger \gg \Delta H^*$. In contrast, the difference in heat capacity change between the transition state and the unfolded state (ΔC_p^{TS-U}) is more significant than the difference relative to the native state (ΔC_p^{N-TS}): $\Delta C_p^{TS-U} \gg \Delta C_p^{N-TS}$. These results support the fundamental hypothesis of the molten-globule and energy landscape models for protein folding. In addition, taking profit of the high irreversibility observed in the force distance curves (FDCs), we have derived the position of the transition state relative to the native ($x^\dagger$) and unfolded (x^*) states based on the Bell-Evans kinetic model. By converting the extension into released amino acids, we have proved that the central assumption of such a kinetic model ($x^\dagger$ and x^* are constant in force) is a direct result of the Leffler-Hammond postulate. In the future, it would be interesting to investigate a protein exhibiting intermediate states along the folding/unfolding process to see if the derived results in this thesis are generally applicable.

In part IV (Chaps. 6, 7, and 8), we have investigated the temperature-dependent thermodynamic, elastic, and kinetic properties of DNA hairpins. In Chap. 6, we have carried out pulling experiments at different temperatures to derive the temperature-dependent elastic properties of DNA. First, we have used short DNA hairpins with different GC content to reconstruct the force-dependent molecular extension of ssDNA molecules. The three studied hairpins are named poly(GC) (100% GC content), poly(AT) ($\sim$ 16% GC content), and mixed (50% GC content). We have proved that the elastic response of the three studied hairpins is the same, demonstrating that the elasticity of ssDNA does not show sequence effects beyond the stacking effects reported for poly-A and poly-G molecules. We have used the inextensible Worm-Like Chain model

(WLC) to determine the persistence length L_p and inter-phosphate distance d_b of ssDNA chains. Mention that L_p is strongly affected by temperature, while d_b is practically constant in the explored temperature range, $5 - 50^{o}$C. Second, we have determined the temperature-dependent elastic properties L_p and d_b of a long dsDNA molecule using stretching experiments. In these experiments, the dsDNA molecule is stretched from its ends. At the same time, the exerted force as a function of the trap position is recorded. Here, we have found that L_p has a strong temperature dependence, whereas d_b remains unchanged by the temperature and equals the crystallographic distance ~ 0.34 nm. Interestingly, we have found that L_p for ssDNA increases when we raise the temperature; in contrast, L_p of the dsDNA is decreases. The temperature dependence of L_p for both molecules is well fitted to a linear function. Besides, it is known that in semi-flexible polymers, e.g., ssDNA and dsDNA, the persistence length has an intrinsic electrostatic contribution that depends on the screening Debye length, which is proportional to $\sqrt{T}$. In the explored temperature range, we have found that L_p is well fitted with a linear or a function of the type, $aT^{1/2} + b$. A discrepancy of only $\sim 10\%$ between both functions is found close to 100^{o}C. In addition, in Chap. 6, we have approached the problem of DNA overstretching varying the temperature. Here, we have shown that the average force at the overstretching transition is proportional to the temperature. Future studies should characterize the overstretching transition using the elastic parameters derived in this chapter to bring light into the still debated overstretching transition's nature.

In Chaps. 7 and 8, we have derived the free energy (ΔG), entropy (ΔS), enthalpy (ΔH) and heat capacity change (ΔC_p) per DNA base pair using the short DNA hairpins from Chap. 6. In Chap. 7, we have derived the thermodynamic potentials from hopping experiments carried out in a wide temperature range, $5 - 50^{o}$C. The three hairpins (poly(GC), poly(AT), and mixed) have different nearest-neighbour base pair (NNBP) motifs containing GC or AT bps in different proportion. The poly(GC) hairpin has been synthesized with only GC bp in the stem. Concretely its stem is formed by five 5'-GC-3' motifs. Therefore, according to the

Nearest-Neighbor model (NN), the poly(GC) hairpin is formed by the NN DNA motifs GC/CG and CG/GC. The second hairpin, poly(AT), is formed by a single 5'-GC-3' motif at the initial bp of the stem, followed by six 5'-AT-3' motifs. Hence, poly(AT) contains a single GC/CG and a CA/GT NN DNA motifs, followed by three AT/TA and TA/AT motifs. The last hairpin, mixed, contains 50% of GC content with two trinucleotide purine-purine motifs an one tetranucleotide purine-purine motif. Here, we have derived ΔG, ΔS, ΔH, and ΔC_p per GC bp and AT bp by dividing the total values by the corresponding number of bps. We have found a remarkable sequence effect in the heat capacity change, which is not apparent in bulk assays due to the difficulty of discerning ΔC_p variations from ssDNA collective effects, such as aggregation and non-specific secondary structure formatio. The values reported here from the poly(GC), poly(AT) and mixed hairpins are: $\Delta C_p = 90 \pm 11$ cal/mol·K, 68 ± 8 cal/mol·K, and 77 ± 7 cal/mol·K per bp, respectively. In addition, we have compared the derived ΔG, ΔS, and ΔH for the mixed hairpin with the averaged values considering the measured energy, entropy, and enthalpy per GC bp and AT bp. We have observed that the energy differences per bp obtained by averaging the values derived from the poly(GC) and poly(AT) hairpins underestimate the values obtained for the mixed hairpin. We hypothesize that such difference arises from the additional seven purine-purine stacks in the mixed hairpin. By comparing the difference in ΔG, we infer that every purine stack (G or A) contributes by roughly 0.3kcal/mol in the stabilization of the dsDNA.

Moreover, in Chap. 7, we have determined the ΔG, ΔS, ΔH, and ΔC_p of two different hairpins with the same stem but different loops. The hairpins are the mixed one (now called mixed-L4), ended with the tetra-loop 5'-GAAA-3', and the same hairpin ended with the octa-loop 5'-GAAAAAAA-3' (mixed-L8). The obtained results from the mixed-L4 and mixed-L8 hairpins suggest that DNA folding follows the same main features of protein folding, i.e., $\Delta S^{\dagger} \gg \Delta S^{*}$, $\Delta H^{\dagger} \gg \Delta H^{*}$), and $\Delta C_p^{N-TS} \ll \Delta C_p^{TS-U}$. In future works, it would be interesting to study larger loops to validate this hypothesis.

In Chap. 8, we have extended the works done in the laboratory [116, 129] by carrying out unzipping experiments varying the temperature. From the obtained data, we derived the free energy difference of the sixteen DNA motifs, i.e., AT/TA, AC/TG, AG/TC, AA/TT, etc. We have compared the measured temperature-dependent ΔG of the different motifs with the prediction of the nearest-neighbour oligonucleotide database [127, 128] and with previous unzipping experiments carried out in the small biosystems lab at different ionic strengths [116, 129]. Our results agree with the existing literature based on the NN model. In addition, we have compared the results from the unzipping experiments with those derived in Chap. 7. Mention that the agreement between both methods opens the door to derive ΔC_p values for the NNBP (which in this thesis was taken to vanish).

Finally, in part V, we present two collaborations done during the thesis. On the one hand, in Chap. 9, we have extended the continuous effective barrier approach (CEBA) to investigate molecules involving intermediates states during folding. To illustrate the advantages of the introduced method, we have determined the folding free energy and kinetic attempt of DNA hairpins with three different kinds of intermediates: a hairpin with an inner-loop and a single intermediate (HI1); a two-hairpin structure with a doubly degenerate intermediate (HI2); and a three-way junction with three intermediates (HI3). In addition, we have used CEBA to determine the temperature-dependent free energy difference of two short DNA hairpins. The first used hairpin (mixed-L4), while the second one, named H3s, has an intermediate state with short lifetimes. We have compared the measured entropy and enthalpy with the predictions based on the NN model using the MFold database [127, 128]. In the future, we plan to develop a method to determine the attempt rate without the necessity to adopt a model to predict the mFEL.

To conclude, in Chap. 10 we have studied the binding of a minor groove binder named Netropsin. From the existing literature, it is known that netropsin binds DNA in two different modes: non-specifically and specifically. The non-specific mode is associated with the binding in

AT-rich regions. In contrast, netropsin specifically binds to 5′-AATT-3′ motifs. Here, we have derived the binding energy and binding rates of netropsin using two 20bp DNA hairpins. One hairpin contains an AT-rich domain in the middle of the stem, and the other contains a 5′-AATT-3′ motif. We have introduced a novel approach to study ligand binding interactions from pulling experiments. In particular, our kinetic based approach can discriminate specific from non-specific binding. From the survival probability of the folded state (compresing the native and ligand bound state), we have derived the dissociation constant, $K_d = k^{off}/k^{on}$, for both binding modes. The measured Gibbs binding energies agree with the reported ones, being -11.9 and -12.7 kcal/mol for non-specific and specific binding modes, respectively, at 1M of Netropsin. In future studies, it would be interesting to study long DNA hairpins in presence of netropsin to investigate the preference binding DNA motifs.

Part VII

APPENDIXES

A

MOLECULAR SYNTHESIS OF DNA AND PROTEIN CONSTRUCTS

This appendix briefly summarizes the steps to synthesize the DNA hairpins and protein used throughout the thesis. This includes short DNA hairpins, long DNA hairpins, long double-stranded DNA chains, and protein Barnase.

A.1 SHORT DNA HAIRPINS WITH SHORT DNA HANDLES

The short DNA hairpins used throughout this thesis have been synthesized using a protocol based on the steps described in reference [113]. The protocol we will describe in this section is valid either for fully-complementary DNA hairpins and hairpins presenting unpaired bases.

DNA hairpins are synthesized by hybridizing the oligonucleotides shown in Fig. A.1. The primary oligonucleotide (black dotted box), which we name *Oligo A*, contains the sequence of the hairpin (which is suitable for each assay) flanked by the sequence of one strand of the DNA handles. The handles are the same for every hairpin and their sequence is:

$$5'-\text{AGTTAGTGGTGGAAACACAGTGCCAGCGC}-3'$$

The second oligo, which we name splint, has the complementary sequence of the handles in oligo A. All these oligonucleotides are supplied by a specialized company, such as Sigma-Aldrich or Invitrogen.

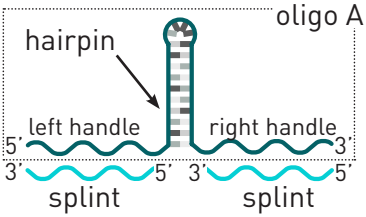

Figure A.1.: **Sketch of the oligonucleotides composing the short DNA constructs.** The box denotes the oligonucleotide containing the DNA hairpin sequence.

The sequences of the used short DNA hairpins are summarized in Tab. A.1. The loop sequences are in bold.

name	sequence
poly(GC)	5'-GCGCGCGCGC **GAAA** GCGCGCGCGC-3'
poly(AT)	5'-GCATATATATATAT **GAAA** ATATATATATATGC-3'
mixed / H2s /H1	5'-GCGAGCCATAATCTCATCTG **GAAA** CAGATGAGATTATGGCTCGC-3'
H3s	5'-GCAGCACATAGCGACATCTG **GAAA** CAGATGTCGCTATGTGCTGC-3'
H2	5'-GCGAGCCTAATTCTCATCTG **GAAA** CAGATGAGAATTAGGCTCGC-3'
HI1	5'-GCGTCGCAGCGCCAAAAGGCAGGCG **GAAA** **AAGAG** CGCCTGCCTTTTACCCGCTGCGACGC-3'
HI3	5'- GCGTCGCAGCGCGATATCTACCACGCGGCG TTAAATA **GAAA** TATTTTAACGCCGCGTGGTGC GGCGCAATTAAT **AAAG** ATTAATTGCGCCGCAC GTAGATATCGCGCTGCGACGC-3'

To connect the molecular construct to the dielectric bead used in the SMEs, the 5'- end of the left handle is labeled with a biotin, whereas

the 3'- end of the right handle is modified with a digoxigenin tail. The biotin labelling is indicated when buying the oligonucleotides, but the digoxigenin tailing is done in the laboratory.

Digoxigenin tailing

Here, we have used the *Oligonucleotide Tailing Kit* (Roche) to tail a Digoxigenin-dUTP at the 3'-end of Oligo A. The steps to do so are the following:

i. Dissolve the supplied oligonucleotide with double distilled water (ddH_2O) until a 100mM concentration is reached. The required water is specified in the oligonucleotide tube. After that, spin down the oligonucleotide tube.

ii. Mix the following components in a sterile Eppendorf tube:

8μl	ddH_2O
1μl	oligo A 100μM
4μl	$CoCl_2$
4μl	Reaction Buffer X5
1μl	dATP
1μl	Digoxigenin-dUTP
1μl	Terminal transferase (enzyme)
20μl	

iii. Incubate for 15 minutes at 37^oC.

iv. Purify the mixture using the *Qiaquick Nucleotide Purification Kit* (QUIAGEN). Follow the instructions of the kit.

v. Keep the final sample at -20^oC to store it.

Annealing the whole molecular construction

The protocol that follows these lines describes how to hybridize the Oligo A and splint to obtain the molecular construct available for performing SMEs.

i. Spin dow all the oligonucleotide tubs (Oligo A obtain from the previous protocol and splint).

ii. Mix the following components in a sterile Eppendorf tube:

10μl	Dig-tailed oligo A (5pmol)
4μl	Splint (10pmol)
1μl	Tris 1M pH 7.5
1μl	NaCl 5M
13μl	ddH_2O
1μl	Digoxigenin-dUTP
30μl	

iii. Incubate for 2 hours at 42°C.

iv. Reduce the temperature at constant rate, 1°C/min, until 25°C is reached.

v. Keep the DNA at 4°C to store it.

A.2 LONG DNA HAIRPINS WITH SHORT DNA HANDLES

The long DNA hairpins used throughout this thesis were prepared as described in reference [157]. The protocol has been designed in a way that it can be used to synthesize hairpins of arbitrary sequence and length using the λ-phage (New England Biolabs) as a template.

The 48kbp λ-DNA template is cutted using the selected restriction enzymes to obtain the desired dsDNA segments, which will be the core of the stem of the hairpin. Moreover, the enzymes have been selected to create long cohesive ends that permit efficient annealing and ligation of the DNA handles and end-loop after restriction. The restriction enzymes used to obtain the 480bp and 3.6kbp are the following: TspRI and Tsp45I for the 480bp (Chap. 10) and the EcoRI for the 3.6kbp (Chap. 8).

The obtained DNA segment is PCR amplified using the designed flanking primers. The desired DNA hairpin is obtained by assembling

five different oligonucleotides. These are: the oligonucleotide containing the loop, the right and left ssDNA handle (HandBio and HandDig, respectively), and the two splint oligos to generate the 29bp dsDNA handles (splint and inverted splint oligos). The resulting structure is shown in Fig. A.2.

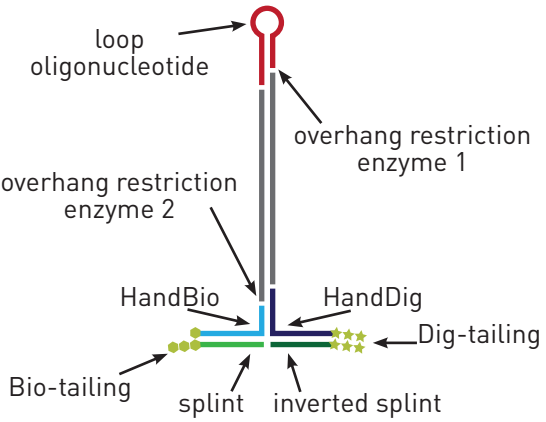

Figure A.2.: **Sketch of the oligonucleotides composing the long DNA constructs.** The overhang restruction is selected specifically to obtain the desired length.

The oligos used to synthesize the 480bp and 3.6kbp DNA hairpins are summarized in Tab. A.2.

Briefly, to create the doubly biotinylated dsDNA handle, the splint oligonucleotide complementary to the unpaired region of HandBio was also purchased. To generate the digoxigenin dsDNA handle, we used a modified oligonucleotide (inverted-splint, *Thermo Scientific*) complementary to the unpaired region of HandDig. This oligonucleotide contains two modifications: a C3 spacer at its 3′- end to block this end in tailing reactions and a polarity inversion at its 5′- end using a 5′-5′ linkage. In this way, both ends of the handle could be tailed with multiple digoxigenins. The oligonucleotides HandDig and inverted-splint are then tailed with digoxigenin-labeled dUTP's and splint with biotin-labeled dUTP's. The

assembly is then ligated in an overnight reaction using T_4 DNA ligase (NEB) at 16°C.

name	sequence
loop oligo: 480bp	5′-Pho-GTC ACT TAG TAA CTA ACA TGA TAG TTAC **TTTT** GTA ACT ATC ATG TTA GTT ACT AA-3′
loop oligo: 3594bp	5′-Pho-AAT TGC CAG TTC GCG TTC GCC AGC ATC CG **ACTA** CGG ATG CTG GCG AAC GCG AAC TGG C-3′
HandBio	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	5′-Pho-AAG ATC TAT TAT ATA TGT GTC TCT ATT AGT TAG TGG TGG AAA CAC AGT GCC AGC GC-3′
Splint	5′-TCC CTA TAG TGA GTC GTA TTA GTG AAG TC-3′
Inverted splint	3′-AAAAA-5′ -5′-GCG CTG GCA CTG TGT TTC CAC CAC TAAC (SpC3)-3′

A.3 BARNASE WITH LONG DNA HANDLES

Protein barnase is inserted between two identical 500bp dsDNa handles following the synthesis steps explained in reference [171]. The protein is attached to the dsDNA handles through disulfide bonds. In order to insert the thiol groups (-S) into the N- and C- terminals of the protein, the initial and the final amino acids of Barnase have been mutated to cysteins. Moreover, the 3′- end of both right and left handles are labeled with one thiol group. The 5′- end of the handles are labeled with a biotin and a digoxigenin molecule to link the molecular construction to the

polystyrene beads used in the SMEs. Furthermore, a tail of histidines is included at the C- terminal of barnase in order to simplify its purification. In what follows we will explain the chemical steps to synthesize the protein barnase flanked with the dsDNA handles.

Barnase expression and purification

1. Bacterial growth
   i. Prepare a pre-culture of BL21 [pET-Bar]: working on ice under sterile conditions, touch the BL21 pET-Bar glycerol from -80°C with a toothpick and inoculate 10ml of LB+00μM ampicillin media. Grow at 250 rpm, 37°C, overnight.
   ii. Inoculate 1l of LB+100μM ampicillin media with 1ml of preculture. Grow at 250 rpm, 37°C, until an O.D. of $0.6 - 0.8$ is reached.
   iii. Induce with 0.5mM IPTG. Let the bacterial cells grow and express the protein during 5h at 37°C.
   iv. Transfer the culture to 80ml centrifuge tubes and centrifuge at 3000 rpm for 15 min at 25°C. Discard the supernatant (i.e., the liquid lying above the solid residue) and repeat the centrifugation until all cells are collected.
2. Bacterial lysis
   i. Resuspend the cells with the Wash-Equilibration Buffer.
   ii. Sonicate on ice.
   iii. Transfer the cell suspension to several 1.5ml Eppendorf tubes. Centrifuge with the microcentrifuge for 10 min at maximum speed. Collect the supernatant (i.e., the lysate containing the soluble over-expressed protein) and discard the cell debris.
3. Barnase-His purification

To carry out purification, the metal affinity chromatography resin from *TALON* (Clontech, Ref. 635501) is used. Co^{2+} has a high affinity for polyhistidine tagged proteins (as is the case of Barnase-His) while it exhibits very low affinity for non-6xHis tagged protein.

i. Load the lysate into the affinity column (follow instructions from the commercial product for its preparation).

ii. Wash resin with Wash-Equilibration Buffer. At this step, Barnase-His and other histidine-rich proteins are attached to the resin.

iii. Elute the protein from the column in 10ml fractions of Elution buffer supplemented with 500mM imidazole.

iv. Analyze the fractions by SDS-PAGE and pool those containing the protein.

v. Concentrate the selected fractions with the protein until $200-400\mu$l.

vi. Load the concentrated protein to a size exclusion. Carry out purification using 100mM NaPi Buffer + 1mM DTT.

vii. Analyze the resulting fraction with SDS-PAGE.

viii. Concentrate the sample again reaching 2ml or less. Add 20 mM DTT and wait 20 minutes.

ix. Carry out a Bradford test to determine the concentration.

4. Barnase-His activation

i. Unfreeze a DTDP stock aliquot and add 8.5ml of 0.1M NaPi.

ii. Add 200μl of freshly prepared DTDP to 200μl of protein. Leave the reaction overnight at 25ºC in the dark.

iii. Carry out a dialysis against 500ml 100mM NaPi buffer for 2 hours. Repeat 3 this step times.

Preparation of the handles

As mentioned before both handles are two identical 500bp segments of a dsDNA molecule. They are amplified by PCR using pBR322 plasmid as template. Both handles are labeled with a Thiol group on the 3'- end, and a biotin or digoxigenin on the 5'- end. These labels are introduced by the oligonucleotides on the PCR reaction. The oligonucleotides are ordered to *MWG Operon/Eurofin*, in a scale of 1μmol and are resuspended in water at a final concentration of 100mM. The oligonucleotides used are:

Bio-500: Bio - GGA ATC TTG CAC GCC CTC GC

Dig-500: Dig - GGA ATC TTG CAC GCC CTC GC

SH-500: Thiol - CAG TTC TCC GCA AGA ATT G

i. Synthesize the Bio-SH 500bp handle: Prepare a 9ml PCR reaction mixing all the components in a 50ml Falcon tube. Transfer 100μl aliquots to PCR tubes and start amplification. Use the standard polymerase from *Biotools*.

ii. Synthesize the Dig-SH handle following the steps described in step i.

iii. Transfer the PCR fractions to a 50ml Falcon tube. Check the PCR with a 1.5% agarose gel.

iv. Use the *Hispeed Plasmid Maxi Kit* (Qiagen, Ref. 12663) starting at step vi and follow the instructions. Perform final elution with 1ml of 15mM NaPi and 3mM DTT.

v. Measure DNA concentration using a biophotometer (Abs260). It is important to obtain a concentration of at least 300μg/ml.

Reaction between protein and handles

i. Mix equal amounts of Bio-SH and Dig-SH handles, typically 0.5ml, trying to get a final concentration of 150μg/ml for each handle ($\sim$ 1μM total DNA).

ii. Concentrate handles mixture with Microcon to a final volume of $50 - 60\mu$l.

iii. Equilibrate 3 Bio-Spin columns with 0.1 NaPi/EDTA (5 x 0.5ml, 1 minute at 1000 g). Exchange the buffer by passing through 3 consecutive Bio-Spin columns.

iv. Prepare reaction:

   a) Place $\sim$ 100pmol of protein in the reaction Eppendorf.

   b) Add 10μl of 500bp DNA handles (just after DTT removal). Reaction starts meanwhile final concentrations are adjusted.

   c) Determine the 500bp DNA handles concentration by Abs260. Use a 1 : 50 dilution. Concentration should be around 20μM.

   d) Adjust the concentration of the 500bp handles and DTDP-protein to $\sim$ 12μM and $\sim$ 4μM respectively. The reaction ratio should be 4 : 1.

v. Let the reaction run overnight.

vi. Analyze the reaction by 4% SDS-PAGE stained with silver nitrate to follow DNA.

In Tab. A.1 are detailed the composition of the buffers required to synthesize the molecular construct with barnase.

LB	Dissolve 10 g Tryptone, 5 g Yeast extract and 10 g NaCl in 900ml. Adjust volume to 1l and autoclave. Store at room temperature ($\sim 25^{o}$C).
Ampillcin 00 mg/ml	Dissolve 1 g of ampicillin in 10 ml of sterile water. Filter the solution through 0.22μm disposable filters under sterile conditions. Store at -20^{o}C in 1ml aliquots.
IPTG 100 mM	Dissolve 238 mg of IPTG into 10ml of sterile water. Filter through a 0.22μm disposable filter under sterile conditions. Store at -20^{o}C in aliquots.
Wash-Equilibration Buffer	Dilute 5x the commercial stock solution (pH 7.5 or pH 8.0) with water. Add 3 volumes of water to 1 volume of 5x commercial buffer. Adjust the pH to 7.5 with HCl and adjust the final volume with water.
Elution buffer	Dilute 10x the commercial stock solution with water.
NaPi	Mix 50 ml of 1M NaH_2PO_4 with 50ml of Na_2HPO_4 and add 350ml of water. Adjust the pH to 7.0 with HCl and bring the final volume to 5000ml. Autoclave and store at 4^{o}C.
NaPi/EDTA	Mix 19ml of 1M NaH_2PO_4 with 10ml of Na_2HPO_4 and 0.5ml of 0.2M EDTA and add 70ml of water. Adjust pH to 8.0 with HCl and level to 100ml. Autoclave and store at 4^{o}C.

Table A.1.: Composition of the buffer required to synthesize protein barnase

B

FIRST-TIME FEEDBACK FT DERIVATION AND CTF LIMIT

Here, we present the derivation of the detailed and full work fluctuation theorem (work-FT) for the first-time feedback (1stTF) protocol. As a corollary, we derive the continuous-time feedback (CTF) limit.

In pulling experiments the force is ramped with a constant loading rate r_F. Measurements are made as a function of time or trap position λ (the natural control parameter in our optical tweezers setup). In the 1stTF protocol measurements are taken at a predetermined set of trap positions along the pulling curve, $\{\lambda_k; 0 \leq k \leq M\}$, i.e., at given times $\{t_k; 0 \leq k \leq M\}$ starting from an initial time, $t_0 = 0$, up to a final time t_M. Therefore, a total number of $M-1$ observations are made for each trajectory (the initial and final times are excluded) implying that $M \geq 2$.

The force and λ limits in pulling experiments are such that the molecule is always folded (F) at λ_0 and unfolded (U) at λ_M. At each predetermined λ_k position the force is measured, and the state of the molecule, F or U, is determined depending on whether it falls in the folded or unfolded branch ($f_F(\lambda)$, $f_U(\lambda)$). Therefore, each stochastic state trajectory Γ is defined by a sequence of F and U, $\Gamma \equiv \{F^0, ...F^{k^*-1}, U^{k^*}, ..., U^M\}$ ($1 \leq k^* \leq M$). The 1stTF protocol changes the loading rate from the initial value r_F to a second value r_U the first time t_{k^*} an unfolding event is observed at λ_{k^*}. It is important to stress the notion of *first time* event. In the above trajectory Γ, the first part of the sequence of measurements

until position λ_k, $\{F^0, ...F^{k^*-1}\}$, only contains F symbols, whereas the second part between λ_{k^*} and the limit λ_M, , $\{U^{k^*}, ...U^M\}$, always starts in U at λ_{k^*} and ends in U at λ_M with either none or multiple (even) hopping transitions ($F \rightleftarrows U$) in-between.

To derive the detailed work-FT, first we define the total forward work probability, $\rho_{\rightarrow}(W|k^*)$, conditioned to the first unfolding event taking place at λ_{k^*} ($1 \leq k^* \leq M$):

$$\rho_{\rightarrow}(W|k^*) = \int \Big(\prod_{k=0}^{k^*-1} \left[\rho_{\lambda_k \rightarrow \lambda_{k+1}}(W_{k+1}) dW_{k+1}\right] \times \rho_{\lambda_{k^*} \rightarrow \lambda_N}(W') \delta(W - W_1 - ... - W_{k^*} - W') \Big) . \tag{B.1}$$

In Eq.(B.1) $\rho_{\lambda_k \rightarrow \lambda_{k+1}}(W_{k+1})$ is the forward work distribution for the section $\lambda_k \rightarrow \lambda_{k+1}$ in the first part of the trajectory $0 \leq k \leq k^* - 1$ and $\rho_{\lambda_{k^*} \rightarrow \lambda_M}(W')$ is the forward work distribution for the second part of the trajectory, $\lambda_{k^*} \rightarrow \lambda_M$. Following the notation in Chap. 3, $\lambda_0 \equiv \lambda_{min}$ and $\lambda_M \equiv \lambda_{max}$.

Second, we define the reverse work distributions along the two parts of trajectory Γ (before and after the first unfolding event at λ_{k^*}). By definition, the reverse process is the time reverse of the forward one, i.e., the loading rate in the reverse process equals r_U between λ_M and λ_{k^*} changing to r_F between λ_{k^*} and λ_0. To determine the reverse work distributions, we used the extended version of Crooks-FT introduced in [50]:

1. Case $0 \leq k < k^* - 1$:

$$\rho_{\lambda_k \rightarrow \lambda_{k+1}}(W_{k+1}) = \rho_{\lambda_k \leftarrow \lambda_{k+1}}(-W_{k+1}) \exp\left[\frac{W_{k+1} - \Delta G_{\sigma_k, \sigma_{k+1}} + k_B T \log\left(\frac{\phi^{\leftarrow}_{\sigma_k}}{\phi^{\rightarrow}_{\sigma_{k+1}}}\right)}{k_B T}\right] \tag{B.2}$$

In equation (B.2), $\phi^{\rightarrow}_{\sigma_{k+1}}$ is the fraction of forward trajectories conditioned to begin at state σ_k in λ_k and end in state σ_{k+1} at λ_{k+1}. $\phi^{\leftarrow}_{\sigma_k}$ is

the fraction of reverse trajectories conditioned to be in σ_{k+1} at λ_{k+1} and end in state σ_k at λ_k. These fractions (forward and reverse) are measured at pulling rate r_F. $\Delta G_{\sigma_k,\sigma_{k+1}} = G_{\sigma_{k+1}}(\lambda_{k+1}) - G_{\sigma_k}(\lambda_k)$ is the partial free energy difference between states σ_{k+1} at λ_{k+1} and σ_k at λ_k. The partial free energy $G_\sigma(\lambda)$ of any state ($\sigma = F, U$) at a given λ equals $G_\sigma(\lambda) = -k_B T \log Z_\sigma(\lambda)$ where $Z_\sigma(\lambda)$ is the partition function restricted to the set of configurations of state σ at the trap position λ. Remember that in this first part of Γ, $\sigma_k = F$ for $0 \leq k \leq k^* - 1$ and $\sigma_{k^*} = U$.

2. Case $k^* \leq k \leq N$:

$$\rho_{\lambda_{k^*} \to \lambda_N}(W') = \rho_{\lambda_{k^*} \leftarrow \lambda_N}(-W') \exp\left[\frac{W' - \Delta G_{\sigma_{k^*},\sigma_N} + k_B T \log\left(\frac{\phi^{\leftarrow}_{\sigma_{k^*}}}{\phi^{\rightarrow}_{\sigma_N}}\right)}{k_B T}\right] \tag{B.3}$$

Here, $\phi^{\rightarrow}_{\sigma_M}$ is the fraction of forward trajectories conditioned to start in σ_{k^*} at λ_{k^*} and end in σ_M at λ_M. $\phi^{\leftarrow}_{\sigma_{k^*}}$ is the fraction of reverse trajectories conditioned to start in σ_M at λ_M and end in σ_{k^*} at λ_{k^*}. In contrast with the previous case, these fractions (forward and reverse) are measured at pulling rate r_U. By definition, the state of the molecule at λ_{k^*} is always U, so $\sigma_{k^*} = U$. Moreover, all the forward trajectories end in U at λ_M, so $\sigma_M = U$ and $\phi^{\rightarrow}_{\sigma_M} = 1$ in Eq.(B.3). Analogously, $\Delta G_{\sigma_{k^*},\sigma_M} = \Delta G_{U,U} = G_U(\lambda_M) - G_U(\lambda_{k^*})$ is the partial free energy difference between state U at λ_M) and λ_{k^*}.

In Eqs. (B.2) and (B.3), k_B is the Boltzmann constant and T is the temperature. In the following lines, to lighten notation, we keep σ_k, σ_{k^*}, σ_M as free variables. Only at the end, we replace them with $\sigma_k = F$ for

$0 \leq k \leq k^* - 1$ and σ_{k^*}, $\sigma_M = U$. Inserting Eqs. (B.2) and (B.3) into Eq.(B.1) leads

$$\rho_{\rightarrow}(W|k^*) = \int \left[\prod_{k=0}^{k^*-1} dW_{k+1}\right] \cdot$$
$$\cdot dW' \delta(W - \sum_{k=0}^{k^*-1} W_{k+1} - W') \cdot A \cdot B \tag{B.4}$$

where $\beta = 1/k_B T$ and,

$$B = \prod_{k=0}^{k^*-1} \rho_{\lambda_k \leftarrow \lambda_{k+1}}(-W_{k+1}) \rho_{\lambda_{k^*} \leftarrow \lambda_M}(-W') \tag{B.5}$$

$$\begin{aligned} A &= \prod_{k=0}^{k^*-1} e^{\beta\left[W_{k+1} - \Delta G_{\sigma_k,\sigma_{k+1}} + k_B T \log\left(\frac{\phi^{\leftarrow}_{\sigma_k}}{\phi^{\rightarrow}_{\sigma_{k+1}}}\right)\right]} \cdot \\ &\quad \cdot e^{\beta\left[W' - \Delta G_{\sigma_{k^*},\sigma_M} + k_B T \log\left(\frac{\phi^{\leftarrow}_{\sigma_{k^*}}}{\phi^{\rightarrow}_{\sigma_M}}\right)\right]} = \\ &= A' \cdot A'' \cdot A''' \end{aligned} \tag{B.6}$$

with

$$A' = \prod_{k=0}^{k^*-1} e^{\beta W_{k+1}} \cdot e^{\beta W'} = e^{\beta\left(\sum_{k=0}^{k^*-1} W_{k+1} + W'\right)} = e^{\beta W} \tag{B.7}$$

$$\begin{aligned} A'' &= \prod_{k=0}^{k^*-1} e^{-\beta \Delta G_{\sigma_k,\sigma_{k+1}}} \cdot e^{-\beta \Delta G_{\sigma_{k^*},\sigma_M}} = \\ &= e^{-\beta \sum_{k=0}^{k^*-1} \Delta G_{\sigma_k,\sigma_{k+1}} - \beta \Delta G_{\sigma_{k^*},\sigma_M}} = e^{-\beta \Delta G_{F,U}} \end{aligned} \tag{B.8}$$

$$
\begin{aligned}
A''' &= \prod_{k=0}^{k^*-1} e^{\beta k_B T \log\left(\frac{\phi^{\leftarrow}_{\sigma_k}}{\phi^{\rightarrow}_{\sigma_{k+1}}}\right)} \cdot e^{\beta k_B T \log\left(\frac{\phi^{\leftarrow}_{\sigma_{k^*}}}{\phi^{\rightarrow}_{\sigma_M}}\right)} = \\
&= \prod_{k=0}^{k^*-1} \frac{\phi^{\leftarrow}_{\sigma_k}}{\phi^{\rightarrow}_{\sigma_{k+1}}} \cdot \frac{\phi^{\leftarrow}_{\sigma_{k^*}}}{\phi^{\rightarrow}_{\sigma_M}} = \frac{\prod_{k=0}^{k^*-1} \phi^{\leftarrow}_{\sigma_k} \cdot \phi^{\leftarrow}_{\sigma_{k^*}}}{\prod_{k=0}^{k^*-1} \phi^{\rightarrow}_{\sigma_{k+1}} \cdot \phi^{\rightarrow}_{\sigma_M}} = \\
&= \frac{\prod_{k=0}^{k^*-2} \phi^{F,F}_{\lambda_k \leftarrow \lambda_{k+1}}}{\prod_{k=0}^{k^*-2} \phi^{F,F}_{\lambda_k \rightarrow \lambda_{k+1}}} \cdot \frac{\phi^{U,U,r_F}_{\lambda_{k^*} \leftarrow \lambda_M}}{\phi^{U,U,r_F}_{\lambda_{k^*} \leftarrow \lambda_M}} \cdot \frac{\phi^{F,U}_{\lambda_{k^*-1} \leftarrow \lambda_{k^*}} \phi^{U,U}_{\lambda_{k^*} \leftarrow \lambda_M}}{\phi^{F,U}_{\lambda_{k^*-1} \rightarrow \lambda_{k^*}} \phi^{U,U}_{\lambda_{k^*} \rightarrow \lambda_M}} = \\
&= \frac{\widetilde{\psi}_{k^*}}{\psi_{k^*}} \cdot \frac{p^U_{\leftarrow,k^*}(r_U)}{p^U_{\leftarrow,k^*}(r_F)} \qquad \text{(B.9)}
\end{aligned}
$$

where in Eqs. (B.8) and (B.9), we used $\sigma_0 = F$, $\sigma_M = U$. In addition, in the last line of Eq.(B.9), we have introduced a multiplicative factor equal to 1 ($\phi^{U,U,r_F}_{\lambda_{k^* \leftarrow \lambda_M}} / \phi^{U,U,r_F}_{\lambda_{k^* \leftarrow \lambda_M}}$). Moreover, in the last line of Eq.(B.9) we adopted a specific notation for the conditional probabilities or fractions $\phi^{\leftarrow}_{\sigma_k}$, $\phi^{\rightarrow}_{\sigma_{k+1}}$, $\phi^{\leftarrow}_{\sigma_{k^*}}$ and $\phi^{\rightarrow}_{\sigma_M}$ previously introduced in Eqs. (B.2) and (B.3) (see Tab. B.1). The conditional probabilities are:

1. $\phi^{F,F}_{\lambda_k \leftarrow \lambda_{k+1}}$ is the fraction of reverse trajectories where $\sigma_k = F$ at λ_k conditioned to $\sigma_{k+1} = F$ at λ_{k+1}. This fraction is measured with the unloading rate r_F.

2. $\phi^{F,F}_{\lambda_k \rightarrow \lambda_{k+1}}$ is the fraction of forward trajectories where $\sigma_{k+1} = F$ at λ_{k+1} conditioned to $\sigma_k = F$ at λ_k. This fraction is measured with the loading rate r_F.

3. $\phi^{U,U,r_F}_{\lambda_{k^*} \leftarrow \lambda_M}$ is the fraction of reverse trajectories where $\sigma_{k^*} = U$ at λ_{k^*} starting at $\sigma_M \equiv U$ at λ_M. As explicitly indicated in the notation, this fraction is measured at the unloading rate r_F.

4. $\phi^{F,U}_{\lambda_{k^*-1} \leftarrow \lambda_{k^*}}$ is the fraction of reverse trajectories where $\sigma_{k^*-1} = F$ at λ_k conditioned to $\sigma_{k^*} = U$ at λ_{k^*}. This fraction is measured with the unloading rate r_F

5. $\phi^{F,U}_{\lambda_{k^*-1}\to\lambda_{k^*}}$ is the fraction of forward trajectories where $\sigma_{k^*} = U$ at λ_{k^*} conditioned to $\sigma_{k^*-1} = F$ at λ_{k^*-1}. This fraction is measured with the unloading rate r_F;

6. $\phi^{U,U}_{\lambda_{k^*}\leftarrow\lambda_M}$ is the fraction of reverse trajectories where $\sigma_{k^*} = U$ at λ_{k^*} starting at $\sigma_M \equiv U$ at λ_M. This fraction is measured with the unloading rate r_U.

7. $\phi^{U,U}_{\lambda_{k^*}\to\lambda_M}$ is the fraction of forward trajectories where $\sigma_M = U$ at λ_M conditioned to $\sigma_{k^*} = U$ at λ_{k^*}. This fraction is measured with the loading rate r_U. Note that, because all trajectories end in U at λ_M the fraction $\phi^{U,U}_{\lambda_{k^*}\to\lambda_M}$ equals 1.

Note that in items 4 and 5 we have introduced the quantities $\phi^{F,U}_{\lambda_{k^*-1}\leftarrow\lambda_{k^*}}$ and $\phi^{F,U}_{\lambda_{k^*-1}\to\lambda_{k^*}}$, both measured with the loading rate r_F.

notation in Eq. (B.9)	notation in Eqs. (B.2) and (B.3)	measured at
$\phi^{F,F}_{\lambda_k\leftarrow\lambda_{k+1}}$	$\phi^{\leftarrow}_{\sigma_k}$ with $\sigma_k = \sigma_{k+1} = F$	r_F
$\phi^{F,F}_{\lambda_k\to\lambda_{k+1}}$	$\phi^{\rightarrow}_{\sigma_{k+1}}$ with $\sigma_k = \sigma_{k+1} = F$	r_F
$\phi^{U,U,r_F}_{\lambda_{k*}\leftarrow\lambda_M}$	$\phi^{\leftarrow}_{\sigma_{k^*}}$ with $\sigma_k* = \sigma_M = U$	r_F
$\phi^{F,U}_{\lambda_{k^*-1}\leftarrow\lambda_{k*}}$	$\phi^{\leftarrow}_{\sigma_{k^*-1}}$ with $\sigma_{k^*-1} = F; \sigma_{k^*} = U$	r_F
$\phi^{F,U}_{\lambda_{k^*-1}\to\lambda_{k*}}$	$\phi^{\rightarrow}_{\sigma_{k^*}}$ with $\sigma_{k^*-1} = F; \sigma_{k^*} = U$	r_F
$\phi^{U,U}_{\lambda_{k^*}\leftarrow\lambda_M}$	$\phi^{\leftarrow}_{\sigma_{k*}}$ with $\sigma_k* = \sigma_M = U$	r_U
$\phi^{U,U}_{\lambda_{k*}\to\lambda_M}$(=1)	$\phi^{\rightarrow}_{\sigma_M}$ with $\sigma^*_k = \sigma_M = U$	r_U

Table B.1.: **Notations.**

To demonstrate the last equality in the last line of Eq.(B.9) we group into a single product all fractions regarding reverse transitions at the un-

loading rate r_F in the numerator ($\phi^{F,F}_{\lambda_k\leftarrow\lambda_{k+1}}$, $\phi^{F,U}_{\lambda_{k^*-1}\leftarrow\lambda_{k^*}}$, $\phi^{U,U,r_F}_{\lambda_{k^*}\leftarrow\lambda_M}$), and all fractions of forward transitions at the unloading rate r_F in the denominator ($\phi^{F,F}_{\lambda_k\rightarrow\lambda_{k+1}}$, $\phi^{F,U}_{\lambda_{k^*-1}\rightarrow\lambda_{k^*}}$). We define

$$\prod_{k=0}^{k^*-2}\phi^{F,F}_{\lambda_k\rightarrow\lambda_{k+1}}\cdot\phi^{F,U}_{\lambda_{k^*-1}\rightarrow\lambda_{k^*}}\phi^{U,U}_{\lambda_{k^*}\rightarrow\lambda_M}=\psi_{k^*} \tag{B.10}$$

$$\prod_{k=0}^{k^*-2}\phi^{F,F}_{\lambda_k\leftarrow\lambda_{k+1}}\cdot\phi^{F,U}_{\lambda_{k^*-1}\leftarrow\lambda_{k^*}}\phi^{U,U,r_F}_{\lambda_{k^*}\leftarrow\lambda_M}=\widetilde{\psi}_{k^*}. \tag{B.11}$$

$\psi_{k^*}(1\leq k^*\leq M)$ is the fraction of forward trajectories that start in F at λ_0 and are observed to be in U for the *first* time at λ_{k^*} with loading rate r_F. $\widetilde{\psi}_{k^*}(1\leq k^*\leq M)$ is the fraction of reverse trajectories that start in U at λ_M and are observed to be in U for the *last* time at λ_{k^*} with unloading rate r_F. Notice that for $k^*=1$, the products $\prod_{k=0}^{k^*-2}$ in Eqs. (B.10) and (B.11) are equal to 1. We stress two facts: 1) both ψ_{k^*}, $\widetilde{\psi}_{k^*}$ are fractions measured at the single pulling rate r_F without feedback and; 2) the notion of *first* and *last* time is bound to trajectories Γ defined as sequences of observations at the predetermined measurement positions λ_k as they are defined in the 1$^{\text{st}}$TF protocol, irrespective of what is the state of the molecule at other intermediate (unobserved) positions.

Finally, in Eq.(B.9), $\phi^{U,U}_{\lambda_{k^*}\leftarrow\lambda_M}=p^U_{\leftarrow,k^*}(r_U)$ is the fraction of reverse trajectories that start in U at λ_M and are observed to be in U at λ_{k^*} at the unloading rate r_U. According to this definition, we also have $\phi^{U,U,r_F}_{\lambda_{k^*}\leftarrow\lambda_M}=p^U_{\leftarrow,k^*}(r_F)$, a term that also appears in the denominator of the last fraction in Eq.(B.9).

Inserting Eqs. (B.7, B.8, B.9) in Eq.(B.6) and then in Eq.(B.4), we remark that A can be taken out of the integral in Eq.(B.4). The remaining integral in Eq.(B.4) contains only the term B from Eq.(B.5), which yields the reverse work distribution $\rho_{\leftarrow}(-W|k^*)$. We stress that the reverse work distribution is conditioned to forward process, through the first

unfolding event observed at λ_{k^*} along that process. Putting everything together we get the detailed work-FT for the 1$^{\text{st}}$TF protocol,

$$\frac{\rho_{\rightarrow}(W|k^*)}{\rho_{\leftarrow}(-W|k^*)} = \exp\left[\beta\left(W - \Delta G_{F,U} + k_B T J_{k^*}\right)\right]$$
$$J_{k^*} = \log\left(\frac{\widetilde{\psi}_{k^*}}{\psi_{k^*}} \cdot \frac{p^U_{\leftarrow,k^*}(r_U)}{p^U_{\leftarrow,k^*}(r_F)}\right) (1 \leq k^* \leq M)\,. \tag{B.12}$$

In equation (B.12), J_{k^*} denotes for the partial thermodynamic information and depends on four basic quantities (ψ_{k^*}, $\widetilde{\psi}_{k^*}$, $p^U_{\leftarrow,k^*}(r_F)$, $p^U_{\leftarrow,k^*}(r_U)$). These quantities are experimentally measured in protocols without feedback at the pulling rate r_F along the forward process (ψ_{k^*}) and the reverse process ($\widetilde{\psi}_{k^*}$, $p^U_{\leftarrow,k^*}(r_F)$), and at the pulling rate r_U along the reverse process ($p^U_{\leftarrow,k^*}(r_U)$).

From Eq.(B.12), we derive the full work-FT for the 1$^{\text{st}}$TF protocol. The full work-distribution in the forward process is given by:

$$\begin{aligned}
\rho_{\rightarrow}(W) &= \sum_{k=1}^{M} \rho_{\rightarrow}(W|k)\cdot\psi_k = \\
&= \sum_{k=1}^{M} \rho_{\leftarrow}(-W|k) e^{\beta(W-\Delta G_{FU}+k_B T J_k)}\psi_k = \\
&= e^{\beta(W-\Delta G_{FU})}\cdot\sum_{k=1}^{M}\rho_{\leftarrow}(-W|k)e^{J_k}\cdot\psi_k = \\
&= e^{\beta(W-\Delta G_{FU})}\cdot\frac{\sum_{k=1}^{M}\rho_{\leftarrow}(-W|k)e^{J_k}\psi_k}{\sum_{k=1}^{M}e^{J_k}\cdot\psi_k}\cdot \\
&\quad\cdot\sum_{k=1}^{M}e^{J_k}\cdot\psi_k
\end{aligned} \tag{B.13}$$

where, in the last line we have multiplied and divided by the term $\sum_{k=1}^{M} e^{J_k}\cdot\psi_k$. This allows us to define the reverse full-work distribution for the 1$^{\text{st}}$TF protocol,

$$\rho_{\leftarrow}(-W) = \frac{\sum_{k=1}^{M}\rho_{\leftarrow}(-W|k)e^{J_k+\log\psi_k}}{\sum_{k=1}^{M}e^{J_k+\log\psi_k}}\,. \tag{B.14}$$

Notice that $\rho_{\leftarrow}(W)$ is properly normalized. Finally we get:

$$\frac{\rho_{\rightarrow}(W)}{\rho_{\leftarrow}(-W)} = e^{\beta(W-\Delta G_{FU}+k_B T \Upsilon_M)} \tag{B.15}$$

The term Υ_M in Eq. (B.15) is the thermodynamic information and equals

$$\Upsilon_M = \log\left[\sum_{k=1}^{M} \psi_k e^{J_k}\right] = \log\left[\sum_{k=1}^{M} \psi_k \frac{p^U_{\leftarrow,k}(r_U)}{p^U_{\leftarrow,k}(r_F)} \cdot \frac{\widetilde{\psi}_k}{\psi_k}\right]$$

$$\Upsilon_M = \log\left(\sum_{k=1}^{M} \frac{p^U_{\leftarrow,k}(r_U)}{p^U_{\leftarrow,k}(r_F)} \widetilde{\psi}_k\right) \tag{B.16}$$

Continus-time limit

To conclude this appendix, we determine the partial and full thermodynamic information, $J(\lambda)$ and Υ_∞, for the Continuous-Time Feedback (CTF) case corresponding to the limit $M \to \infty$.

In this limit Eqs.(B.2-B.16) hold but with the continuous variable λ replacing the discrete variable k. The partial thermodynamic information J_k becomes the continuous function $J(\lambda)$ defined as:

$$J(\lambda) = \log\left(\frac{p^U_{\leftarrow}(\lambda, r_U)}{p^U_{\leftarrow}(\lambda, r_F)} \frac{\widetilde{\psi}(\lambda)}{\psi(\lambda)}\right) . \tag{B.17}$$

with equivalent definitions for the continuous first unfolding $\psi(\lambda)$ and last folding $\widetilde{\psi}(\lambda)$ distributions, and $p^U_{\leftarrow}(\lambda, r_U)$, and $p^U_{\leftarrow}(\lambda, r_F)$. The full thermodynamic information Υ_∞ is determined by taking the continuous limit, $\lambda_{k+1} = \lambda_k + \Delta\lambda$, where $\Delta\lambda \to 0$, and writing the sum in Eq.(B.16) as an integral:

$$\Upsilon_\infty = \log\left(\int_{\lambda_{min}}^{\lambda_{max}} \frac{p^U_{\leftarrow}(\lambda, r_U)}{p^U_{\leftarrow}(\lambda, r_F)} \widetilde{\psi}(\lambda) d\lambda\right). \tag{B.18}$$

C

FLUCTUATION THEOREM FOR DISCRETE-TIME FEEDBACK

Here, we derive the detailed and full FT for the discrete-time feedback (DTF) using two different methods, either as the 1stTF-FT limit for $M = 2$ or by directly applying the extended-FT [50] by classifying trajectories according to the measurement outcome at the intermediate position λ_1.

First, we derived the full and partial fluctuation theorem for DTF by considering the 1stTF-FT with $M = 2$ (see App. B). In this case state measurement sequences are of the type $\Gamma = \{\sigma_0 = F,\ \sigma_1,\ \sigma_2 = U\}$ corresponding to the three different measurements trap positions (λ_k): λ_0, λ_1 and λ_2. The relevant quantities in Eqs. (B.12), (B.15), and (B.16) are ψ_k, $\widetilde{\psi}_k$, $p^{U}_{\leftarrow,k}(r_F)$, $p^{U}_{\leftarrow,k}(r_U)$, which for the relevant trap positions are:

1. For $k = 0$ ($\lambda_0 \equiv \lambda_{min}$): By definition all the forward (reverse) trajectories start (end) are in F at λ_0, i.e. $p^{U}_{\leftarrow,0}(r_F) = 0$ and $p^{U}_{\leftarrow,0}(r_U) = 0$. For the same reason, the probability to see the last refolding event at λ_0 is zero ($\widetilde{\psi}_0 = 0$).

2. For $k = 1$ (λ_1): $p^{U}_{\leftarrow,1}(r_F) \neq 0$ and $p^{U}_{\leftarrow,1}(r_U) \neq 0$. Moreover, the fact that the molecule always starts in F (U) and ends in U (F) during the forward (reverse) process implies that the probability to observe the first (last) unfolding (refolding) event at λ_1 equals the probability to be in U at λ_1 during the forward (reverse)

process. This holds for all pulling rate values: $\psi_1 = p^U_{\rightarrow,1}(r_F)$, $\widetilde{\psi}_1 = p^U_{\leftarrow,1}(r_F)$.

3. For $k = 2$ ($\lambda_2 \equiv \lambda_{max}$): By definition the probability to be in U at λ_2 equals 1 because all forward (reverse) trajectories end (start) in U, i.e., $p^U_{\leftarrow,2}(r_F) = p^U_{\leftarrow,2}(r_U) = 1$. Moreover, the fact that the molecule always starts in F (U) and ends in U (F) during the forward (reverse) process implies that the probability to observe the first (last) unfolding (refolding) event at λ_2 equals the probability that the molecule is in F at λ_1 during the forward (reverse) process. This holds for all pulling rate values: $\psi_2 = p^F_{\rightarrow,1}(r_F)$, $\widetilde{\psi}_2 = p^F_{\leftarrow,1}(r_F)$.

The different values of $p^U_{\leftarrow,k}$, ψ_k, $\widetilde{\psi}_k$ ($k = 1, 2$) are presented in Tab. C.1.

λ_0	λ_1	λ_2
$p^U_{\leftarrow,0}(r_F) = 0$	$p^U_{\leftarrow,1}(r_F) \neq 0$	$p^U_{\leftarrow,2}(r_F) = 1$
$p^U_{\leftarrow,0}(r_U) = 0$	$p^U_{\leftarrow,1}(r_U) \neq 0$	$p^U_{\leftarrow,2}(r_U) = 1$
$\psi_0 = 0$	$\psi_1 = p^U_{\rightarrow,1}(r_F)$	$\psi_2 = p^F_{\rightarrow,1}(r_F)$
$\widetilde{\psi}_0 = 0$	$\widetilde{\psi}_1 = p^U_{\leftarrow,1}(r_F)$	$\widetilde{\psi}_2 = p^F_{\leftarrow,1}(r_F)$

Table C.1.: Fractions for DTF.

From the results presented in Tab. C.1, we calculate the partial and full thermodynamic information, $J_1 \equiv I_U$, $J_2 \equiv I_F$ (Eq.(B.12)) and Υ_2 (Eq.(B.16)):

$$
\begin{aligned}
I_U = J_1 &= \log\left(\frac{p^U_{\leftarrow,1}(r_U)}{p^U_{\rightarrow,1}(r_F)}\right) \\
I_F = J_2 &= \log\left(\frac{p^F_{\leftarrow,1}(r_F)}{p^F_{\rightarrow,1}(r_F)}\right) \\
\Upsilon_2 &= \log\left(p^U_{\leftarrow,1}(r_U) + p^F_{\leftarrow,1}(r_F)\right)
\end{aligned}
\tag{C.1}
$$

Second, we derived the detailed and full work-FT in DTF by classifying the trajectories in two classes depending on the observation made at λ_1 : 1) the system is in F at λ_1 or, 2) the system is in U at λ_1. We use the extended-FT [50] to calculate the detailed work-FT for each class of trajectories, first between $\lambda_0 = \lambda_{min}$ and λ_1, next between λ_1 and $\lambda_2 = \lambda_{max}$. These results are then combined to extract the detailed work-FT, for each class of trajectories, for the full pulling cycle between $\lambda_0 = \lambda_{min}$ and $\lambda_2 = \lambda_{max}$.

The detailed work-FT, in the range $\lambda_0 \to \lambda_1$, is given by [50]:

$$\frac{p^F_{\rightarrow}(\lambda_1)}{p^F_{\leftarrow}(\lambda_1)} \frac{\rho_{\lambda_0,\lambda_1}(W|F)}{\rho_{\lambda_0,\lambda_1}(-W|F)} = \exp\left(\beta(W - \Delta G^{\lambda_0,\lambda_1}_{F,F})\right) \tag{C.2a}$$

$$\frac{p^U_{\rightarrow}(\lambda_1)}{p^U_{\leftarrow}(\lambda_1)} \frac{\rho_{\lambda_0,\lambda_1}(W|U)}{\rho_{\lambda_0,\lambda_1}(-W|U)} = \exp\left(\beta(W - \Delta G^{\lambda_0,\lambda_1}_{F,U})\right) \tag{C.2b}$$

where $\Delta G^{\lambda_0,\lambda_1}_{F,F(U)}$ is the free energy difference between state F (U) at λ_1 and state F at λ_0. $\rho_{\lambda_0,\lambda_1}(W|F(U))$ is the work distribution for the class of forward trajectories that start in F at λ_0 and end in F (U) at λ_1. $\rho_{\lambda_0,\lambda_1}(-W|F(U))$ is the corresponding reverse work distribution. $p^{F(U)}_{\rightarrow}(\lambda_1)$ are the probabilities in the forward process to be in F (U) at λ_1 conditioned to start in F at λ_0. $p^{F(U)}_{\leftarrow}(\lambda_1)$ are the probabilities in the reverse process to be in F at λ_0 conditioned to start in F (U) at λ_1. As mentioned before, by definition $p^{F(U)}_{\leftarrow}(\lambda_1) = 1$ because the system always ends in F at λ_0. All quantities in Eqs. (C.2a) and (C.2b) are measured at the pulling rate r_F.

Analogously, the detailed work-FT in the range $\lambda_1 \to \lambda_2$, is given by:

$$\frac{p^F_{\rightarrow}(\lambda_1)}{p^F_{\leftarrow}(\lambda_1|r_F)} \frac{\rho_{\lambda_1,\lambda_2}(W|F)}{\rho_{\lambda_1,\lambda_2}(-W|F,r_F)} = \exp\left(\beta(W - \Delta G^{\lambda_1,\lambda_2}_{F,U})\right) \tag{C.3a}$$

$$\frac{p^U_{\rightarrow}(\lambda_1)}{p^U_{\leftarrow}(\lambda_1,r_U)} \frac{\rho_{\lambda_1,\lambda_2}(W|U)}{\rho_{\lambda_1,\lambda_2}(-W|U,r_U)} = \exp\left(\beta(W - \Delta G^{\lambda_1,\lambda_2}_{U,U})\right) \tag{C.3b}$$

where $\Delta G^{\lambda_1,\lambda_2}_{F(U),U}$ is the free energy difference between state U at λ_2 and state F (U) at λ_1, $\rho_{\lambda_1,\lambda_2}(W|F(U))$ is the work distribution for the class of forward trajectories that start in F (U) at λ_1 and end in U at λ_2. $\rho_{\lambda_1,\lambda_2}(-W|F(U), r_{F(U)})$ is the corresponding reverse work distribution at the corresponding unloading rate, r_F (Eq. (C.3a)) and r_U (Eq. (C.3b)). $p^{F(U)}_{\rightarrow}(\lambda_1)$ are the probabilities in the forward process to be in U at λ_2 conditioned to start at F (U) in λ_1. Remember, by definition $p^{F(U)}_{\rightarrow}(\lambda_1) = 1$ because the molecule always ends in U at λ_2. $p^{F(U)}_{\leftarrow}(\lambda_1|v_{F(U)})$ are the probabilities in the reverse process to be in F (U) at λ_1 conditioned to start in U at λ_2 with unloading rate r_F (r_U). The unloading rate value for quantities in the reverse process are explicitly indicated in Eqs. (C.2a) and (C.2b)).

From Eqs.(C.2a, C.2b, C.3a, C.3b), we calculated the partial forward work distributions across the whole range $\lambda_{min} \rightarrow \lambda_{max}$ for the two classes of trajectories:

$$\rho_{\rightarrow}(W|F) = \int \rho_{\lambda_0,\lambda_1}(W_1|F)\rho_{\lambda_1,\lambda_2}(W_2|F) \times\delta(W - W_1 - W_2)dW_1dW_2 \tag{C.4a}$$

$$\rho_{\rightarrow}(W|U) = \int \rho_{\lambda_0,\lambda_1}(W_1|U)\rho_{\lambda_1,\lambda_2}(W_2|U) \times\delta(W - W_1 - W_2)dW_1dW_2\ . \tag{C.4b}$$

Putting everything together, we obtain the *detailed* work-FTs for DTF.

$$\frac{\rho_{\rightarrow}(W|F)}{\rho_{\leftarrow}(-W|F, r_F)} = \exp\left(\beta(W - \Delta G_{FU} + k_BTI_F)\right) \quad \text{with} \quad I_F = \log\left(\frac{p^F_{\leftarrow}(r_F)}{p^F_{\rightarrow}}\right) \tag{C.5a}$$

$$\frac{\rho_{\rightarrow}(W|U)}{\rho_{\leftarrow}(-W|U, r_U)} = \exp\left(\beta(W - \Delta G_{FU} + k_BTI_U)\right) \quad \text{with} \quad I_U = \log\left(\frac{p^U_{\leftarrow}(r_U)}{p^U_{\rightarrow}}\right) \tag{C.5b}$$

where $\Delta G_{FU} = \Delta G^{\lambda_0,\lambda_1}_{F,U(F)} + \Delta G^{\lambda_1,\lambda_2}_{U(F),U} = G_U(\lambda_{max}) - G_F(\lambda_{min})$ is the full free-energy difference. To simplify Eqs. (C.5a) and (C.5b), we dropped the argument λ_1 from the fractions $p^{F(U)}_{\rightarrow(\leftarrow)}(\lambda_1)$. Equations (C.5a) and (C.5b) are the detailed feedback-FTs for the two set of trajectories measured using the DTF.

From Eqs. (C.5a) and (C.5b), we computed the full work-FT as follows:

$$\begin{aligned}\rho_{\rightarrow}(W) &= \rho_{\rightarrow}(W|F)p^{F}_{\rightarrow} + \rho_{\rightarrow}(W|U)p^{U}_{\rightarrow} \\ &= \rho_{\leftarrow}(-W)e^{\left(\beta\left(W-\Delta G_{FU}+k_BT\log\left[p^{F}_{\leftarrow}(r_F)+p^{U}_{\leftarrow}(r_U)\right]\right)\right)}\end{aligned}$$

where:

$$\rho_{\leftarrow}(-W) = \frac{\left(p^{F}_{\leftarrow}(r_F)\rho_{\leftarrow}(-W|F,r_F) + p^{U}_{\leftarrow}(r_U)\rho_{\leftarrow}(-W|U,r_U)\right)}{\left(p^{F}_{\leftarrow}(r_F) + p^{U}_{\leftarrow}(r_U)\right)}. \tag{C.6}$$

Therefore, we obtain the *full* work-FT for DTF:

$$\begin{aligned}&\frac{\rho_{\rightarrow}(W)}{\rho_{\leftarrow}(-W)} = \exp\left(\beta(W - \Delta G_{FU} + k_BT\Upsilon_2)\right) \\ &\text{with } \Upsilon_2 = \log\left(p^{F}_{\leftarrow}(r_F) + p^{U}_{\leftarrow}(r_U)\right)\end{aligned} \tag{C.7}$$

D

PHENOMENOLOGICAL SIMULATION DETAILS

Here, we present how we simulated the pulling experiments where the trap position, λ, is the control parameter. The forward and reverse trajectories are generated as a first-order Markov chain where, for a given value of λ and force, f, it is satisfied that:

$$\lambda(f) = x^{\sigma}_{DNA}(f) + x_h(f) + x_b(f) \; . \tag{D.1}$$

In Eq. (D.1), $x_b(f)$ is the position of the trapped bead relative to the center of the optical trap; $x_h(f)$ is the end-to-end distance of the dsDNA handles; and x^{σ}_{DNA} is the end-to-end distance of the DNA hairpin, which depends on the state of the molecule ($\sigma = F, U$). The latter defines two force branches, one when the DNA hairpin is folded and the other when the hairpin is unfolded.

To determine the two force branches, first we need to determine the extension of each element at a given λ, i.e., at a given force. On one hand, $x_{trap}(f)$ satisfies:

$$|f_{\text{Hooke}}| = k_{trap} x_b \tag{D.2}$$

being k_{trap} the trap stiffness ($k_{trap} = 0.06 - 0.068$ pN/nm). In this simple approximation (Hooke's law), the bead behaves as a Brownian particle under the action of two opposing springs (the optical trap and the

molecular construct), hence the white noise due to thermal fluctuations satisfies:

$$\langle \delta x^2 \rangle = \frac{k_B T}{k^{\sigma}_{mol} + k_b} \,, \quad \langle \delta f^2 \rangle = \frac{k_B T k_b^2}{k^{\sigma}_{mol} + k_b} \tag{D.3}$$

where k^{σ}_{mol} is the stiffness of the molecular construct. It is given by $1/k^{\sigma}_{mol} = 1/k_h + 1/k^{\sigma}_{DNA}$) where k_h and k_{DNA} are the stiffness of the ds-DNA handle and the DNA hairpin. These stiffness values are calculated as $\partial x(f)/\partial f$, being x the molecular extension of the dsDNA handles and the DNA hairpin. Furthermore, the elastic response of the dsDNA handles and the unfolded DNA hairpin is well described by the Worm-Like Chain model (experimental details can be found in Chap. 6),

$$f_{\mathrm{WLC}} = \frac{k_B T}{4 L_p} \left(\left(1 - \frac{x}{L_c} \right)^{-2} + 4 \frac{x}{L_c} - 1 \right) \tag{D.4}$$

where k_B is the Boltzmann's constant and T is the temperature. f_{WLC} is the measured force at a given end-to-end extension x of the dsDNA handles and the unfolded hairpin; L_p is the persistence length (L_p = 10nm for the dsDNA handles and L_p = 1.34nm for the unfolded DNA hairpin at 25ºC 6) and L_c is the contour length ($L_c = 58\mathrm{bp} \cdot 0.34\mathrm{nm/bp}$ = 19.72nm for the dsDNA handles and $L_c = Ndb$ being N the number of bases and d_b the inter-phosphate distance (0.59 nm at 25ºC) for the unfolded DNA hairpin). Besides, the folded DNA hairpin behaves as a dipole of length (d_0 = 2 nm) of extension:

$$x_d(f) = d_0 \left[\coth\left(\frac{d_0 f}{k_B T} \right) - \frac{k_B T}{d_0 f} \right] \tag{D.5}$$

To simulate pulling experiment, we calculate the unfolding, $k_{\rightarrow}(\lambda)$, and refolding, $k_{\leftarrow}(\lambda)$, kinetic rates as described in [112]. The kinetic rates are defined as:

$$k_{\rightarrow}(\lambda) = k_0 \exp\left(-\frac{B(\lambda)}{k_B T} \right) \tag{D.6a}$$

$$k_{\leftarrow}(\lambda) = k_0 \exp\left(-\frac{B(\lambda) - \Delta G_{FU}(\lambda)}{k_B T} \right) \tag{D.6b}$$

where $B(\lambda)$ is the kinetic barrier, k_0 is the attempt rate ($k_0 = 8 \cdot 10^3 s^{-1}$) and $\Delta G_{FU}(\lambda)$ is the energy difference between states F and U at a given λ. The latter is determined as:

$$\Delta G_{FU}(\lambda) = \Delta G_{FU} + \Delta W_{FU}^{DNA} + \Delta W_{FU}^{h} + \Delta W_{FU}^{b} \tag{D.7}$$

where ΔG_{FU} is the folding free energy of the DNA hairpin ($\Delta G_{FU} =$ 51 k_BT for the hairpin used in Chap. 3); ΔW_{FU}^{DNA} is the stretching contribution of the unfolded hairpin ($W_U^{DNA} = \int_0^{x_{DNA}^U} f_{\mathrm{WLC}}(x')dx'$) minus the orientation contribution of the folded DNA hairpin ($W_F^{DNA} = \int_0^{x_{DNA}^F} f_{\mathrm{d}}(x')dx'$) with $f_d(x)$ the inverse function of Eq. (D.5); ΔW_{FU}^{h} is the stretching contribution of the handles ($\Delta W_{FU}^{h} = \int_{x_h^F}^{x_h^U} f_{\mathrm{WLC}}(x')dx'$) where the limits are the end-to-end distances at a given λ evaluated at the different forces, when the hairpin is folded and unfolded; ΔW_{FU}^{b} is the work done to displace the bead with respect to the center of the optical trap between the folded and unfolded branches ($\Delta W_{FU}^{b} = \int_{x_b^F}^{x_b^U} f_{\mathrm{Hooke}}(x')dx'$).

Finally the kinetic barrier $B(\lambda)$ is derived using the Kramers solution to the one-dimensional diffusion problem in the unzipping molecular free energy landscape described by the nearest-neighbour model [112]:

$$B(\lambda) = k_B T \log\left(\sum_{n=0}^{N}\sum_{n'=0}^{n} \exp\left[\frac{\Delta G_n(\lambda) - \Delta G_{n'}(\lambda)}{k_B T}\right]\right) . \tag{D.8}$$

In Eq.(D.8), $\Delta G_n(\lambda)$ is the free energy of the hairpin determined by the number of opened base-pairs, n, at a given λ. This energy is calculated as

$$\Delta G_n(\lambda) = \Delta G_n + \int_0^{x_n^{DNA}} f_{\mathrm{WLC}}(x')dx' + \int_0^{x_h^n} f_{\mathrm{WLC}}(x')dx' + \int_0^{x_b^n} f_{\mathrm{Hooke}}(x')dx' . \tag{D.9}$$

Here, x_n^{DNA} is the extension of DNA when only n base-pairs are opened, x_h^n and x_b^n are the extension of the handles and the bead position when only n base-pairs are opened.

The simulated trajectories are generated as follows: the forward trajectory is initialized at the folded state ($n = 0$) at $\lambda_{min} = 0$ nm, while the reverse one is initialized at the unfolded state ($n = N$) at λ_{max}, e.g., $\lambda_{max} = 300$ nm. During the simulation the relaxation of the handles and the bead is assumed to be instantaneous. The steps of the algorithm for the forward (reverse) process are:

1. λ increases (decreases) by the amount $r\Delta t/k_b$, were r is the loading rate and Δt is the inverse of the data acquisition frequency ($\Delta t = 0.001$s).

2. Eq.(D.1) is solved according to the state of the hairpin to find the value of the force f acting on the experimental setup. Moreover, to be more realistic we added a Gaussian noise of zero mean and variance given by Eq.(D.3) to the measured λ and f.

3. We calculate the probability to observe an unfolding or folding transition as: $k_{\rightarrow}(\lambda) \cdot \Delta t$ if the molecule is folded or as $k_{\leftarrow}(\lambda) \cdot \Delta t$ if the molecule is unfolded.

4. We compare the transition probability with an uniformly distributed random number between 0 and 1. If the probability is larger than the random number we change the state of the molecule in the next step, otherwise the molecule remains in the same state.

E

FOLDING FREE ENERGY, ENTROPY AND ENTHALPY

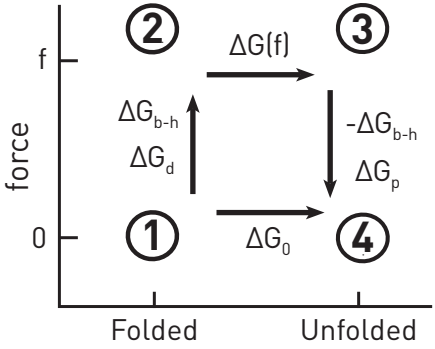

Figure E.1.: **Scheme of the different thermodynamic steps to measure ΔG_0.** Stretching of the folded molecule ($1 \rightarrow 2$), unfolding at a given force ($2 \rightarrow 3$), releasing of the unfolded molecule ($3 \rightarrow 4$), and unfolding at zero force ($1 \rightarrow 4$).

To derive the folding free energy at zero force, ΔG_0, from single-molecule force spectroscopy experiments we consider four different configurations or states of the molecule (Fig.E.1): (1) folded molecule at zero force; (2) folded molecule at a given force f; (3) unfolded molecule at the same force f; and (4) unfolded molecule at zero force. Notice that the direct unfolding pathway at zero force $1 \rightarrow 4$, observed in bulk experiments, can be decomposed as the sum of three sequential steps ($1 \rightarrow 4 = 1 \rightarrow 2 + 2 \rightarrow 3 + 3 \rightarrow 4$).

The three steps are as follows. *Step 1→2:* folded molecule is reversibly pulled from zero force to a final force f. In this case the free energy difference equals the sum of the reversible work of stretching the handles and displacing the bead in the optical trap, $\Delta G_{\mathrm{h-b}}(0 \to f)$, and the reversible work to orient the folded molecule, $\Delta G_{\mathrm{d}}(0 \to f)$. This contribution is analogous to orient a dipole of length 2nm diameter for the dsDNA hairpin and $\simeq$ 3nm for barnase. *Step 2→3:* folded molecule is reversibly unfolded (denaturated) at a constant force f. In this step the free energy difference, $\Delta G(f)$, equals the free energy of the stretched polypeptide chain or ssDNA minus the folding free energy of folded molecule, at force f. *Step 3→4:* the stretched polypeptide chain or ssDNA is reversibly relaxed from f to zero force. The free energy difference in this step equals the reversible work of releasing the polypeptide chain or ssDNA from force f to o (-$\Delta G_{\mathrm{p}}(0 \to f)$) plus the reversible work of relaxing the handles and the bead in the optical trap from f to zero (equal to -$\Delta G_{\mathrm{h-b}}(0 \to f)$, from step 1→2). Thermodynamic energy differences are path-independent so $\Delta G(1 \to 4) = \Delta G(1 \to 2) + \Delta G(2 \to 3) + \Delta G(3 \to 4)$. This gives,

$$\Delta G_0 = \Delta G_{\mathrm{d}}(0 \to f) + \Delta G(f) - \Delta G_{\mathrm{p}}(0 \to f) \tag{E.1}$$

The same balance equation holds for enthalpy (ΔH_0) and entropy (ΔS_0). Notice that $\Delta G_{\mathrm{h-b}}(0 \to f)$ does not appear in Eq.E.1 as it cancels out when adding steps 1→2 and 3→4.

It is worth mentioning that the unfolding transition at constant force can be measured in instruments where the force is the natural control parameter (e.g., in magnetic tweezers). In contrast, in optical tweezers force cannot be controlled unless force feedback is applied. As a consequence, in most of the single-molecule experiments using optical tweezers (as the hopping and pulling presented above), the unfolding transition does not occur at fixed force f but at fixed trap position (λ). Indeed, when pulling with optical tweezers the unfolding transition is observed as a sudden force rip which occurs at fixed λ. Therefore free energy differ-

ences in the force-ensemble, $\Delta G(f)$, are Legendre transforms of those measured in the λ-ensemble.

A major contribution in Eq.E.1 is the elastic term $\Delta G_{\rm p}$ for unfolded molecule, which is often modelled as a semiflexible polymer. The term $\Delta G_{\rm d}$ stands for the elastic energy of aligning a molecular-sized dipole along the force axis. As the dipole extension is much shorter than the contour length of the polypeptide chain, $\Delta G_{\rm p}(0 \to f) \gg \Delta G_{\rm d}(0 \to f)$ at all forces. The relative magnitude of $\Delta G_{\rm p}(0 \to f)$ and $\Delta G(f)$ depends on the difference between f and the coexistence force f_c, which is defined as the force at which the folded and unfolded molecule have equal free energies, i.e., $\Delta G(f_c) = 0$. Equation E.1 gives for $f = f_c$,

$$\Delta G_0 = -\Delta G_{\rm p}(0 \to f_c) + \Delta G_{\rm d}(0 \to f_c) \; . \tag{E.2}$$

The stretching free energy of the different elastic elements at a given force f can be obtained by using the well-known expression

$$\Delta G_{\rm i}(0 \to f) = -\int_0^f x_{\rm i}(f')df' \tag{E.3}$$

where i≡p,d, whereas $x_{\rm i}(f)$ stands for the molecular extension of the unfolded and folded states. The difference in molecular extension between both states, $\Delta\lambda$, is given by the difference in trap position in the FDCs obtained from pulling experiments at a fixed force,

$$\Delta\lambda(f) = x_{\rm p}(f) - x_{\rm d}(f) \; . \tag{E.4}$$

Combining the previous equations we obtain the relation,

$$\Delta G_0(T) = \int_0^{f_c(T)} \Delta\lambda(f,T)df \; , \tag{E.5}$$

showing that the knowledge of $f_c(T)$ and the measured $\Delta\lambda(f,T)$ permits to determine ΔG_0 at a given temperature T.

The folding entropy and enthalpy are directly derived from Eq. E.5. As it is known from thermodynamics, $\Delta S_0 = -\partial \Delta G_0 / \partial T$ and $\Delta H_0 = \Delta G_0 - T\Delta S_0$. For the entropy we find,

$$\Delta S_0(T) = -\frac{\partial f_c(T)}{\partial T}\Delta\lambda(f_c(T)) - \int_0^{f_c(T)} \frac{\partial \Delta\lambda(f', T)}{\partial T} df' \quad . \tag{E.6}$$

The first term of Eq. E.6 is analogous to the Clausius-Clapeyron equation for first-order phase transitions, where f and λ are the equivalent of pressure and volume. The second term in Eq.E.6 is the entropic contribution to stretch and orient the unfolded and folded (dipole) molecule from zero force to f_c.

Equations E.5 and Eq.E.6 are the basic thermodynamic formulas used throughout this thesis to determine $\Delta G_0(T)$, $\Delta S_0(T)$, $\Delta H_0(T)$ for the protein barnase and different DNA hairpins.

F

FOLDING FREE ENERGY DETERMINATION FROM SMES

In this appendix, we would provide the mathematical details of the derived energy shown in App. E. In SMEs where the force is controlled, e.g., magnetic tweezers or optical tweezers instruments with force feedback algorithms, the free energy difference between the unfolded and native states of the molecule under study are measured at a given force, $\Delta G(f)$, which is defined as

$$\Delta G(f) = G_U(f) - G_N(f) \,, \tag{F.1}$$

where $G_U(f)$ and $G_N(f)$ are the free energies of the unfolded (U) and native (N) states at force f, respectively. Taking the native molecule, e.g., protein or DNA/RNA hairpins, at zero force as reference state, $G_U(f)$ and $G_N(f)$ are calculated as in [25],

$$G_U(f) = -\int_0^f \lambda_U(f')df' + \Delta G_0 \tag{F.2a}$$

$$G_N(f) = -\int_0^f \lambda_N(f')df' \tag{F.2b}$$

with $\Delta G_0 (> 0)$ the free energy of the random coil state relative to the native state. Here, λ_U and λ_N are the extension of the system when the

molecule is in U and N at a given force f. Moreover, the extension of the system λ is decomposed as:

$$\lambda_U(f) = x_b(f) + x_h(f) + x_U(f) \tag{F.3a}$$

$$\lambda_N(f) = x_b(f) + x_h(f) + x_N(f) \tag{F.3b}$$

where $x_b(f)$ is the displacement of the bead with respect to the center of the optical trap; $x_h(f)$ is the extension of the dsDNA handles; $x_U(f)$ is the extension of the unfolded molecule; and, $x_N(f)$ is the extension of the native state along the force axis.

Introducing Eqs. (F.2a) and (F.2b) to Eq.F.1 and taking into consideration Eqs.(F.3a) and (F.3b), we obtain,

$$\begin{aligned}\Delta G(f) &= -\sum_{i=b,h,U}\int_0^f x_i(f')df' + \Delta G_0 + \sum_{i=b,h,N}\int_0^f x_i(f')df' = \\ &= \Delta G_0 - \int_0^f x_U(f')df' + \int_0^f x_N(f')df'\end{aligned} \tag{F.4}$$

where the bead and handles contributions cancel out. The two integrals of $x_U(f)$ and $x_N(f)$ in Eq.(F.4) equal the free energies required to stretch the unfolded molecule and orient the native state, respectively. In a more compact form,

$$\Delta G(f) = \Delta G_0 - \int_0^f (x_U(f') - x_N(f'))df' = \Delta G_0 - \int_0^f \Delta\lambda(f')df' \tag{F.5}$$

where $x_U(f) - x_N(f)$ stands for the difference in trap position at a given force $\Delta\lambda(f)$. In addition, Eq. (F.5) evaluated at the coexistence force gives,

$$\Delta G(f_c) = \Delta G_0 - \int_0^f \Delta\lambda(f')df' = \Delta G_0 - \int_0^{f_c} \Delta\lambda(f')df' = 0 \tag{F.6}$$

In contrast, in SMEs where the position of one the ends of the molecular construct is controlled, e.g., the optical tweezers instrument used throughout this thesis, the energy difference between N and U is measured at two different forces, when the molecule is U and N. In par-

ticular, in pulling experiments the molecular construction is driven out of equilibrium from an initial position (f_{min}, λ_{min}) where the molecule is always in its native state to a final position (f_{max}, λ_{max}) where it is always unfolded. The energy difference between these two points, ΔG_λ, is defined as

$$\Delta G_\lambda = G_U^{\lambda_{max}} - G_N^{\lambda_{min}} \tag{F.7}$$

where $G_U^{\lambda_{max}}$ and $G_N^{\lambda_{min}}$ are the free energies of the unfolded and native states relative to the native state at zero force (corresponding to $\lambda = 0$) which is taken as reference state. Free energies in Eq.F.7 are the Legendre transforms of Eq.F.1,

$$G_U^{\lambda_{max}} = G_U^{f_{max}} + \lambda_U(f_{max}) f_{max} \tag{F.8a}$$

$$G_N^{\lambda_{min}} = G_N^{f_{min}} + \lambda_N(f_{min}) f_{min} \ . \tag{F.8b}$$

where $G_{N/U}^{f_{min}/f_{max}}$ is obtained from Eqs. (F.2a) and (F.2b),

$$G_U^{f_{max}} = -\int_0^{f_{max}} \lambda_U(f')df' + \Delta G_0 \tag{F.9a}$$

$$G_N^{f_{min}} = -\int_0^{f_{\min}} \lambda_N(f')df' \ . \tag{F.9b}$$

Introducing Eqs.(F.8a) and (F.8b) and Eqs. (F.9a) and (F.9b) into Eq.F.7 we obtain

$$\begin{aligned} \Delta G_\lambda = & \\ & -\int_0^{f_{max}} \lambda_U(f')df' + \Delta G_0 + \lambda_U(f_{max}) f_{max} + \\ & + \int_0^{f_{min}} \lambda_N(f')df' - \lambda_N(f_{min}) f_{min} \ . \end{aligned} \tag{F.10}$$

Considering the definitions of λ_U and λ_N in Eqs.(F.3a) and (F.3b), and integrating by parts, we obtain the final expression for ΔG_λ:

$$\begin{aligned} \Delta G_\lambda = & \\ & = \Delta G_0 + \sum_{i=b,h} \left[\int_{x_i(f_{min})}^{x_i(f_{max})} f_i(x')dx' \right] + \int_0^{x_U(f_{max})} f_U(x')dx' - \int_0^{x_N(f_{min})} f_N(x')dx' \end{aligned}$$

or, in abbreviated form,

$$\Delta G_\lambda = \Delta G_0 + W_\lambda^{\text{elas}} \tag{F.11a}$$

$$W_\lambda^{elas} = \Delta G_b^{\lambda_{min}\to\lambda_{max}} + \Delta G_h^{\lambda_{min}\to\lambda_{max}} + \Delta G_U^{0\to\lambda_{max}} - \Delta G_N^{0\to\lambda_{min}} \tag{F.11b}$$

being W_λ^{elas} the stretching contributions of the different elements involved in the SMEs. The different terms in Eq.(F.11b) are defined as

$$\Delta G_b^{\lambda_{min}\to\lambda_{max}} = \int_{x_b(f_{min})}^{x_b(f_{max})} f_b(x')dx' \tag{F.12a}$$

$$\Delta G_h^{\lambda_{min}\to\lambda_{max}} = \int_{x_h(f_{min})}^{x_h(f_{max})} f_h(x')dx' \tag{F.12b}$$

$$\Delta G_U^{0\to\lambda_{max}} = \int_{0}^{x_U(f_{max})} f_U(x')dx' \tag{F.12c}$$

$$\Delta G_N^{0\to\lambda_{min}} = \int_{0}^{x_N(f_{min})} f_N(x')dx' \tag{F.12d}$$

where $f_i(x)$ is the inverse function of the elastic response, $x_i(f)$, of the different elements (b, h, U, and N).

BIBLIOGRAPHY

[1] S. Carnot, "Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance," in *Annales scientifiques de l'École Normale Supérieure*, vol. 1, pp. 393–457, 1872.

[2] R. Diesel, "Die entstehung des dieselmotors," in *Jahrbuch der Schiffbautechnischen Gesellschaft*, pp. 267–367, Springer, 1913.

[3] K. Kinosita, R. Yasuda, H. Noji, and K. Adachi, "A rotary molecular motor that can work at near 100% efficiency," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 355, no. 1396, pp. 473–489, 2000.

[4] F. Ritort, "Nonequilibrium fluctuations in small systems: from physics to biology," *Journal of Physics: Condensed Matter*, vol. 18, no. 32, pp. R531–, 2006.

[5] F. Crick and A. Hughes, "The physical properties of cytoplasm," *Experimental Cell Research*, vol. 1, no. 1, pp. 37–80, 1950.

[6] A. Ashkin, "Acceleration and trapping of particles by radiation pressure," *Phys. Rev. Lett.*, vol. 24, pp. 156–159, Jan 1970.

[7] G. Binnig, C. F. Quate, and C. Gerber, "Atomic force microscope," *Physical review letters*, vol. 56, no. 9, p. 930, 1986.

[8] K. C. Neuman and A. Nagy, "Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy," *Nature methods*, vol. 5, no. 6, pp. 491–505, 2008.

[9] C. Gosse and V. Croquette, "Magnetic tweezers: micromanipulation and force measurement at the molecular level," *Biophysical journal*, vol. 82, no. 6, pp. 3314–3329, 2002.

[10] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, "Verification of the crooks fluctuation theorem and recovery of rna folding free energies," *Nature*, vol. 437, no. 7056, pp. 231–234, 2005.

[11] D. T. Edwards, M.-A. LeBlanc, and T. T. Perkins, "Modulation of a protein-folding landscape revealed by afm-based force spectroscopy notwithstanding instrumental limitations," *Proceedings of the National Academy of Sciences*, vol. 118, no. 12, 2021.

[12] G. M. Blackburn, M. J. Gait, D. Loakes, D. M. Williams, J. A. Grasby, M. Egli, A. Flavell, S. Allen, J. Fisher, A. M. Pyle, *et al.*, *Nucleic acids in chemistry and biology*. Royal Society of Chemistry, 2006.

[13] H. Lodish, *Biología celular y molecular*. Ed. Médica Panamericana, 2005.

[14] A. Annunziato, "Dna packaging: nucleosomes and chromatin," *Nature Education*, vol. 1, no. 1, p. 26, 2008.

[15] J. D. Watson and F. H. Crick, "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[16] J. Randall, "Crystallinity in sperm heads: molecular structure of nucleoprotein in vivo.," *Biochimica et biophysica acta*, vol. 10, no. 1, pp. 192–193, 1953.

[17] R. E. Franklin and R. G. Gosling, "Molecular configuration in sodium thymonucleate," *Nature*, vol. 171, no. 4356, pp. 740–741, 1953.

[18] J. C. Wang, "Helical repeat of dna in solution," *Proceedings of the National Academy of Sciences*, vol. 76, no. 1, pp. 200–203, 1979.

[19] V. A. Mitkevich, A. A. Schulga, Y. S. Ermolyuk, V. M. Lobachov, V. O. Chekhov, G. I. Yakovlev, R. W. Hartley, C. N. Pace, M. P. Kirpichnikov, and A. A. Makarov, "Thermodynamics of denatura-

tion of complexes of barnase and binase with barstar," *Biophysical chemistry*, vol. 105, no. 2-3, pp. 383–390, 2003.

[20] A. Matouschek, J. T. Kellis, L. Serrano, and A. R. Fersht, "Mapping the transition state and pathway of protein folding by protein engineering," *Nature*, vol. 340, no. 6229, p. 122, 1989.

[21] C. Martin, V. Richard, M. Salem, R. Hartley, and Y. Mauguen, "Refinement and structural analysis of barnase at 1.5 å resolution," *Acta Crystallographica Section D: Biological Crystallography*, vol. 55, no. 2, pp. 386–398, 1999.

[22] S. d. Lorenzo Ros, *Single molecule studies in a temperature-jump optical trap*. PhD thesis, Universitat de Barcelona. Departament de Física Fonamental, January 2015.

[23] S. B. Smith, Y. Cui, and C. Bustamante, "[7] optical-trap force transducer that operates by direct measurement of light momentum," in *Methods in Enzymology*, vol. 361, pp. 134–162, Elsevier, 2003.

[24] M. Rico-Pasto, I. Pastor, and F. Ritort, "Force feedback effects on single molecule hopping and pulling experiments," *The Journal of Chemical Physics*, vol. 148, no. 12, p. 123327, 2018.

[25] A. M.Monge, M. Manosas, and F. Ritort, "Experimental test of ensemble inequivalence and the fluctuation theorem in the force ensemble in dna pulling experiments," *Physical Review E*, vol. 98, no. 3, p. 032146, 2018.

[26] H. S. Leff and A. F. Rex, *Maxwell's Demon: Entropy, Information, Computing*. Adam Hilger, 1990.

[27] H. Leff and A. F. Rex, *Maxwell's Demon 2 Entropy, Classical and Quantum Information, Computing*. CRC Press, 2002.

[28] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, "Experimental demonstration of information-to-energy conversion and validation of the generalized jarzynski equality," *Nat. Phys.*, vol. 6, no. 12, pp. 988–992, 2010.

[29] G. Paneru, D. Y. Lee, T. Tlusty, and H. K. Pak, "Lossless brownian information engine," *Phys. Rev. Lett.*, vol. 120, no. 2, p. 020601, 2018.

[30] G. Paneru, D. Y. Lee, J.-M. Park, J. T. Park, J. D. Noh, and H. K. Pak, "Optimal tuning of a brownian information engine operating in a nonequilibrium steady state," *Phys. Rev. E*, vol. 98, no. 5, p. 052119, 2018.

[31] T. Admon, S. Rahav, and Y. Roichman, "Experimental realization of an information machine with tunable temporal correlations," *Phys. Rev. Lett.*, vol. 121, no. 18, p. 180601, 2018.

[32] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin, "Experimental realization of a szilard engine with a single electron," *Proceedings of the National Academy of Sciences*, vol. 111, no. 38, pp. 13786–13789, 2014.

[33] J. V. Koski, V. F. Maisi, T. Sagawa, and J. P. Pekola, "Experimental observation of the role of mutual information in the nonequilibrium dynamics of a maxwell demon," *Phys. Rev. Lett.*, vol. 113, no. 3, p. 030601, 2014.

[34] M. D. Vidrighin, O. Dahlsten, M. Barbieri, M. S. Kim, V. Vedral, and I. A. Walmsley, "Photonic maxwell's demon," *Phys. Rev. Lett.*, vol. 116, no. 5, p. 050401, 2016.

[35] K. Chida, S. Desai, K. Nishiguchi, and A. Fujiwara, "Power generator driven by maxwell's demon," *Nature Commun.*, vol. 8, no. 1, pp. 1–7, 2017.

[36] A. Kumar, T.-Y. Wu, F. Giraldo, and D. S. Weiss, "Sorting ultracold atoms in a three-dimensional optical lattice in a realization of maxwell's demon," *Nature*, vol. 561, no. 7721, pp. 83–87, 2018.

[37] M. Ribezzi-Crivellari and F. Ritort, "Large work extraction and the landauer limit in a continuous maxwell demon," *Nat. Phys.*, vol. 15, no. 7, pp. 660–664, 2019.

[38] N. Cottet, S. Jezouin, L. Bretheau, P. Campagne-Ibarcq, Q. Ficheux, J. Anders, A. Auffèves, R. Azouit, P. Rouchon, and B. Huard, "Observing a quantum maxwell demon at work," *Proc. Natl Acad. Sci. USA*, vol. 114, no. 29, pp. 7561–7564, 2017.

[39] Y. Masuyama, K. Funo, Y. Murashita, A. Noguchi, S. Kono, Y. Tabuchi, R. Yamazaki, M. Ueda, and Y. Nakamura, "Information-to-work conversion by maxwell's demon in a superconducting circuit quantum electrodynamical system," *Nature Commun.*, vol. 9, no. 1, pp. 1–6, 2018.

[40] M. Naghiloo, J. J. Alonso, A. Romito, E. Lutz, and K. Murch, "Information gain and loss for a quantum maxwell's demon," *Phys. Rev. Lett.*, vol. 121, no. 3, p. 030604, 2018.

[41] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM J. Res. Develop.*, vol. 5, no. 3, pp. 183–191, 1961.

[42] C. H. Bennett, "Logical reversibility of computation," *IBM J. Res. Develop.*, vol. 17, no. 6, pp. 525–532, 1973.

[43] S. Ciliberto, "Experiments in stochastic thermodynamics: Short history and perspectives," *Phys. Rev. X*, vol. 7, no. 2, p. 021051, 2017.

[44] T. Sagawa and M. Ueda, "Generalized jarzynski equality under nonequilibrium feedback control," *Phys. Rev. Lett.*, vol. 104, no. 9, p. 090602, 2010.

[45] M. Ponmurugan, "Generalized detailed fluctuation theorem under nonequilibrium feedback control," *Phys. Rev. E*, vol. 82, no. 3, p. 031129, 2010.

[46] J. M. Horowitz and J. M. R. Parrondo, "Designing optimal discrete-feedback thermodynamic engines," *New J. Phys.*, vol. 13, no. 12, p. 123019, 2011.

[47] M. Palassini and F. Ritort, "Improving free-energy estimates from unidirectional work measurements: theory and experiment," *Phys.*

Rev. Lett., vol. 107, no. 6, p. 060601, 2011.

[48] T. Sagawa, "Thermodynamic and logical reversibilities revisited," *J. Stat. Mech.*, vol. 2014, no. 3, p. P03025, 2014.

[49] G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," *Phys. Rev. E*, vol. 60, no. 3, p. 2721, 1999.

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

[51] A. Mossa, M. Manosas, N. Forns, J. M. Huguet, and F. Ritort, "Dynamic force spectroscopy of DNA hairpins: I. force kinetics and free energy landscapes," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, p. P02060, feb 2009.

[52] C. B. Anfinsen, E. Haber, M. Sela, and F. White Jr, "The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 47, no. 9, p. 1309, 1961.

[53] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 4096, pp. 223–230, 1973.

[54] C. Levinthal, "Are there pathways for protein folding?," *Journal de chimie physique*, vol. 65, pp. 44–45, 1968.

[55] K. A. Dill, "Dominant forces in protein folding," *Biochemistry*, vol. 29, no. 31, pp. 7133–7155, 1990.

[56] O. Ptitsyn, "Molten globule and protein folding," *Advances in Protein Chemistry*, vol. 47, pp. 83–229, 1995.

[57] R. L. Baldwin and G. D. Rose, "Molten globules, entropy-driven conformational change and protein folding," *Current opinion in structural biology*, vol. 23, no. 1, pp. 4–10, 2013.

[58] M. Dijkstra, W. Fokkink, J. Heringa, E. van Dijk, and S. Abeln, "The characteristics of molten globule states and folding pathways

strongly depend on the sequence of a protein," *Molecular Physics*, vol. 116, no. 21-22, pp. 3173–3180, 2018.

[59] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, "The energy landscapes and motions of proteins," *Science*, vol. 254, no. 5038, pp. 1598–1603, 1991.

[60] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, "Funnels, pathways, and the energy landscape of protein folding: a synthesis," *Proteins: Structure, Function, and Bioinformatics*, vol. 21, no. 3, pp. 167–195, 1995.

[61] C. J. Camacho and D. Thirumalai, "Kinetics and thermodynamics of folding in model proteins," *Proceedings of the National Academy of Sciences*, vol. 90, no. 13, pp. 6369–6372, 1993.

[62] K. A. Dill and J. L. MacCallum, "The protein-folding problem, 50 years on," *Science*, vol. 338, no. 6110, pp. 1042–1046, 2012.

[63] A. Borgia, P. M. Williams, and J. Clarke, "Single-molecule studies of protein folding," *Annu. Rev. Biochem.*, vol. 77, pp. 101–125, 2008.

[64] D. Thirumalai, E. P. O'Brien, G. Morrison, and C. Hyeon, "Theoretical perspectives on protein folding," *Annual Review of Biophysics*, vol. 39, pp. 159–183, 2010.

[65] C. Bustamante, L. Alexander, K. Maciuba, and C. M. Kaiser, "Single-molecule studies of protein folding with optical tweezers," *Annual Review of Biochemistry*, vol. 89, no. 1, pp. 443–470, 2020. PMID: 32569525.

[66] P. Cossio, G. Hummer, and A. Szabo, "Transition paths in single-molecule force spectroscopy," *The Journal of Chemical Physics*, vol. 148, no. 12, p. 123309, 2018.

[67] C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco, and C. Bustamante, "The ribosome modulates nascent protein folding," *Science*, vol. 334, no. 6063, pp. 1723–1727, 2011.

[68] E. Infante, A. Stannard, S. J. Board, P. Rico-Lastres, E. Rostkova, A. E. Beedle, A. Lezamiz, Y. J. Wang, S. G. Breen, F. Panagaki, V. Sundar, C. Shanahan, P. Roca-Cusachs, and S. Garcia-Manyes, "The mechanical stability of proteins regulates their translocation rate into the cell nucleus," *Nature Physics*, vol. 15, no. 9, pp. 973–981, 2019.

[69] C. Cecconi, E. A. Shank, C. Bustamante, and S. Marqusee, "Direct observation of the three-state folding of a single protein molecule," *Science*, vol. 309, no. 5743, pp. 2057–2060, 2005.

[70] F. Stockmar, A. Y. Kobitski, and G. U. Nienhaus, "Fast folding dynamics of an intermediate state in rnase h measured by single-molecule fret," *The Journal of Physical Chemistry B*, vol. 120, no. 4, pp. 641–649, 2016.

[71] J. C. M. Gebhardt, T. Bornschlögl, and M. Rief, "Full distance-resolved folding energy landscape of one single protein molecule," *Proceedings of the National Academy of Sciences*, vol. 107, no. 5, pp. 2013–2018, 2010.

[72] K. Neupane, A. P. Manuel, and M. T. Woodside, "Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes," *Nature Physics*, vol. 12, no. 7, pp. 700–703, 2016.

[73] J. Stigler, F. Ziegler, A. Gieseke, J. C. M. Gebhardt, and M. Rief, "The complex folding network of single calmodulin molecules," *Science*, vol. 334, no. 6055, pp. 512–516, 2011.

[74] P. J. Elms, J. D. Chodera, C. Bustamante, and S. Marqusee, "The molten globule state is unusually deformable under mechanical force," *Proceedings of the National Academy of Sciences*, vol. 109, no. 10, pp. 3796–3801, 2012.

[75] S. Lindhoud, M. Pirchi, A. H. Westphal, G. Haran, and C. P. van Mierlo, "Gradual folding of an off-pathway molten globule detected at the single-molecule level," *Journal of molecular biology*,

vol. 427, no. 19, pp. 3148–3157, 2015.

[76] T. Sosnick, L. Mayne, R. Hiller, and S. Englander, "The barriers in protein folding," *Nature Structural Biology*, vol. 1, no. 3, pp. 149–156, 1994.

[77] A. R. Fersht, "Characterizing transition states in protein folding: an essential step in the puzzle," *Current opinion in structural biology*, vol. 5, no. 1, pp. 79–84, 1995.

[78] A. Tikhomirova, N. Taulier, and T. V. Chalikian, "Energetics of nucleic acid stability: The effect of δ cp," *Journal of the American Chemical Society*, vol. 126, no. 50, pp. 16387–16394, 2004.

[79] C. M. Johnson, "Differential scanning calorimetry as a tool for protein folding and stability," *Archives of Biochemistry and Biophysics*, vol. 531, no. 1-2, pp. 100–109, 2013.

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

[81] W. Stephenson, S. Keller, R. Santiago, J. E. Albrecht, P. N. Asare-Okai, S. A. Tenenbaum, M. Zuker, and P. T. Li, "Combining temperature and force to study folding of an rna hairpin," *Physical Chemistry Chemical Physics*, vol. 16, no. 3, pp. 906–917, 2014.

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

[83] C. Jarzynski, "Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale," *Annu. Rev. Condens. Matter Phys.*, vol. 2, no. 1, pp. 329–351, 2011.

[84] M. Palassini and F. Ritort, "Improving free-energy estimates from unidirectional work measurements: theory and experiment," *Physical Review Letters*, vol. 107, no. 6, p. 060601, 2011.

[85] G. I. Bell, "Models for the specific adhesion of cells to cells," *Science*, vol. 200, no. 4342, pp. 618–627, 1978.

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

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

[88] W. Pfeil, *Protein stability and folding supplement 1: a collection of thermodynamic data*. Springer Science & Business Media, 2012.

[89] J. J. Galano-Frutos and J. Sancho, "Accurate calculation of barnase and snase folding energetics using short molecular dynamics simulations and an atomistic model of the unfolded ensemble: Evaluation of force fields and water models," *Journal of Chemical Information and Modeling*, vol. 59, no. 10, pp. 4350–4360, 2019.

[90] Y. Bai, T. Sosnick, L. Mayne, and S. Englander, "Protein folding intermediates: native-state hydrogen exchange," *Science*, vol. 269, no. 5221, pp. 192–197, 1995.

[91] R. L. Baldwin, "The nature of protein folding pathways: The classical versus the new view," *Journal of Biomolecular NMR*, vol. 5, pp. 103 – 109, 1995.

[92] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, "Reversible unfolding of individual titin immunoglobulin domains by afm," *Science*, vol. 276, no. 5315, pp. 1109–1112, 1997.

[93] M. S. Z. Kellermayer, S. B. Smith, H. L. Granzier, and C. Bustamante, "Folding-unfolding transitions in single titin molecules characterized with laser tweezers," *Science*, vol. 276, no. 5315, pp. 1112–1116, 1997.

[94] A. A. Rebane, L. Ma, and Y. Zhang, "Structure-based derivation of protein folding intermediates and energies from optical tweezers," *Biophysical Journal*, vol. 110, no. 2, pp. 441–454, 2016.

[95] D. Sharma, O. Perisic, Q. Peng, Y. Cao, C. Lam, H. Lu, and H. Li, "Single-molecule force spectroscopy reveals a mechanically stable protein fold and the rational tuning of its mechanical stability," *Proceedings of the National Academy of Sciences*, vol. 104, no. 22, pp. 9278–9283, 2007.

[96] P. Zheng, Y. Wang, and H. Li, "Reversible unfolding–refolding of rubredoxin: A single-molecule force spectroscopy study," *Angewandte Chemie International Edition*, vol. 53, no. 51, pp. 14060–14063, 2014.

[97] D. H. Goldman, C. M. Kaiser, A. Milin, M. Righini, I. Tinoco, and C. Bustamante, "Mechanical force releases nascent chain–mediated ribosome arrest in vitro and in vivo," *Science*, vol. 348, no. 6233, pp. 457–460, 2015.

[98] H. Yu, M. G. W. Siewny, D. T. Edwards, A. W. Sanders, and T. T. Perkins, "Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins," *Science*, vol. 355, no. 6328, pp. 945–950, 2017.

[99] R. Merkel, P. Nassoy, A. Leung, K. Ritchie, and E. Evans, "Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy," *Nature*, vol. 397, no. 6714, p. 50, 1999.

[100] E. Evans, "Probing the relation between force—lifetime—and chemistry in single molecular bonds," *Annual Review of Biophysics and Biomolecular structure*, vol. 30, no. 1, pp. 105–128, 2001.

[101] A. A. Makarov, I. I. Protasevich, N. V. Kuznetsova, B. B. Fedorov, S. V. Korolev, N. K. Struminskaya, N. P. Bazhulina, I. B. Leshchinskaya, R. W. Hartley, M. P. Kirpichnikov, G. I. Yakovlev, and N. G. Esipova, "Comparative study of thermostability and structure of close homologues - barnase and binase," *Journal of Biomolecular Structure and Dynamics*, vol. 10, no. 6, pp. 1047–1065, 1993. PMID: 8357541.

[102] Y. V. Griko, G. I. Makhatadze, P. L. Privalov, and R. W. Hartley, "Thermodynamics of barnase unfolding," *Protein Science*, vol. 3,

no. 4, pp. 669–676, 1994.

[103] J. E. Leffler, "Parameters for the description of transition states," *Science*, vol. 117, no. 3039, pp. 340–341, 1953.

[104] G. S. Hammond, "A correlation of reaction rates," *Journal of the American Chemical Society*, vol. 77, no. 2, pp. 334–338, 1955.

[105] B. Alberts, "Dna replication and recombination," *Nature*, vol. 421, no. 6921, pp. 431 – 435, 2003.

[106] E. C. Friedberg, G. C. Walker, W. Siede, and R. D. Wood, *DNA repair and mutagenesis*. American Society for Microbiology Press, 2005.

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

[108] T. T. Le and H. D. Kim, "Probing the elastic limit of DNA bending," *Nucleic Acids Research*, vol. 42, pp. 10786–10794, 08 2014.

[109] A. Idili, F. Ricci, and A. Vallée-Bélisle, "Determining the folding and binding free energy of DNA-based nanodevices and nanoswitches using urea titration curves," *Nucleic Acids Research*, vol. 45, pp. 7571–7580, 06 2017.

[110] J. Camunas-Soler, A. Alemany, and F. Ritort, "Experimental measurement of binding energy, selectivity, and allostery using fluctuation theorems," *Science*, vol. 355, no. 6323, pp. 412–415, 2017.

[111] X. Viader-Godoy, M. Manosas, and F. Ritort, "Sugar-pucker force-induced transition in single-stranded dna," *International Journal of Molecular Sciences*, vol. 22, no. 9, 2021.

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

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

[114] D. C. Brydges and P. A. Martin, "Coulomb systems at low density: A review," *Journal of Statistical Physics*, vol. 96, no. 5-6, pp. 1163–1330, 1999.

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

[116] J. M. Huguet, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort, "Single-molecule derivation of salt dependent base-pair free energies in dna," *Proceedings of the National Academy of Sciences*, vol. 107, no. 35, pp. 15431–15436, 2010.

[117] A. Cooper, M. A. Nutley, and A. Wadood, "Differential scanning microcalorimetry," *Protein-ligand interactions: Hydrodynamics and calorimetry*, pp. 287–318, 2000.

[118] E. Freire, O. L. Mayorga, and M. Straume, "Isothermal titration calorimetry," *Analytical chemistry*, vol. 62, no. 18, pp. 950A–959A, 1990.

[119] L. A. Marky and K. J. Breslauer, "Calorimetric determination of base-stacking enthalpies in double-helical dna molecules," *Biopolymers: Original Research on Biomolecules*, vol. 21, no. 11, pp. 2185–2194, 1982.

[120] K. J. Breslauer, R. Frank, H. Blöcker, and L. A. Marky, "Predicting dna duplex stability from the base sequence," *Proceedings of the National Academy of Sciences*, vol. 83, no. 11, pp. 3746–3750, 1986.

[121] G. Vesnaver and K. J. Breslauer, "The contribution of dna single-stranded order to the thermodynamics of duplex formation.," *Proceedings of the National Academy of Sciences*, vol. 88, no. 9, pp. 3569–

3573, 1991.

[122] J. A. Holbrook, M. W. Capp, R. M. Saecker, and M. T. Record, "Enthalpy and heat capacity changes for formation of an oligomeric dna duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices," *Biochemistry*, vol. 38, no. 26, pp. 8409–8422, 1999.

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

[124] A. Tikhomirova, N. Taulier, and T. V. Chalikian, "Energetics of nucleic acid stability: The effect of δ cp," *Journal of the American Chemical Society*, vol. 126, no. 50, pp. 16387–16394, 2004.

[125] T. V. Chalikian, J. Völker, G. E. Plum, and K. J. Breslauer, "A more unified picture for the thermodynamics of nucleic acid duplex melting: a characterization by calorimetric and volumetric techniques," *Proceedings of the National Academy of Sciences*, vol. 96, no. 14, pp. 7853–7858, 1999.

[126] P. Vaitiekunas, C. Crane-Robinson, and P. L. Privalov, "The energetic basis of the dna double helix: a combined microcalorimetric approach," *Nucleic acids research*, vol. 43, no. 17, pp. 8577–8589, 2015.

[127] J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics," *Proceedings of the National Academy of Sciences*, vol. 95, no. 4, pp. 1460–1465, 1998.

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

[129] J. M. Huguet, M. Ribezzi-Crivellari, C. V. Bizarro, and F. Ritort, "Derivation of nearest-neighbor DNA parameters in magnesium from single molecule experiments," *Nucleic Acids Research*, vol. 45, pp. 12921–12931, 11 2017.

[130] A. Alemany and F. Ritort, "Force-dependent folding and unfolding kinetics in dna hairpins reveals transition-state displacements along a single pathway," *The Journal of Physical Chemistry Letters*, vol. 8, no. 5, pp. 895–900, 2017.

[131] K. D. Whitley, M. J. Comstock, and Y. R. Chemla, "Elasticity of the transition state for oligonucleotide hybridization," *Nucleic Acids Research*, vol. 45, pp. 547–555, 11 2016.

[132] A. Kornberg, T. A. Baker, *et al.*, *DNA replication*, vol. 3. Wh Freeman New York, 1992.

[133] R. K. Saiki, T. L. Bugawan, G. T. Horn, K. B. Mullis, and H. A. Erlich, "Analysis of enzymatically amplified β-globin and hla-dqα dna with allele-specific oligonucleotide probes," *Nature*, vol. 324, pp. 163–166, 1986.

[134] H. DeVoe and I. Tinoco Jr, "The stability of helical polynucleotides: base contributions," *Journal of Molecular Biology*, vol. 4, no. 6, pp. 500–517, 1962.

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

[136] A. Serganov and E. Nudler, "A decade of riboswitches," *Cell*, vol. 152, no. 1-2, pp. 17–24, 2013.

[137] J. Kortmann and F. Narberhaus, "Bacterial rna thermometers: molecular zippers and switches," *Nature reviews microbiology*, vol. 10, no. 4, pp. 255–265, 2012.

[138] S. S. Krajewski and F. Narberhaus, "Temperature-driven differential gene expression by rna thermosensors," *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1839, no. 10, pp. 978–988, 2014.

[139] H. Maity, M. Maity, M. M. G. Krishna, L. Mayne, and S. W. Englander, "Protein folding: The stepwise assembly of foldon units,"

Proceedings of the National Academy of Sciences, vol. 102, no. 13, pp. 4741–4746, 2005.

[140] R. L. Baldwin, "Clash between energy landscape theory and foldon-dependent protein folding," *Proceedings of the National Academy of Sciences*, vol. 114, no. 32, pp. 8442–8443, 2017.

[141] F. S. Ruggeri, J. Habchi, A. Cerreta, and G. Dietler, "Afm-based single molecule techniques: unraveling the amyloid pathogenic species," *Current pharmaceutical design*, vol. 22, no. 26, pp. 3950–3970, 2016.

[142] D. Sluysmans, P. Lussis, C.-A. Fustin, A. Bertocco, D. A. Leigh, and A.-S. Duwez, "Real-time fluctuations in single-molecule rotaxane experiments reveal an intermediate weak binding state during shuttling," *Journal of the American Chemical Society*, vol. 143, no. 5, pp. 2348–2352, 2021.

[143] F. Kriegel, N. Ermann, and J. Lipfert, "Probing the mechanical properties, conformational changes, and interactions of nucleic acids with magnetic tweezers," *Journal of structural biology*, vol. 197, no. 1, pp. 26–36, 2017.

[144] N. Dahal, J. Nowitzke, A. Eis, and I. Popa, "Binding-induced stabilization measured on the same molecular protein substrate using single-molecule magnetic tweezers and heterocovalent attachments," *The Journal of Physical Chemistry B*, vol. 124, no. 16, pp. 3283–3290, 2020.

[145] D. Choudhary, A. Mossa, M. Jadhav, and C. Cecconi, "Biomolecular applications of recent developments in optical tweezers," *Biomolecules*, vol. 9, no. 1, p. 23, 2019.

[146] C. J. Bustamante, Y. R. Chemla, S. Liu, and M. D. Wang, "Optical tweezers in single-molecule biophysics," *Nature Reviews Methods Primers*, vol. 1, no. 1, pp. 1–29, 2021.

[147] R. B. Best and G. Hummer, "Reaction coordinates and rates from transition paths," *Proceedings of the National Academy of Sciences*,

vol. 102, no. 19, pp. 6732–6737, 2005.

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

[149] J.-D. Wen, M. Manosas, P. T. Li, S. B. Smith, C. Bustamante, F. Ritort, and I. Tinoco, "Force unfolding kinetics of rna using optical tweezers. i. effects of experimental variables on measured results," *Biophysical Journal*, vol. 92, no. 9, pp. 2996–3009, 2007.

[150] M. Manosas, J.-D. Wen, P. Li, S. Smith, C. Bustamante, I. Tinoco, and F. Ritort, "Force unfolding kinetics of rna using optical tweezers. ii. modeling experiments," *Biophysical Journal*, vol. 92, no. 9, pp. 3010–3021, 2007.

[151] M. Manosas, D. Collin, and F. Ritort, "Force-dependent fragility in rna hairpins," *Physical Review Letters*, vol. 96, no. 21, p. 218301, 2006.

[152] C. V. Bizarro, A. Alemany, and F. Ritort, "Non-specific binding of na+ and mg 2+ to rna determined by force spectroscopy methods," *Nucleic acids research*, vol. 40, no. 14, pp. 6922–6935, 2012.

[153] H. A. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions," *Physica*, vol. 7, no. 4, pp. 284–304, 1940.

[154] R. Mishra, A. Kumar, R. Chandra, and D. Kumar, "A review on theoretical studies of various types of drug-dna interaction," *Int J Sci Technol Soc*, vol. 3, pp. 11–27, 2017.

[155] S. R. Hengel, M. A. Spies, and M. Spies, "Small-molecule inhibitors targeting dna repair and dna repair deficiency in research and cancer therapy," *Cell Chemical Biology*, vol. 24, no. 9, pp. 1101–1119, 2017.

[156] V. Rives, M. del Arco, and C. Martín, "Intercalation of drugs in layered double hydroxides and their controlled release: A review," *Applied Clay Science*, vol. 88, pp. 239–269, 2014.

[157] J. Camunas-Soler, M. Manosas, S. Frutos, J. Tulla-Puche, F. Albericio, and F. Ritort, "Single-molecule kinetics and footprinting of dna bis-intercalation: the paradigmatic case of thiocoraline," *Nucleic Acids Research*, vol. 43, no. 5, pp. 2767–2779, 2015.

[158] C. Zimmer, "Effects of the antibiotics netropsin and distamycin a on the structure and function of nucleic acids," in *Progress in nucleic acid research and molecular biology*, vol. 15, pp. 285–318, Elsevier, 1975.

[159] J. Portugal, "Footprinting analysis of sequence-specific dna-drug interactions," *Chemico-Biological Interactions*, vol. 71, no. 4, pp. 311–324, 1989.

[160] A. J. Hampshire, D. A. Rusling, V. J. Broughton-Head, and K. R. Fox, "Footprinting: a method for determining the sequence selectivity, affinity and kinetics of dna-binding ligands," *Methods*, vol. 42, no. 2, pp. 128–140, 2007.

[161] K. Teilum, M. B. A. Kunze, S. Erlendsson, and B. B. Kragelund, "(s) pinning down protein interactions by nmr," *Protein Science*, vol. 26, no. 3, pp. 436–451, 2017.

[162] A. Kellett, Z. Molphy, C. Slator, V. McKee, and N. P. Farrell, "Molecular methods for assessment of non-covalent metallodrug–dna interactions," *Chemical Society Reviews*, vol. 48, no. 4, pp. 971–988, 2019.

[163] D. Gendrel, J. Raymond, J. Coste, F. Moulin, M. Lorrot, S. Guerin, S. Ravilly, H. Lefevre, C. Royer, C. Lacombe, *et al.*, "Comparison of procalcitonin with c-reactive protein, interleukin 6 and interferon-alpha for differentiation of bacterial vs. viral infections," *The Pediatric Infectious Disease Journal*, vol. 18, no. 10, pp. 875–881, 1999.

[164] T. Suren, D. Rutz, P. Mößmer, U. Merkel, J. Buchner, and M. Rief, "Single-molecule force spectroscopy reveals folding steps associated with hormone binding and activation of the glucocorticoid receptor," *Proceedings of the National Academy of Sciences*, vol. 115, no. 46, pp. 11688–11693, 2018.

[165] M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura, and R. E. Dickerson, "The molecular origin of dna-drug specificity in netropsin and distamycin," *Proceedings of the National Academy of Sciences*, vol. 82, no. 5, pp. 1376–1380, 1985.

[166] M. Freyer, R. Buscaglia, D. Cashman, S. Hyslop, W. Wilson, J. Chaires, and E. Lewis, "Binding of netropsin to several dna constructs: Evidence for at least two different 1:1 complexes formed from an –aatt-containing ds-dna construct and a single minor groove binding ligand," *Biophysical Chemistry*, vol. 126, no. 1, pp. 186–196, 2007.

[167] M. Durand, N. T. Thuong, and J. C. Maurizot, "Binding of netropsin to a dna triple helix.," *Journal of Biological Chemistry*, vol. 267, no. 34, pp. 24394–24399, 1992.

[168] J. Lah and G. Vesnaver, "Binding of distamycin a and netropsin to the 12mer dna duplexes containing mixed at·gc sequences with at most five or three successive at base pairs," *Biochemistry*, vol. 39, no. 31, pp. 9317–9326, 2000.

[169] M. Rettig, M. W. Germann, M. A. Ismail, A. Batista-Parra, M. Munde, D. W. Boykin, and W. D. Wilson, "Microscopic rearrangement of bound minor groove binders detected by nmr," *The Journal of Physical Chemistry B*, vol. 116, no. 19, pp. 5620–5627, 2012.

[170] Y. S. Babayan, V. Kalantaryan, S. Hakobyan, and R. Ghazaryan, "Interaction of netropsin with double-stranded nucleic acids irradiated with nonionizing athermal millimeter electromagnetic waves," *Armenian Journal of Physics*, vol. 9, no. 2, pp. 167–172, 2016.

[171] C. Cecconi, E. A. Shank, S. Marqusee, and C. Bustamante, "Dna molecular handles for single-molecule protein-folding studies by optical tweezers," in *DNA Nanotechnology*, pp. 255–271, Springer, 2011.