Dynamic force spectroscopy and folding kinetics in molecular systems

Anna Alemany i Arias

Dynamic force spectroscopy and folding kinetics in molecular systems

Anna Alemany i Arias

Director: Dr. Fèlix Ritort i Farran
Tutor: Dr. Fèlix Ritort i Farran
Programa de doctorat en Física
Memòria presentada per optar al títol de Doctor en Física.
Setembre de 2014

Haig d'anar-me'n, però començaria a estudiar-ho tot.

Carlos Alemany Martínez
Barcelona, agost de 2012

Agraïments

En primer lloc voldria agrair al Dr. Felix Ritort el seu suport durant els anys que ha durat la realització d'aquesta tesi. Una de les primeres coses que vaig aprendre d'ell, i que en el seu moment em va sobtar moltíssim, és que la diversió comença quan teoria i experiments estan en aparent desacord. La perseverança, la paciència, i la passió que sempre transmet han estat un gran estímul per continuar endavant en els millors i els pitjors moments, tant a escala personal com professional. Voldria també agrair-li la confiança que des d'un bon inici va dipositar en mi, i totes les oportunitats docents i investigadores que m'ha donat. De tot cor, gràcies Felix.

També voldria agrair a totes les persones que han passat pel *Small Biosystems Lab*. D'una manera o altra, tots i cadascun d'ells m'ha ajudat a tirar endavant la meva feina. L'Alessandro Mossa i la Núria Forns van ser uns pilars fonamentals en els meus inicis: amb ells vaig aprendre molts dels secrets de l'anàlisi de dades i de la síntesi de les molècules de DNA, ambdós han resultat imprescindibles per tirar endavant la meva feina. A la Sara li vull donar les gràcies per la seva afectuosa acollida al despatx, i per totes aquelles hores de diversió i història que vam passar juntes escrivint el projecte de màster. Al Josep Maria i al Cristiano els hi dec els seus consells per afrontar les hores d'experiments; realment no hi ha ningú com vosaltres per transmetre'm la calma i els ànims necessaris per continuar al laboratori quan tot sembla que es torça. Val a dir que en aquests moments també m'ajuda recordar les hores que vam passar amb la Gloria mesurant. Sense la Carol, la Blanca i la Sílvia, sintetitzar i estirar la Barnasa no hauria estat possible. Jo esperava i donava ànims, però realment sense la seva feinada no hauria pogut tirar endavant aquest projecte. A l'Ivan li vull agrair els seus ànims a l'hora d'afrontar segons quins reptes. Realmente escribir ese artículo fue un buen ejercicio para mi. De l'exigència i curiositat de la Sandra en vaig aprendre que encara que una cosa estigui acabada sempre pot millorar-se. D'altra banda, amb el Marco hem passat hores parlant de física i de projectes de tot tipus que sempre ajuden a retrobar la inspiració. Segurament mai s'oblidarà el punt àlgid quan, amb ell i el Lorenzo, vam compartir una idíl·lica copa davant del Pacífic tot veient la posta de sol. *Grande*. Amb la Maria Mañosas vam fer un viatge molt i molt llarg. Per mi eren dies de buidor, però la Maria em va ajudar a trobar bellesa i alegria entre tanta neu, tant de fred, i tantes hamburgueses –dues– a l'*Annie's*. La Maria Martí va arribar quan jo duia mitja tesi. En ella he trobat una companya de laboratori i una gran amiga. No sé fins a quin punt se li pot agrair a algú així totes les valuoses estones que hem passat juntes. Finalment, el

Joan, el gran company de viatge. Amb ell vam començar i ara acabem la gran empresa que és el doctorat. Junts ens hem animat, ens hem desfogat, ens hem plantejat un munt de reptes, i no hem parat fins a resoldre'ls. És gràcies a la seva insistència que em vaig tornar a apassionar per la ciència quan absolutament res aconseguia despertar el meu interès. I ara deixem pas als nouvinguts! A l'Elisa, la Isabel, el Marc i l'Álvaro, que arriben i revolucionen, comencen nous projectes i contagien el seu entusiasme. I a la Lara perquè he après un munt de coses i m'ho he passat genial amb tu i el teu projecte final de grau. A tots vosaltres, gràcies per les estones compartides a despatxos, passadissos, laboratoris, dinars, sopars, excursions, calçotades, on sigui.

Durant els anys que ha durat aquesta tesi doctoral he tingut l'oportunitat de dedicar part del meu temps a la docència i a engegar projectes com les *Flash Talks* o la JIPI, que m'han permès explorar facetes de la ciència més enllà de l'estricta investigació. Vull agrair, encara que sigui una mica anònimament, a tots els estudiants de llicenciatura i grau que han aguantat amb bon humor les meves classes a hores que sempre són intempestives (perquè tots sabem que a primera hora és difícil, que abans de dinar costa, i que després de dinar és impossible). A més a més, vull donar les gràcies al Pau, Marta, Josep i Enric perquè junts vam iniciar un projecte estimulant i enriquidor, i a tots els JIPIs –Albert, Marc, Joan, Blai, David, Narcís, Rubén, Daniel P., José Manuel, Oleguer, Maria, Anna, Daniel R., i Daniel S.– amb qui hem construït un congrés ambiciós i inclusiu, que personalment penso que està ple d'idees engrescadores. Gràcies a tots vosaltres he après a valorar una bona organització i a apreciar la feina en equip ben feta.

Al David, més que donar-li les gràcies, voldria recordar-li els primers mesos de tesi. Cada dia descobríem una cosa nova i corríem l'un al despatx de l'altra a explicar-nos-la. La il·lusió d'aleshores és un motor que encara funciona. A ell, junt amb la Marta, el Víctor, el Saül, el Miguel, l'Ausiàs, la Maria, el Joan, la Griselda, la Macarena i el Ricardo els hi vull agrair molt sincerament la quantitat de divendres que hem passat junts. Perquè hem menjat, hem begut, hem ballat, hem rigut, hem jugat, i hem renovat forces setmanalment. I aquesta és una d'aquelles tradicions que tant de bo duri per sempre. En la Magalí també hi he trobat sempre una font de renovació important (i sovint poc convencional): un viatge a New York, una ascensió al Pedraforca, un estiu a Menorca, i ara amb el Marc i la petita Arlet comença una nova i gran aventura que espero seguir de molt aprop. A la Júlia, la Sílvia, l'Artur, el Rai, l'Alba, la Carol, el Jörg, la Marina, el Joan Antoni i l'Àngel, us he vist menys del que hauria volgut, però heu estat sempre un punt de suport i valoro molt totes les estones que hem passat junts, que heu suportat els meus daltabaixos i que heu dit "oh que interessant" o "ànims Anna que tu pots" quan per mi era important escoltar-ho. Finalment també vull donar-li les gràcies a la Matilde, perquè la seva curiositat sense límits des que era ben petita sempre ha estat molt estimulant. A ella li vull donar molts ànims ara que comença l'etapa universitària.

També voldria mencionar en aquests agraïments a la Milagros, per veure tan clar que la física m'agradaria; al Jordi Barba, per suportar diàriament i amb un somriure les meves bromes matineres; i a l'oncle Josep, a la Mercè, a l'Annie i al Joan Josep, que han acompanyat la meva família sempre.

En darrer lloc, vull agrair a la meva família el seu suport i la seva paciència en el transcurs de la meva tesi. Al Carles li vull agrair aquells mails que m'envia on apareixen paraules com "encara", o "espavila" o "ànims". Al Jordi li vull donar les gràcies per la seva companyia, els seus ànims, la seva poca mandra a anar amunt i avall quan li he demanat, i la magnífica portada que m'ha fet. A ma mare li vull agrair tot el temps que m'ha dedicat, la seva fortalesa, i la seva comprensió quan he estat massa capficada per a veure que potser ella em necessitava a mi. I al meu pare li vull agrair pràcticament tot. Els seus llibres oportuns en els moments més adequats, la seva set inspiradora i inacabable de coneixement, la seva impressionant capacitat de treball, la tenacitat excepcional davant de qualsevol entrebanc, les hores de companyia que ens hem fet l'un a l'altre, la confiança que mai va dissimular en mi i en nosaltres... Així que Pàpet, aquí tenim la nostra tesi.

Portopetro, agost de 2014.

Contents

Resum de la tesi en català

La codificació de la informació genètica, la regulació de l'expressió dels gens, el transport de nutrients dins la cèllula, la protecció immunològica contra agents infecciosos... Aquests són alguns dels processos moleculars que s'esdevenen en organismes vius i són crucials per la seva supervivència. Com més aprenem sobre com les molècules duen a terme les seves tasques biològiques associades, més ens sorprèn la seva gran precisió i eficiència. Entendre el funcionament d'aquests fenòmens és important des de com a mínim dos punts de vista. (1) Des d'una perspectiva biològica, l'origen de moltes malalties podria entendre's. Per exemple, actualment es creu que algunes malalties neurodegeneratives com l'Alzheimer s'inicien amb el plegament incorrecte i la posterior oligomerització d'unes proteïnes que en condicions normals són solubles. Així, l'estudi dels mecanismes que regulen el plegament molecular pot ser decisiu per no només descobrir l'origen, sinó també prevenir i fins i tot trobar la cura de moltes malalties. (2) Des del punt de vista de la física teòrica, un coneixement precís de les lleis que governen el món microscòpic és interessant de per si. A més, aquests descobriments podrien produir una revolució tecnològica amb dissenys de motors microscòpics altament eficients o tècniques computacionals basades en l'ADN.

Els avenços aconseguits en les últimes dècades en l'àmbit de la nanotecnologia han fet possible l'observació i la manipulació de processos moleculars a escala de molècula i trajectòria individuals, amb una resolució espaio-temporal sense precedents. Una tècnica experimental paradigmàtica per dur a terme aquest tipus de mesures són les "pinces òptiques", consistents en un feix de llum coherent i focalitzat capaç d'atrapar partícules de plàstic microscòpiques utilitzant la conservació del moment. Aquest instrument permet manipular una única biomolècula amb precisió nanomètrica, i exercir-hi forces en el rang entre 0 i 100 pN.

La diversitat d'experiments que poden dur-se a terme utilitzant les pinces òptiques augmenta cada dia. Des de l'estirament d'una molècula d'ADN per extreure'n la resposta elàstica, fins a l'observació a temps real del cicle operatiu d'un motor molecular en un substrat *in vitro*, l'estudi de sistemes moleculars mitjançant l'ús de les pinces òptiques ha promogut una revolució en el camp de la biofísica molecular. En aquesta tesi, s'utilitzen per desxifrar els mecanismes del plegament i desplegament d'àcids nucleics i d'una proteïna quan s'hi aplica una força als extrems. A més, les propietats de la unió entre un antigen i un anticòs s'investiguen de manera qualitativa, mesurant la correlació entre l'afinitat i l'elasticitat de la interacció. En tots els experiments realitzats en molècules individuals les fluctuacions tèrmiques tenen un paper molt important. De fet, són crucials en l'escala microscòpica i contenen informació molt valuosa sobre la termodinàmica i la cinètica dels sistemes moleculars.

Quan s'apliquen forces mecàniques a sistemes moleculars, aquests poden experimentar transicions entre diferents estats en un rang de forces característic. Per exemple, les proteïnes i les forquilles d'àcids nucleics es desnaturalitzen quan la força aplicada és prou gran, i es tornen a formar en el moment que la força és suficientment baixa. Així doncs, podem induir mecànicament transicions mecàniques entre l'estat natiu i l'estat

desplegat d'una donada molècula. En aquests processos de plegament i desplegament poden també observar-se, per determinades molècules, estats intermedis i mal plegats. L'aplicació d'una força pot promoure també la desunió d'un lligand adherit a la doble hèlix d'una molècula d'ADN, o pot provocar el trencament d'un enllaç entre un antigen i un anticòs. L'estudi d'aquestes forces característiques per a cada sistema rep el nom d'"espectrosòpia dinàmica de forces", i ocupa la primera part d'aquesta tesi. Primer s'han obtingut les propietats elàstiques de molècules curtes d'ADN d'única cadena, cosa que fins ara no s'havia aconseguit amb les tècniques tradicionals. L'elasticitat juga un paper important en molts processos moleculars, i és sabut que les molècules petites d'àcids nucleics estan involucrades en molts fenòmens de regulació genòmica. És per això que aquests resultats poden resultar molt útils a la comunitat científica. En segon lloc, la velocitat cinètica del desplegament i plegament mecànic de forquilles d'àcids nucleics s'han obtingut utilitzant models de Markov a partir de les forces de desplegament i plegament, cosa que permet extreure informació important sobre el paisatge d'energia lliure molecular i el camí de reacció seguit per la molècula. En aquest context, s'han estudiat diversos processos que impliquen la presència de forquilles d'ADN i ARN, com l'efecte no-específic de ions monovalents i divalents en l'estabilitat molecular o la interacció específica d'un lligand a la doble hèlix de l'ADN. A més, s'han mesurat les propietats cinètiques i termodinàmiques d'estats intermedis que només poden generar-se en processos dinàmics. En tercer lloc, s'ha investigat la correlació entre la força d'un enllaç (a través de les forces de trencament) i la flexibilitat de l'enllaç (extretes de les corbes experimentals de força-extensió) utilitzant un antigen determinat i anticossos obtinguts en diferents moments de la línia de maduració del sistema immune d'un conill. Els resultats suggereixen que com major és la força, major és l'afinitat i més rígid és l'enllaç. Finalment, s'ha investigat el desplegament i plegament mecànic de la proteïna Barnasa utilitzant experiments de no-equilibri i equilibri, respectivament. Els resultats obtinguts permeten caracteritzar l'estat de transició i obtenir una estima de l'energia de formació de la molècula.

En la segona part d'aquesta tesi doctoral s'estudien les "relacions de fluctuació" (RF), que són igualtats fonamentals entre el treball realitzat en un sistema termodinàmic i la diferència d'energia lliure ΔG entre l'estat inicial i el final. De fet, quan una molècula és "estirada" o "alliberada" amb l'acció d'una força utilitzant pinces òptiques, s'efectua una transformació termodinàmica entre dos estats molt ben definits: l'estat natiu a forces baixes, i l'estat desnaturalitzat a forces elevades. Si es realitzen repeticions idèntiques i independents dels protocols d'"estirament" i "alliberació" adiabàticament, els resultats són idèntics i el treball realitzat és igual a la diferència d'energia lliure entre l'estat inicial i l'estat final. En canvi, quan els experiments es realitzen en condicions d'irreversibilitat (és a dir, fora de l'equilibri) el sistema molecular pot seguir trajectòries diferents i el treball realitzat pot variar en cada repetició a causa de les fluctuacions tèrmiques. D'acord amb la segona llei de la termodinàmica, la mitja dels valors obtinguts pels treballs ha de ser més gran que ΔG. En aquesta situació, es mostra en aquesta tesi que les RF són molt útils per obtenir mesures d'energia a partir dels treballs mesurats en infinites realitzacions de processos idèntics duts a terme fora de l'equilibri. L'únic requeriment

per tal de poder utilitzar aquestes relacions per extreure ΔG de les nostres mesures és iniciar el protocol experimental des d'una condició d'equilibri termodinàmic. Les RF poden extendre's en cas que aquesta condició no es satisfaci i el protocol s'iniciï en condicions d'equilibri parcial. Això esdevé un avantatge per tal de mesurar les branques d'energia lliure i l'energia lliure de formació d'estats cinètics, que són aquells estats generats dinàmicament i sovint difícils d'observar i caracteritzar en condicions d'equilibri. Exemples clàssics d'estats cinètics són els estats moleculars intermitjos i els estats mal plegats, que juguen un rol important en molts processos metabòlics.

Els resultats d'aquesta tesi obren la porta a la caracterització termodinàmica i cinètica de processos moleculars complexos que s'esdevenen en condicions d'equilibri parcial (com ocorre en organismes vius) utilitzant l'espectroscòpia dinàmica de forces (és a dir, l'estudi de les forces característiques per induir transicions moleculars) i els teoremes de fluctuació (que proporcionen estimacions de l'energia lliure mitjançant mesures irreversibles). Això permet vèncer una limitació molt habitual de tècniques experimentals com la calorimetria o l'absorbància de rajos UV, que proporcionen mesures mitjades entre una població molt gran de molècules i, per tant, rarament proporcionen informació sobre les configuracions transitòries. Alguns exemples on al força aplicada juga un paper important poden trobar-se en els estats cinètics i metaestables relacionats amb el plegament mecànic –com els estats intermedis i mal plegats–, en la interacció intermolecular, o en estats metaestables que es donen en reaccions de polimerització –com la translocació dels motors moleculars.

Part I

Preliminaries

Chapter 1

General introduction

The advent of the Industrial Revolution, between the late eighteenth and mid-nineteenth centuries, gave birth to thermodynamics as the science that studies motor engines. From the initial rudimentary understanding of thermally equilibrated systems to the theoretical modeling of the Carnot cycle, thermodynamics led to an unprecedented development of the technical capacity of humanity.

The operation of a motor engine is based on the conversion of heat absorbed from the thermal bath into useful work in a cyclic process. It has been experimentally observed that such conversion is never 100% efficient: a fraction of the input thermal energy is always lost as dissipated energy [Zema 97]. This result constitutes the second law of thermodynamics, which states that the work required to bring a macroscopic system from an initial to a final state is equal (for reversible processes) or larger (for irreversible processes) than the difference in free energy between the two states.

However, as intriguing as it may seem, the mesoscopic world is governed by different rules. At such scales, energy fluctuations at the level of individual trajectories can be of the same order of magnitude than the energy exchanged between the system and the thermal bath. As a consequence, independent repetitions of an identical experimental process connecting two states may require different values of the work. Rare trajectories where the work performed is less than the free-energy difference between the initial and the final state may even exist [Jarz 97]. Hence, fluctuations contain relevant information about the behavior of the system under any transformation. In any case, the average of the work values obtained over a large number of trajectories is larger than the difference in free energy, hence satisfying the second law of thermodynamics.

In addition, it has been experimentally observed that molecular machines operating at the mesoscopic scale are much more efficient than macroscopic motors [Yasu 98, Kino 00]. Hence, the understanding of the physical rules governing mesoscopic systems under both equilibrium and non-equilibrium conditions might be crucial to unravel the origins of dissipation. Furthermore, a detailed characterization of several crucial processes in order for life to take place –such as the transcription and translation of DNA, the transport of molecules across biomembranes, or the folding of proteins– is now possible.

This thesis focuses on the study of Brownian fluctuations to extract thermodynamic

information and to characterize the kinetic properties of molecular systems under applied mechanical forces. In this chapter, the framework and context to study such molecular systems is presented. In Section 1.1 a brief introduction to single-molecule techniques, which favored the study of mesoscopic systems from a biophysical perspective, is provided. In Section 1.2 the main ideas of thermodynamics of small systems are sketched. Next, the molecular folding problem is introduced in Section 1.3. Finally, a summary of the work presented in this thesis is given in Section 1.4.

1.1 Single-molecule biophysics

Biophysics can be described as that branch of physics that quantitatively studies living organisms from the atomistic level to entire ecosystems. In particular, molecular biophysics focuses the attention on the study of biopolymers that are crucial for life, such as nucleic acids, proteins, lipids, etc. Some issues addressed by this discipline are the biological function, the structural organization, and the dynamical behavior of molecules.

Technical developments in the field of nanotechnology achieved over the last two decades have open the door to the observation and manipulation of single molecules with unprecedented spatial and temporal resolution. Nowadays, this kind of experiments are called "single-molecule experiments" [Rito 06b, Neum 08].

The most common experimental techniques to manipulate single molecules are optical tweezers (OT), magnetic tweezers (MT) and atomic force microscope (AFM). All three of them exert forces in the range of piconewtons and measure molecular distances with nanometric resolution. The range of energies explored is Energy$\simeq$1 nm$\times$1 pN$\simeq$0.24 $k_{\mathrm{B}}T$ at room temperature (25℃). Hence, thermal fluctuations can be measured using these techniques and play an important role in the study of molecular systems. Differences between AFM, MT and OT appear when considering the operating ranges in force and distance:

- The AFM is based on the principle that a cantilever with a tip can sense the roughness of the surface and deflect by an amount which is proportional to the proximity of the tip to the surface [Zlat 08]. After proper calibration, such deflection can be translated into a force acting on the cantilever. The typical experimental setup in single-molecule experiments is as follows: a surface is coated with the desired molecules and the AFM tip is coated with molecules that can bind (either specifically or non-specifically) to the molecules on the substrate. By moving the tip to the substrate a contact is made between the tip and one of the molecules adsorbed on the substrate. Retraction of the tip at a constant speed allows us to measure the deflection of the cantilever in real time, which provides the force acting on the molecule as a function of its end-to-end distance. The main limitation in the use of AFM in single-molecule experiments is the presence of uncontrolled interactions between the tip and the substrate. Different strategies, like the design of polyproteins, have been specifically developed to overcome this effect [Carr 00].

  The AFM covers forces in the range of 20-1000 pN depending on the stiffness of

the cantilever. Typical values are in the range 10-1000 pN/nm. Resolution in the AFM is limited by thermal fluctuations. When the cantilever stage is held at a constant position, both the extension between tip and substrate and the force acting on the tip fluctuate. Respective root mean square displacements are given by the equipartition law $\sqrt{\langle \delta x^2 \rangle} = \sqrt{k_\mathrm{B}T/k} \sim 2$ Å, and $\sqrt{\langle \delta f^2 \rangle} = \sqrt{k_\mathrm{B}Tk} \sim 20$ pN, where k_B is the Boltzmann constant, T is the absolute temperature of the environment and k is the stiffness of the cantilever, taken equal to 100 pN/nm.

The AFM is an ideal technique to investigate strong intermolecular and intramolecular interactions [Hint 06, Gale 10].

- MT are based on the principle that a magnetized bead with dipole moment $\vec{\mu}$ experiences a force $\vec{f}$ when immersed in a magnetic field gradient $\vec{B}$ equal to $\vec{f} = \vec{\mu}\nabla\vec{B}$. The typical experimental setup is as follows: a bead is trapped in the magnetic field gradient generated by two strong magnets and a molecule is attached between the surface of the magnetic bead and a glass surface. Molecules are pulled or twisted by moving the stage that supports the magnets. A microscope objective with a CCD camera is used to determine the position of the bead relative to the surface, equal to the extension of the molecule x. Force is measured by equipartition using the expression $f = k_\mathrm{B}Tx/\left\langle \delta x^2 \right\rangle$ [Zlat 03, Goss 02].

  Typical values for the stiffness of magnetic traps are 10^{-4} pN/nm, which imply large fluctuations in the extension of the molecule, on the order of 20 nm. The typical range of operating forces is 10^{-2}-10 pN, where the maximum value of the force depends on the size of the magnetic bead. When the position of the magnet stage is kept fixed, the force acting on the bead is kept constant because the spatial region occupied by the bead is small enough for the magnetic field gradient to be considered uniform. Therefore, although the bead position fluctuates the force is always constant.

  MT are extensively used to investigate elastic and torsional properties of DNA molecules and processes involving molecular motors [Mosc 09].

- The principle of OT is based on the optical gradient force generated by a focused beam of light acting on an object with an index of refraction higher than that of the surrounding medium (see Section 2.2 for more details) [Ashk 70, Ashk 86, Svob 94b, Ashk 97, Smit 03]. A typical experimental setup is as follows: a micron-sized polystyrene or silica bead is captured in the optical trap, while another bead is immobilized by air suction in the tip of a micropipette. Molecular biology tools are employed to insert the molecular system under study between the two beads.

  The operating force range in optical tweezers is 0.5-100 pN, depending on the size of the bead (on the order of 1-3 μm) and the power of the laser (few hundred milliWatts). To a very good approximation the trapping potential is harmonic, $f = -kx$, where k is the stiffness of the trap and x is the distance of the trapped bead to the center of the optical trap. Values of the stiffness for OT are 10^{-2}-10^{-4} times smaller than stiffnesses for AFM tips. Consequently, force resolution

is at least 10 times better, it the order of 0.1 pN. Spatial resolution can reach the nanometer level only in carefully isolated environments (absence of air currents, mechanical and acoustic vibrations and temperature oscillations).

OT are being widely used to investigate nucleic acids and molecular motors [Moff 08].

Within the cell, molecular motors drive many dynamical processes involving transient melting events of nucleic acids by applying localized mechanical forces. Therefore, the study of these systems via single-molecule experiments can give insight into the thermodynamics and kinetics of many processes that take place in living systems [Bust 04, Tino 10].

1.2 Thermodynamics of small systems

In traditional bulk experiments, like calorimetry or UV absorbance, the samples under study contain a large number of molecules N, of the order of Avogadro's number ($N_{\mathrm{A}} = 6.02 \times 10^{23}$). Experimental measurements are the result of an average over all the molecules and transient states are very difficult to observe. Fluctuations are of the order of $1/\sqrt{N}$ and become negligible in the large N limit. Thermodynamics and statistical physics are both well-established theories useful to describe such large systems.

In contrast, in single molecule systems N is of order 1 (and therefore $1/\sqrt{N} \sim 1$). These systems provide a paradigmatic example of the so-called *small systems*, where energy fluctuations play an important role since they are of the same order of magnitude than the energy exchanged by the system with the environment [Bust 05, Rito 06a]. Small systems are not restricted to single molecules: any macroscopic system operating at short enough timescales can show large energy fluctuations in the measurement of energy exchanges between the system and the surroundings. On the other hand, if measurements in a single molecule are carried out over long times (compared to the heat diffusion time of the system) fluctuations become small and irrelevant and the molecule no longer behaves as a small system.

In any case, the possibility to not only observe but also manipulate small systems provided by single-molecule experiments has revolutionized the field of statistical physics and has inspired the development of new physical theories to understand non-equilibrium phenomena. Two contributions of great importance for the realization of this thesis are the transition state theory (TST) and fluctuation relations.

Transition state theory

Since the late nineteenth century, the theoretical formulation of chemical reactions has been a subject under development. It has been observed that the conversion of reactants into products requires an energy that represents a maximum along the degree of advance of the chemical reaction. Thus, the transition state (TS) is defined as an intermediate configuration between reactants and products in which there is the activation barrier (*i. e.*, the energy maximum) of the chemical reaction. The existence of this TS between

reactants and products has been formulated, but its identification with an observable state is not straightforward [Zhou 10].

One of the main objectives of TST is to predict how a system explores its accessible configurations, parameterized with a reaction coordinate, just taking into account the energy level of each state. A key idea is to find theoretical expressions for the kinetic rates from the so-called free-energy landscape (FEL) of the system (see Chapter 4).

Single-molecule experiments provide an excellent opportunity to experimentally test several results of the TST, which in turn provides the tools to measure the molecular FEL and related kinetic properties.

Fluctuation relations

Suppose a small system initially at equilibrium that is brought under non-equilibrium conditions to a final state where it equilibrates. Taking standard concepts from statistical physics in can be shown that, if the system satisfies detailed balance, it is possible to extract the free-energy difference between the initial and final states from a sufficiently large collection of non-equilibrium work measurements obtained over independent repetitions of the same protocol [Jarz 11]. The mathematical relations that express this result are known as "fluctuation theorems" or "fluctuation relations" (see Chapter 8).

An remarkable result of the fluctuation relations is that, even though the average of the work measured over the different trajectories must be larger than the free-energy difference, there must exist trajectories where this is not satisfied. That is, occasionally the work performed to bring the system from the initial to the final state is smaller than the difference in free energy between the two states. This result is markedly different from the expected behavior of macroscopic systems, where the free-energy difference is always a lower bound to the work required.

Again, single-molecule experiments provide a powerful setup to experimentally test fluctuation relations. In turn, these relations facilitate the measurement of the molecular free energy of formation of native structures and kinetic intermediates that are difficult to sample under equilibrium conditions, or even intermolecular binding free energies.

1.3 The molecular folding problem

In order for a biopolymer (a protein or nucleic acid) to perform its specific biological task, it needs to be in a given conformation. The understanding of how a given polymer chain folds towards this specific state is one of the main challenges of molecular biophysicists.

The problem is specially intriguing for the case of proteins. In 1969, Cyrus Levinthal discussed that a random search of the proper conformation among all the accessible states, which for a 150-aminoacids long polypeptide are of the order of $\sim 10^{80}$, would take more time than the age of the Universe [Levi 69]. This result is known as the Levinthal paradox, to which the author suggested the existence of a specific folding pathway as a solution. By quoting his own words: "protein folding is speeded and

guided by the rapid formation of local interactions which then determine the further folding of the polypeptide."

A few decades after, the concept of the FEL was introduced in the study of the protein folding problem [Andr 94, Karp 97]. In this context, the existence of a specific folding pathway is not required for the Levinthal paradox to be solved. Now, it is well-accepted that both thermodynamic and kinetic factors influence the search of the functional molecular configuration along the free-energy surface [Dinn 00]. What still remains unsolved is how to determine the FEL of a biomolecule given its sequence.

The molecular folding problem has many other faces. For example, it is known that some proteins fold as they are being synthesized in the ribosome; others need to go to the cytoplasm or other cell organelles; in some cases the help of a molecular chaperon is required for the polypeptide to find its functional state. In addition, although it seems straightforward to think that the functional configuration is the most thermodynamically stable one (usually referred to as the native conformation), this is not a general result. In some occasions, intermediate or even unfolded conformations that appear under applied forces are the functionally relevant. An example is found in DNA, that must be unfolded during transcription and translation to release genetic information. In the case of riboswitches, different functions are performed depending on the structure of the molecule. Sometimes a biomolecule can be trapped in a misfolded conformation that prevents it from performing its assigned task. This particular behavior lies at the root of many diseases, such as Parkinson, Alzheimer, the Creutzfeldt-Jakob disorder, and even some cancers. How to include all these effects into the predicted FEL is still work in progress.

Single-molecule experiments provide a novel platform to visualize unfolding and folding individual trajectories of biomolecules *in vitro*. They make it possible to study the relative stability and kinetic accessibility of native, intermediate and misfolded states under different conditions. Due to their structural simplicity, nucleic acids are model systems that allow us to test and enlarge the current knowledge on the modeling of the molecular FEL. The situation becomes more complex when dealing with proteins. Nevertheless, the modeling of proteins and nucleic acids shares several assumptions and thus single-molecule experiments provide a unique opportunity to experimentally test and characterize several complex processes occurring at the molecular level involving proteins, nucleic acids, and molecular ligands.

1.4 Overview of the thesis

In this thesis, non-equilibrium single-molecule methods are employed to investigate and extract accurate information about the thermodynamic, kinetic and elastic properties of a wide variety of molecular systems. Results are presented in four different parts.

Part I provides a description of the research field and the main ingredients that have been used to perform the experimental work. In Chapter 1 an introduction to the field of single-molecule biophysics and the thermodynamics of small systems is presented.

An overview of the molecular folding problem is also provided. Chapter 2 provides a brief biological explanation of the different biomolecules used, such as nucleic acids and proteins, is presented. Next, the discovery and the physical principles of optical trapping and the mode of operation of the optical tweezers instrument are summarized. Finally, the main features of single-molecule experiments carried out with optical tweezers are introduced.

Part II deals with the use of dynamic force spectroscopy (DFS) techniques to characterize different molecular systems, such as DNA and RNA hairpins, protein Barnase and the antibody-antigen interaction. This part includes Chapters 3 to 7.

In Chapter 3 it is shown how to extract the elastic properties of short single-stranded DNA (ssDNA) molecules from pulling experiments by using two-state DNA hairpins. The use of DNA hairpins, whose secondary structure is very well controlled under pulling experiments, allows us to obtain such information without the direct measurement of the force-dependent molecular end-to-end distance. To this end, a toy model of the experimental setup is presented and two different methods to extract the ideal elastic response of ssDNA are shown. An important novelty of the method hereafter presented is that it allows to measure the elastic properties of short ssDNA in a range of forces where it is known that long ssDNA molecules tend to form secondary structure. Parameters derived in this chapter are very important along this thesis because most results are obtained by using short DNA hairpins and the elastic response of ssDNA is important for many calculations.

In Chapter 4 the concept of the molecular FEL is introduced, and its theoretical evaluation for nucleic acid hairpins is explained in detail. In addition, a brief reminder of the most important results of TST is provided. A model DNA hairpin is used to introduce DFS analysis, where valuable information about the molecular FEL is obtained from experiments carried out under non-equilibrium conditions. Results are then compared with theoretical models in order to extract information about the molecular free energy of formation and the attempt frequency at zero force.

Chapter 5 presents several situations where DFS is key to characterize a wide variety of molecular systems involving the presence of nucleic acid hairpins. Thus, different two-state DNA hairpins are used to understand how changes in the position of the kinetic barrier along the reaction coordinate do affect the kinetics of the system. The importance of the Leffler-Hammond postulate on the position of the transition state will be shown. The non-specific binding of monovalent and divalent ions in RNA is also investigated by studying the effects of the ionic strength on the elastic and thermodynamic properties of an RNA hairpin that is mechanically unfolded and folded. Next, DFS tools are pushed beyond their limits to gain accurate information about kinetic states, which are states observed mostly under non-equilibrium conditions. Hence, molecular intermediate and binding states are deeply characterized.

In Chapter 6 the maturation line of the immune system is qualitatively investigated through the characterization of the affinity and flexibility of the antigen-antibody bonds. Results are used to understand how antibodies evolve in order to neutralize a foreign

infectious body. Finally, the profile of the FEL of the antibody-antigen interaction is discussed in terms of the bond affinity.

Finally, in Chapter 7 DFS techniques will be used to unravel the behavior of a small protein –barnase– when pulled by mechanical forces. Elastic, kinetic and thermodynamic properties will be addressed from both equilibrium and non-equilibrium experiments.

In Part III, fluctuations theorems are employed to measure thermodynamic properties of kinetic states, either misfolded or intermediate. This part includes Chapters 8 and 9.

In Chapter 8 fluctuation theorems are pedagogically explained and their application to obtain the free energy of formation of molecules is shown is some detail using a model DNA hairpin. Limitations due to hysteresis effects are also discussed and an example of the situation is provided by using protein Barnase.

Finally, Chapter 9 presents an extension of fluctuation theorems that allows to extract thermodynamic information of kinetic states, such as molecular intermediate and misfolded states. It will be shown how to study the thermodynamic stability of different states and the respective free-energy branches under non-equilibrium pulling experiments and how, in some occasions, a reversible behavior is not necessarily related to full thermodynamic equilibrium.

To finish, in Part IV the main conclusions derived along this work are summarized and future lines of research are presented.

Chapter 2

Description of the experimental setup

In this thesis, experiments in single molecules, such as proteins and nucleic acids, are performed using optical tweezers. The scale and resolution involved in such systems provide an excellent platform to test and develop recent results from non-equilibrium statistical physics and transition state theory. Furthermore, the study of processes occurring at the molecular scale, such as the molecular folding problem or the intermolecular binding, becomes possible. Single-molecule experiments pave the way to the understanding of the structure and kinetics of biomolecules under different conditions.

This chapter presents an introductory description of the elements that constitute the experimental setup used throughout this thesis. These are divided into two main parts:

i. Chemical and structural properties of nucleic acids (DNA & RNA) and proteins are summarized.

ii. Optical tweezers, which is the experimental technique used to unravel the thermodynamic and kinetic properties of single molecules, are briefly explained.

To end, the performance of single-molecule experiments with OT is explained, from how to calibrate force to the operation of the instrument and the most common experimental protocols

2.1 Biomolecules

Biomolecules are organic compounds essential for the correct operation of cells at different levels. To mention a few, nucleic acids are important to store, copy and transmit genetic information; proteins are crucial in many catalytic reactions; carbohydrates are responsible for energy storage; lipids are the main bricks of cellular membranes [Goda 96, Albe 00]. There is not a one-to-one relation between the type of biomolecule and the biological function. In fact, biomolecules usually associate forming macro-

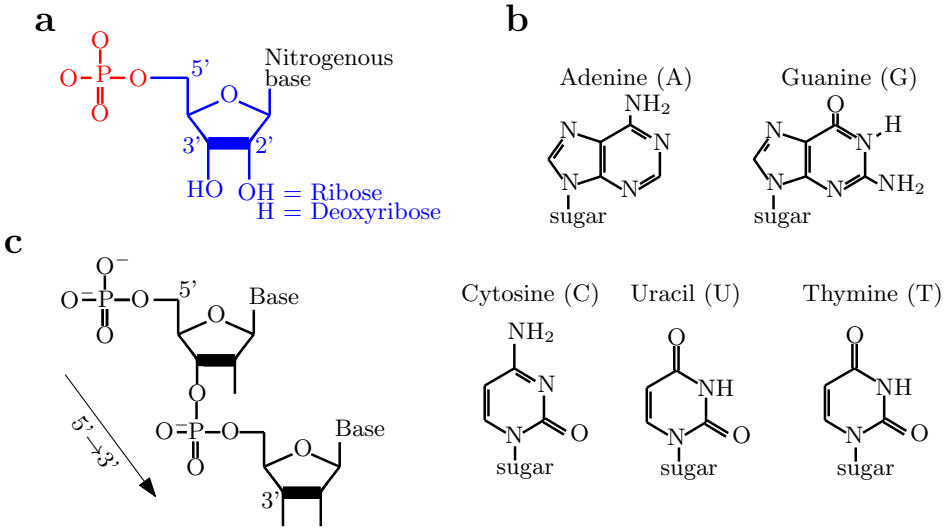

Figure 2.1: Building blocks of nucleic acids. a. Chemical structure of a nucleotide, which consist of a phosphate group (red), a pentose sugar (blue) and a nitrogenous base (black). The radical in the 2'-carbon of the sugar determines whether the sugar is a deoxyribose or a ribose. **b.** Chemical structure of the different nitrogenous bases (A, C, T, G and U). **c.** Phosphate bond between two consecutive nucleotides.

molecules –such as lipoproteins, glycolipids, and so forth– and perform a wide range of functions.

In what follows, a brief chemical and structural description of the biomolecules investigated in this thesis is presented.

2.1.1 Nucleic acids

Nucleic acids are essential molecules for any form of life as they carry the genetic information which is transmitted across generations. They are polymeric macromolecules made of nucleotides, which are organic compounds that consist of a phosphate group, a pentose sugar and a nitrogenous base (Fig. 2.1a). The sugar can be either deoxyribose or ribose. This determines whether the nucleic acid is deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), respectively. The nitrogenous bases adenine (A), cytosine (C) and guanine (G) are found in both DNA and RNA, while thymine (T) only occurs in DNA and uracil (U) in RNA (Fig. 2.1b). The different nucleotides are usually given the same name as their corresponding nitrogenous base. In both DNA and RNA the nucleotides in the polymer chain are linked together through the phosphate bonds that take place between the position 5' of one nucleotide and position 3' of the following (Fig. 2.1c). The polarity of the polynucleotide chain is defined with the direction of the phosphate bond between sequential nucleotides (5'–3' o 3'–5'). For convention, the sequence of any nucleic acid molecule is given in the 5'–3' direction.

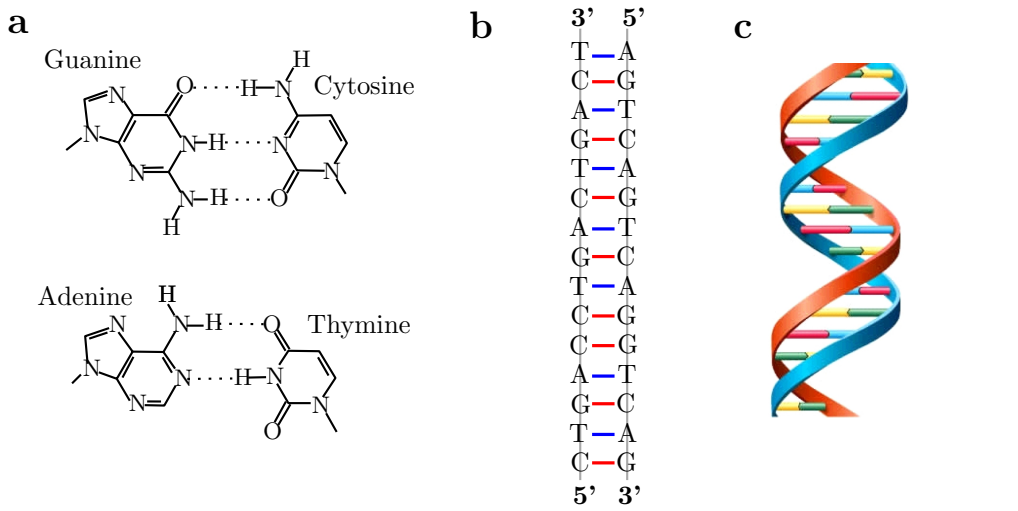

Figure 2.2: Structure of DNA. a. Canonical Watson-Crick base-pairing between C≡G and A=T. **b.** Two complementary polynucleotide chains that run in opposite directions (5'–3' and 3'–5') and that form a DNA molecule. **c.** Sketch of the 3-dimensional structure of the DNA double-helix.

DNA

DNA is mainly found in the cell nucleus and in the mitochondria. Combined with proteins, it forms chromatine and chromosomes. The main function of DNA is to preserve, transmit and reproduce from generation to generation to required information to synthesize the proteins of a living organism.

In 1953, James Watson and Francis Crick, thanks to the crucial experimental work carried out by Maurice Wilkins and Rosalind Franklin, discovered the secondary structure of DNA [Wats 53, Wilk 53, Fran 53]. The main properties are summarized as follows:

i. DNA is made of two polynucleotide chains attached to each other through hydrogen bonds that take place between the nitrogenous bases in each chain. These hydrogen bonds follow very specific rules: A interacts with T through two hydrogen bonds (A=T), and G with C through three hydrogen bonds (C≡G, Fig. 2.2a). The interaction between two different nitrogenous bases is refereed to as a "base pair" (bp), and the specific A=T and C≡G bps are given the name of canonical Watson-Crick base pairs.

ii. The two chains are complementary (Fig. 2.2b): it is possible to reconstruct the sequence of nucleotides of the second chain by knowing the content of the first one thanks to the specificity of canonical Watson-Crick base pairs and the one-to-one relation between nucleotides in each chain.

iii. The two chains run in opposite directions (5' to 3' and 3' to 5', respectively), and form a double-helix structure as depicted in Fig. 2.2c.

Different types of the double-helices are found in nature. The most common is the B-form, which consists of a right-handed double helix with a constant diameter of 2 nm that has 10 bps per turn and where the distance between sequential bps is 3.4 Å. In solutions with alcohol or high salt concentration the A-form (right-handed double helix with 11 bps/turn and 2.1 Å/bp) and the Z-form (left-handed double helix with 12 bps/turn and 3.8 Å/bp) have also been observed.

RNA

RNA is made of a single chain of ribo-nucleotides. Molecules are typically shorter than DNA as they are the result of the transcription of one or few gens (a gen is a segment of DNA that codifies for one protein). There are three types of RNA molecules: the messenger (mRNA), the ribosomal (rRNA) and the transfer RNA (tRNA). The mRNA carries the genetic information from the cell nucleus to the cytoplasm, where the synthesis of the protein takes place. The rRNA is found in association to the ribosome, which is the molecular machine that reads mRNA and synthesizes the coded protein. The tRNA supplies the ribosome with the required aminoacids in order to perform the protein synthesis.

Hydrogen bond interactions can occur between the nucleotides along the polymeric chain. The most common are the canonical Watson-Crick base pairs, where T is replaced by U. However, non-canonical base pairs, known as Wobble base pairs, have also been observed between G−A and G−U. The formation of Watson-Crick and Wobble base pairs gives rise to local secondary structures, summarized in Fig. 2.3.

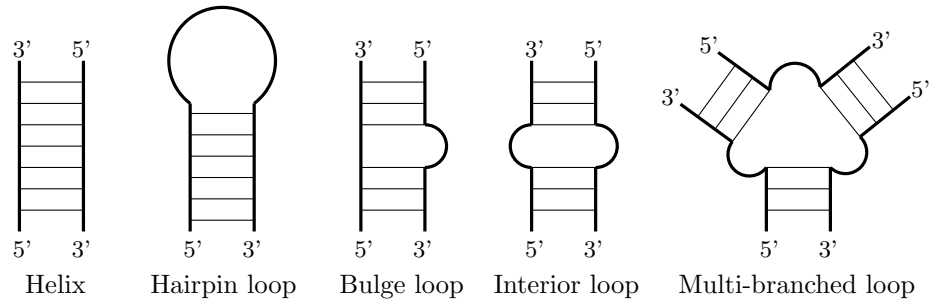

Figure 2.3: RNA secondary structure. Watson-Crick base-pairing interactions between the nitrogenous bases found in RNA promote the local formation of double helices, hairpin loops, bulge loops, interior loops and multi-branched loops.

2.1.2 Proteins

The word "protein" derives from the Greek work $\pi\rho\omega\tau\varepsilon\iota o\varsigma$ (*proteios*), which means "primary" or "of first order". That is because proteins are the most important biological substance after water.

Proteins are key biological elements because they carry out a wide spectrum of functions. For instance, they are important constituents of the cytoskeleton, membranes and

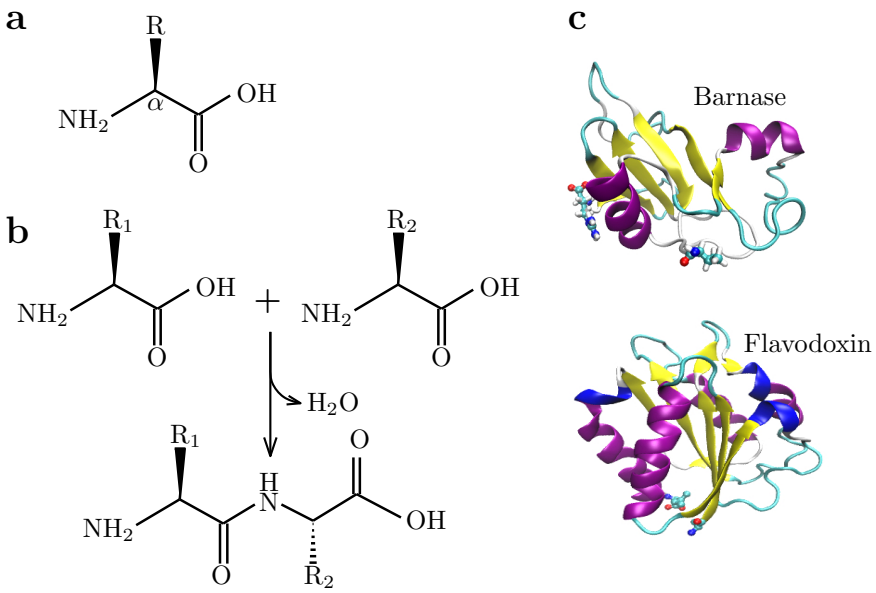

Figure 2.4: Aminoacids and proteins. a. Chemical structure of an aminoacid. **b.** Schematic representation of the formation of a peptide bond between two aminoacids, which implies the releasing of one water molecule. **c.** 3-dimensional representation of the tertiary structure of Barnase and Flavodoxin, two different globular proteins. α-helices are colored in purple, β-sheets in yellow and random coils in cyan. The structure of the first and last aminoacids of the protein chain are also plotted.

tissues. They also play an essential role in metabolic reactions and catalyze and regulate several processes that take place in living organisms. Moreover, they transport different kinds of molecules inside and outside cells and also are the responsible of cell motility. Each living organism has its own proteins. This is particularly important in the immune system, which will be discussed below.

Aminoacids are the building blocks of proteins. They consist of a carbon atom, denoted by C_α, surrounded by an amino and a carboxyl group, and a side chain R (Fig. 2.4a). In nature there are 20 different kinds of aminoacids, and they differ from each other only in the residue chain. Aminoacids are linked together through the peptide bond, which is a covalent bond that takes place between the amino group of one aminoacid and the carboxyl group of another one, with the consequent releasing of a water molecule (Fig. 2.4b). A peptide is the covalent union of a few tens of aminoacids. When the number of aminoacids in a peptidic chain is larger than 50, it is usually referred to as a protein.

The linear sequence of aminoacids is known as the primary structure of proteins. The secondary structure is the spatial organization of the aminoacids along the peptidic chain. The most abundant secondary structures in proteins are the α-helices, the β-sheets, or the random coils. The α-helix structure consists of a right-handed spiral where the carbonyl C=O group of each aminoacid forms an hydrogen bond with the amide N-H group of an aminoacid four residues further. In the β-sheet structure, segments of the aminoacid chain are connected laterally through hydrogen bonds that take place between

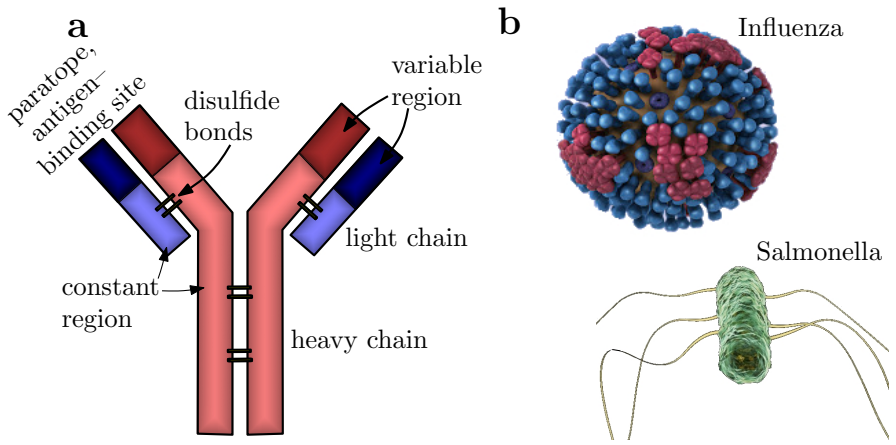

Figure 2.5: Antibody and antigens. a. Antibodies are Y-shaped proteins with a constant region and two variable regions with hyper-mutability that are known as the antigen-binding sites or the paratopes. **b.** Antigens are foreign infectious bodies. In this image a representation of the influenza virus and the salmonella bacteria can be seen.

the C=O and the N-H groups of different confronted aminoacids and, as a result, twisted sheets are formed. The random coil appears when stable structures are not possible and the aminoacid chain collapses randomly. The tertiary structure of a protein refers to the full molecular 3-dimensional structure, and is determined by interactions taking place between aminoacids along the whole peptidic chain. Because secondary structures are local, a single protein can have several α-helices, β-sheets or random coils (Fig. 2.4c). According to the tertiary structure, we distinguish between globular proteins (which are spherical and soluble) and fibrous proteins (which are elongated and insoluble).

2.1.3 Antibodies and antigens

The immune system of vertebrate organisms is a complex network of organs, cells and proteins distributed throughout the body which regulates the growth and development of the organism and protects it from diseases [Poco 02].

Antibodies, also known as immuno-globulins, are proteins that play a very important role in the immune system, as they identify and neutralize foreign objects. All antibodies have a similar chemical structure: they consist of two identical heavy chains and two identical light chains that are linked together through disulfide bonds and that are arranged in a "Y" shape. The two terminal extremes of these Y-shaped proteins are variable regions, usually referred to as paratopes, that constitute the antibody-binding site (Fig. 2.5a).

Antigens are any foreign agents that parasitize a living organism and promote disease or infection. Bacteria, viruses or fungus are the most famous examples of antigens (Fig. 2.5b). In practical terms, there is an immense number of antigens, each one with its characteristic shape and chemical structure.

The innate immune system of a living organism does not have all the information

required to specifically bind and neutralize the infinite amount of external bodies that might be infectious. Lymphatic B cells, which are characteristic cells of the immune system, identify the presence of antigens and trigger a cascade of mutations occurring in the paratope region of antibody in order to improve specificity to each part of the foreign body. This adaptative response is known as the maturation process of the immune system. At the end of the maturation line, new specific antibodies emerge that are able to bind and neutralize the antigen.

2.2 Optical tweezers

The study of light phenomena is very old and is one of the main motors for scientific development. Mirrors were already used in the ancient Egypt. Greek and Roman philosophers –such as Pythagoras, Democritus, Plato, Aristotle or Seneca– did the first attempts to rationally describe light propagation and used crystals as magnifiers and lighters. In the XIII century Roger Bacon (1215–1294) realized that lenses can be used for correcting human vision and 300 years later the first telescopes and microscopes were constructed. In 1619 Johannes Kepler (1571–1630) introduced the concept of pressure of radiation to explain why the tail of a comet points away from the Sun in his work entitled *De Cometis Libelli Tres*.

The appearance of such new optical devices as telescopes and microscopes boosted the observation of several light phenomenon and forced scientists to deeply address the question of "what is light?". Robert Hooke (1635–1703) and Christiaan Huygens (1629–1695), among others, tried to explain reflection, refraction and diffraction by considering that light behave as a wave. On the other hand, Isaac Newton (1642–1727) modelized light as a collection of particles.

This debate took place for a century, being the corpuscular theory of light more accepted than the undulatory theory because of Newton's good scientific reputation. However, when James C. Maxwell (1831–1879) wrote down the mathematical equations that summarized electromagnetism, the theory that light was an electromagnetic wave gained followers despite the fact that it could not explain the black body radiation.

During the first half of the XX century, with the advent of quantum mechanics, it was formulated that undulatory and corpuscular properties could not be separated at the submicroscopic scale. Planck, Einstein, Bohr, Born, Heisenberg, Schrödinger, De Broglie, Pauli, Dirac, among others (Fig. 2.6), stated that light is made of small energy packages called *photons* that simultaneously behave as wave and particles and that have momentum. This theory is nowadays still the most consolidated.

In 1960 the first laser (acronym for Light Amplification by Stimulated Emission of Radiation) was built. This new tool has been very useful to explore the nature of light and also has allowed the discovery of new optical phenomena. Moreover, the use of lasers in most scientific disciplines enables not only the observation of the microscopic world in high detail, but also its manipulation at our will.

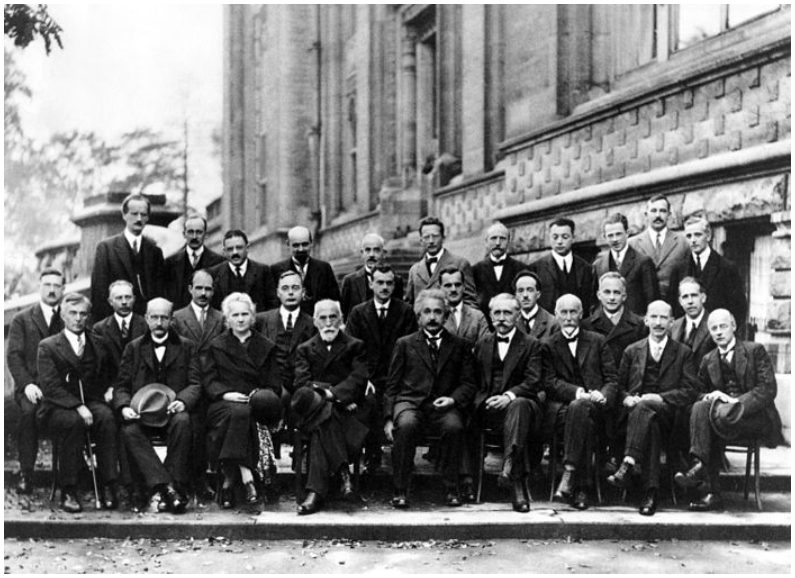

Figure 2.6: Fifth Solvay conference in Physics, 1927, *Electrons et photons***.** Assistants were, from left to right and from top to bottom: A. Piccard, E. Henriot, P. Ehrenfest, E. Herzen, Th. de Donder, E. Schrödinger, J. E. Verschaffelt, W. Pauli, W. Heisenberg, R. H. Fowler, L. Brillouin; P. Debye, M. Knudsen, W. L. Bragg, H. A. Kramers, P. A. M. Dirac, A. H. Compton, L. de Broglie, M. Born, N. Bohr; I. Langmuir, M. Planck, M. Sklodowska-Curie, H. A. Lorentz, A. Einstein, P. Langevin, Ch. - E. Guye, C. T. R. Wilson, O. W. Richardson

2.2.1 Principles of optical trapping

The radiation pressure is defined as the pressure exerted by the electromagnetic radiation to any surface. In 1970 Arthur Ashkin experimentally demonstrated that the pressure of radiation from a visible laser light can be used to accelerate micron-sized particles suspended in liquids or gases [Ashk 70].

In a first experiment, transparent latex spheres of 2.68 μm diameter suspended in a water medium were irradiated with a laser of $\lambda = 514.5$ nm wavelength at a power of $P = 1$ mW (Fig. 2.7a). The index of refraction of the micron-sized particles is $n_{\text{sphere}} = 1.59$, whereas for the water it is $n_{\text{medium}} = 1.33$. Ashkin observed that the illuminated off-center spheres were simultaneously drawn into the beam axis and accelerated in the direction of light. This behavior can be explained using the physical law of conservation of momentum and ray optics.

Light has momentum. Therefore, we can consider that each ray of light has a linear momentum $\vec{p}$ that can be written as:

$$\vec{p} = n_\lambda \frac{h}{\lambda} \hat{e}_p, \tag{2.1}$$

where h is the Planck constant, λ is the wavelength of the laser, $\hat{e}_p$ is the unitary vector that indicates the direction of propagation of the ray, and n_λ is the number of photons of the ray, also denoted as the flux of momentum.

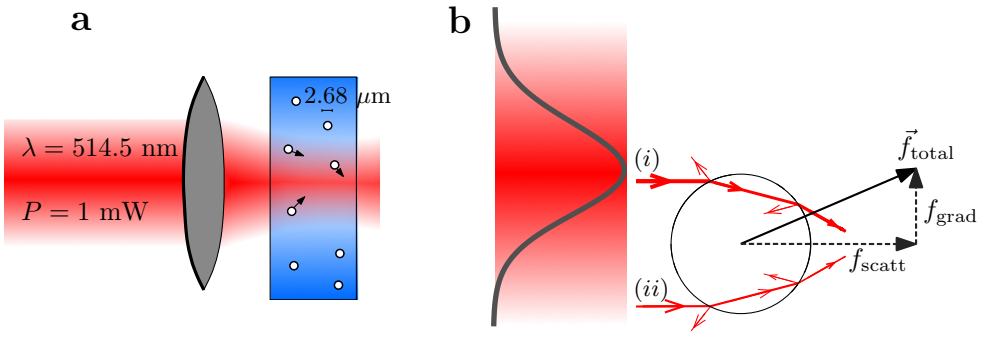

Figure 2.7: Experimental observation of the pressure of radiation. a. Experimental setup used by A. Ashkin in 1970 to experimentally observe the effect of the pressure of radiation on micron-sized beads [Ashk 70]: the illuminated spheres were accelerated towards the center of the beam axis and towards the direction of the incident light. **b.** Ray diagram scheme of the interaction between the bead and the light from a Gaussian laser beam.

Given one ray of light coming from the laser beam, when it reaches the sphere a small fraction of it reflects and the rest refracts according to the Snell law. The refracted ray of light linearly propagates inside the microsphere until it reaches its other facing surface, where reflection and refraction occur again (ray diagram represented in Fig. 2.7b). Therefore, according to Eq. (2.1), the ray of light suffers a change in its linear momentum. There are two reasons for this: the first is the deflection of the ray, *i. e.*, the change in the direction of propagation $\hat{e}_p$ of the incoming and the exiting light to the bead; the second is the loss of photons n_λ of the ray each time it is reflected in the bead surface. Because the momentum of the whole system (bead plus ray of light) must be conserved, the loss of momentum of the ray of light is equal to the momentum gained by the bead. To sum up, each incoming ray of light transfers a certain quantity of momentum to the bead.

If we consider that all the rays of light from the laser beam have the same intensity, the addition of the change in momentum of all the incoming rays to the bead would generate a net force in the direction of the incident light, called "scattering force", f_{scatt}. However, light from a laser beam has a Gaussian profile (Fig. 2.7b), which means that the light coming from the center of the beam has more intensity (larger n_λ) than the light in the extremes. This results in an extra component for the change in momentum of the bead that points toward the center of the laser beam. The force generated due to such contribution is known as the "gradient force", f_{grad}, and is proportional to the gradient of the intensity.

Ashkin repeated the same experiment using other spheres, with diameters of 1.31 and 0.59 μm. He found that the smaller the particle, the more power required to obtain similar forces acting upon the beads [Ashk 70, Ashk 86]. Geometrical optics does not describe optical trapping correctly in the regime where the diameter of the object is similar to the wavelength of light (here equal to $\lambda = 0.5145$ μm), but it still provides a qualitative understanding of the phenomenon. He also proved that the reversion of the indexes of refraction of the bead and medium (*i.e.*, $n_{\text{medium}} > n_{\text{sphere}}$) made the

microsphere to be pushed out of the center of the beam.

2.2.2 First optical traps

From the proper combination of scattering and gradient forces a wide variety of optical traps, based only on radiation pressure, can be designed. In a second experiment, Ashkin did construct an optical potential by using two identical but opposing laser beams [Ashk 70], as sketched in Fig. 2.8a. The scattering forces caused by the change in momentum of each laser beam cancel each other and the only remaining force is the addition of the two gradient forces, giving rise to the aforementioned potential well centered in the optical axis that can successfully trap micron-sized particles. In another example [Ashk 86], Ashkin and collaborators used a high numerical aperture lens that strongly focuses light from a 514.5 nm-wavelength laser (Fig. 2.8b). It was shown that an optical trap close to the focus of the lens was generated, being able to capture particles of diameters ranging from 10 μm to 25 nm. For microspheres whose diameter is larger than the wavelength of the laser –the so-called Mie size regime– ray optics can be used to explain the phenomena. In contrast, when the size of the particles is smaller than the laser wavelength –the so-called Rayleigh size regime–, quantum optics must be used to explain the ability of light to trap beads. A detailed physical explanation of the two regimes can be found in [Kerk 69, Smit 82, Hugu 10a].

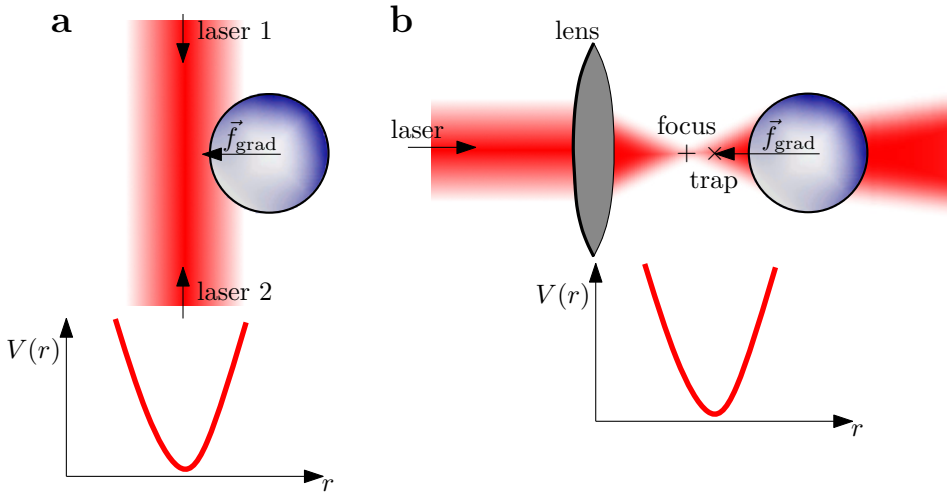

Figure 2.8: Optical trapping by a laser beam and formation of an optical well potential. a. In a two identical counter-propagating laser beams setup the scattering forces acting on the bead coming from the two lasers cancel each other and the two gradient forces are added. As a consequence the resulting force acting on the bead points toward the center of the beam of both lasers [Ashk 70]. **b.** A laser beam is focused with a high-numerical aperture lens and a force towards a position (indicated with $\times$) close to the focus (indicated with $+$) is exerted on the bead [Ashk 86]. For both setups (panels a and b) a schematic representation of the optical potential well $V(r)$ is provided, being r the distance between the center of the bead and the center of the laser beam (a) or the position of the trap close to the focus (b), respectively.

The invention represented in Fig. 2.8b is given the name of "optical tweezers" and

triggered a revolution in the fields of nanotechnology and molecular biophysics [Svob 94a, Moff 08]. This is because the range of forces that a focused laser beam induces in a micron-sized bead is in the order of picoNewtons (10^{-12} N) and instruments have been designed with the capacity of measuring nanometric displacements [Simm 96, Viss 96, Neum 04]. Both picoNewtons and nanometers are in the order of magnitude of biological phenomena occurring at the molecular scale. For instance, forces required to unzip DNA or to unbind intermolecular connections such as antibody-antigen interaction or DNA-binding peptide are typically around 15–30 pN. On the other hand, the energy of formation of hydrogen bonds or Van der Waals interactions cover the interval of 8–50 pN·nm, and thermal fluctuations at 25℃ are of the order of $k_{\mathrm{B}}T \sim 4.11$ pN·nm (where k_{B} is the Boltzmann constant). Both quantities coincide with the range of energies accessible using optical tweezers, which are of the order of force×distance $\sim 10^{-12} \times 10^{-9}$ N·m ~ 1 pN·nm.

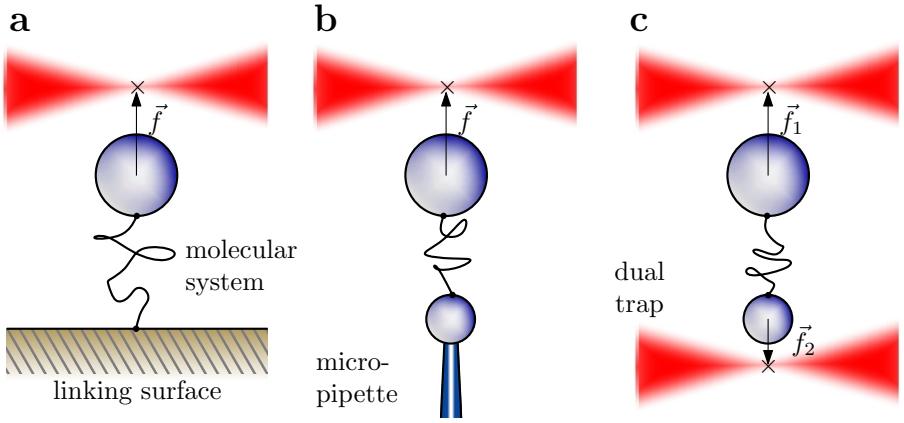

Figure 2.9: Example of setups that use optical tweezers. **a.** *Single trap & linking surface:* the molecular system under investigation is attached between a bead, which is captured with the optical trap, and a linking surface. By changing the distance between the surface and the bead forces will be exerted on the molecular system. **b.** *Single trap & micro-pipette:* the molecular system is attached between two micron-sized beads. One bead is immobilized in the tip of the micro-pipette by air suction whereas the other is captured in the optical trap. Forces can be exerted on the molecular system by changing the distance between the two beads. **c.** *Dual trap setups:* the molecular system is attached between two beads, and each bead is captured in an optical trap. Again, forces can be exerted on the molecular system by changing the distance between the two traps.

Imaginative setups involving the use of optical tweezers have been designed to characterize, from a mechanical point of view, ligand-receptor binding, molecular unfolding, stall forces in molecular motors, and so on. The key idea is to attach one extreme of the molecular system under investigation to a micron-sized bead that is captured in the optical trap, and to immobilize the other extreme. The position of the bead relative to the center of the optical trap gives a measurement of the force acting of the object. In Fig. 2.9 three different examples are shown, where a linking surface [Simm 93, Wang 97], a micro-pipette [Smit 96, Baum 97], or a second bead captured in a second optical trap [Wood 06a, Stig 11] (panels a, b and c respectively) are used. For the sake of simplicity,

in Fig. 2.9 and next figures the position of the trap and the focus of the laser are represented as if they were the same, although it has been mentioned above that this is not the case. This approximation does not have any impact on the results presented throughout this work. Once the molecular system under study is kept under control, different experimental protocols can be applied to investigate its response to applied mechanical forces (Section 2.3.2).

2.2.3 The mini-tweezers setup

The instrument used in the course of this thesis is known as "mini-tweezers", and is based on a miniaturized version of a previous design by Steve B. Smith and Carlos Bustamante in 2003 [Smit 03, Bust 06, Smit 06, Smit 13], where they used two counter-propagating laser beams which are focused on the same spot. The mini-tweezers setup is inspired by the sketch represented in Fig. 2.9b [Hugu 10b, Hugu 10a]: the molecular system is inserted between two polystyrene beads, and one bead is immobilized in the tip of the micro-pipette by air suction whereas the other is captured in an optical trap.

This mini-tweezers have been proved to be very useful to extract thermodynamic properties of biomolecules [Coll 05, Shan 10, Hugu 10b], to study the mechanical folding and unfolding of both nucleic acids and proteins under equilibrium and non-equilibrium conditions [Kell 97, Cui 00, Liph 01, Mano 06], and to investigate the behavior of molecular motors under force [Wuit 00, Smit 01].

In what follows, the mini-tweezers setup is briefly described.

The instrument

The main features of the mini-tweezers are [Hugu 10b, Hugu 10a]:

i. The optical trap is made of two identical counter-propagating laser beams of 845 nm wavelength that are focused on the same point. One of the main advantages of this setup is that the scattering forces of the two lasers cancel each other and only gradient forces are responsible for capturing objects in the focus point of the two lasers.

ii. The lasers have a low numerical aperture but are focused using a high numerical aperture objective, which facilitates the collection of all the light deflected by the trapped object. In fact, the instrument is able to collect more than the 95% of the light exiting the captured bead. This makes it possible to directly measure the change in momentum of the incident light relatively to the exiting light, and to relate it with the force acting on the trapped object.

iii. Finally, this instrument has a small and portable size, hence the name.

The design of the instrument is sketched in Fig. 2.10. In what follows, the path followed by one laser will be described. The second laser follows exactly the same path but in opposite direction.

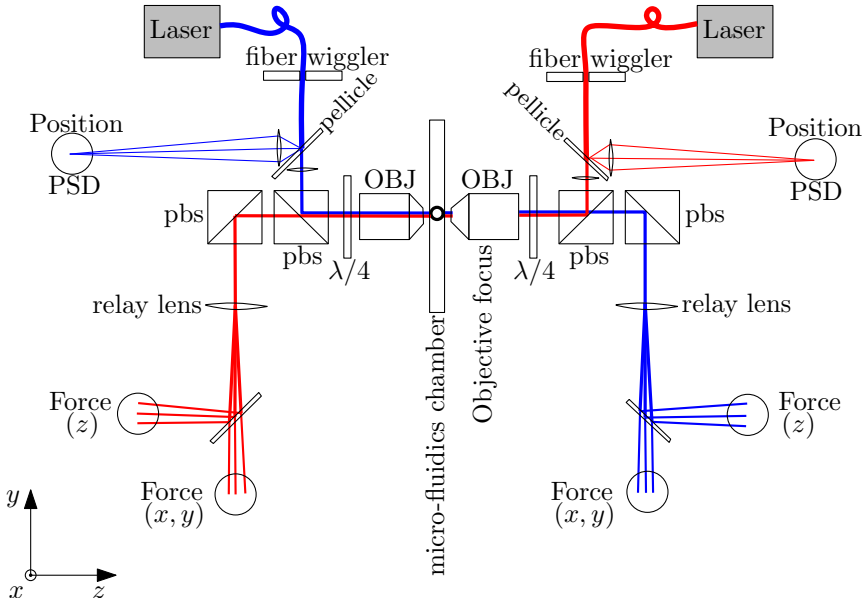

Figure 2.10: Scheme of the dual-beamed mini-tweezers. The path of each laser is indicated in red and blue colors.

The light emitted by the laser is transmitted through a single mode optical fiber. The position of the optical fiber is controlled by a fiber wiggler. When the light comes out of the optical fiber it finds a pellicle, where 5% of light is reflected and directed towards a position-photo sensitive (PSD) detector that provides information about the position of the trap. The remaining 95% of the light first finds a collimating lens that turns it into a collimated beam of light. Such beam is next completely reflected by a polarizing beam splitter (pbs) and immediately afterwards it finds a quarter-wave ($\lambda/4$) plate, which converts linearly polarized light into circularly polarized light. The light is then focused by a microscope objective lens (OBJ) to a focal point inside the micro-fluidics chamber. Light coming out, which is still circularly polarized, is collected by a second objective lens and converted into linearly polarized light through a second $\lambda/4$ plate. This linearly polarized light pass through a first pbs and it is entirely transmitted to a second pbs that changes the direction of propagation of the light-beam. Finally, light finds a relay lens that redirects it into a force-PSD that measures the x and y components of force performed by the optical tweezers. In order to determine the force along the z axis, a beam splitter is used that redirects part of the light exiting the relay lens and allows to measure the diameter of the trapped bead or, equivalently, to size of the light cone coming out of the optical trap.

The micro-fluidics chamber

Experiments are carried out in the micro-fluidics chamber that is located between the two objectives (Fig. 2.10). Its dimensions are 6 cm long, 2.6 cm height and 2 mm width.

The chamber is made of three parallel fluidics channels, as sketched in Fig. 2.11. Buffer and a solution of beads are flowed through on end of the central and the lateral channels, respectively, using syringes. The other extremes are connected to a waste bottle. The two lateral channels are connected with the central one through dispensing tubes, so part of the beads flowed through the lateral channels will reach the central one. Finally, the chamber also contains a micro-pipette inside the central channel through which are is flown using a syringe. The lasers are focused inside the central channel of the chamber, far enough from the surfaces in order to avoid hydrodynamic effects on the trapped objects.

The initial steps of any single-molecule experiment using the mini-tweezers are performed as follows:

i. Buffer is flowed through the three channels. The buffer depends on the experiment. For example, Tris-HCl-EDTA buffer with variable ionic conditions is used to study DNA hairpins, whereas PBS is used to study antibody-antigen interactions.

ii. Beads of type 1 are flowed through one of the lateral channels (the upper one in Fig. 2.11). These beads travel along the channel and some of them end up at the central channel by going through the dispensing tube. Using the optical trap, one of such beads in the central channel is captured and brought to the tip of the micro-pipette, where it is immobilized by air suction.

iii. Beads of type 2 are flowed through the other lateral channel (the one in the bottom in Fig. 2.11). As before, these beads travel along the channel and some of them end up at the central channel traveling through the corresponding dispensing tube. Using the optical trap, one of such beads in the central channel is captured and brought close to the micro-pipette.

The experiment is performed close to the micro-pipette, in the red-circled region in Fig. 2.11. The beads are brought close together by moving either the chamber or the optical trap until a single molecular connection is made.

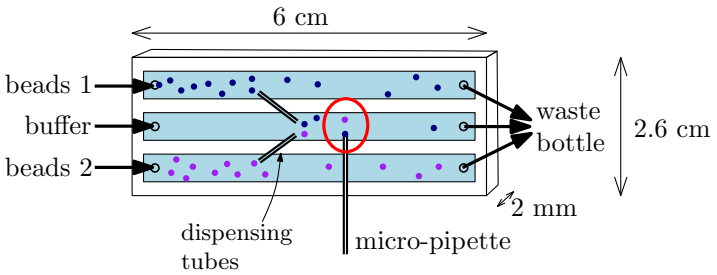

Figure 2.11: Micro-fluidics chamber. The chamber is made of three channels through which beads and buffer are flowed. The lateral channels are connected with the central one through dispensing tubes. Experiments are carried out close to the tip of the micro-pipette, in the central channel (circled region).

2.3 Single-molecule experiments with the mini-tweezers

The key idea is to tether the molecular system under study between the two beads. Such molecular system may consist of a DNA hairpin or protein tethered between two DNA double helix handles (Fig. 2.12a) or an antibody-antigen connection (Fig. 2.12b). In any of the cases, one bead is immobilized in the tip of the micro-pipette and the other is captured in the optical trap. The mini-tweezers instrument allows to control the distance λ between the center of the optical trap and the tip of the micro-pipette. By increasing or decreasing this distance, force is applied to the molecular system and the experimental force-distance curve (FDC) can be recorded.

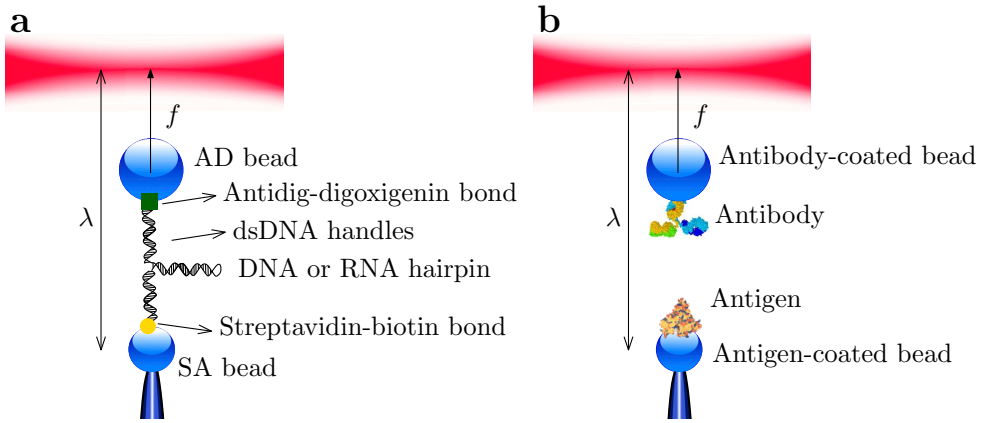

Figure 2.12: Single-molecule experiments with the mini-tweezers. a. Setup to study the unfolding/folding under mechanical force of a nucleic acid hairpin (DNA or RNA). **b.** Setup to investigate the antibody-antigen interaction.

To study the unfolding and folding of a DNA hairpin or a protein, beads of type 1 (upper channel in Fig. 2.11) are streptavidin-coated beads (SA beads), and beads of type 2 (bottom channel in Fig. 2.11) are antidigoxigenin-coated beads (AD beads). Before flowing AD beads through the micro-fluidics chamber, they are incubated together with the DNA hairpins that have been modified so that they contain tail of digoxigenin molecules on one of their ends. Consequently, some DNA molecules bind to the surface of AD beads through digoxigenin-antidigoxigenin bonds, which are very strong. The other extreme of the DNA molecules are modified with a biotin molecule. Therefore, by bringing the two beads close together in the chamber a biotin-streptavidin connection will be occasionally formed, thus having a single molecule between the two captured beads (Fig. 2.12a).

To study the antibody-antigen interaction, beads of type 1 are coated with antigens and beads of type 2 are coated with antibodies. By bringing the two beads close together in the chamber an antibody-antigen bond will be occasionally formed, giving rise to the possibility of applying forces on it (Fig. 2.12b).

2.3.1 Force calibration

The setup described in Fig. 2.10 allows to measure the change in light momentum for the two counter-propagating laser beams before and after interacting with the trapped bead. Therefore, the calibration of force can be performed from the physical law of conservation of momentum,

$$\vec{f} = \frac{d\vec{p}}{dt}. \tag{2.2}$$

A second way to calibrate the force is to use the Stokes law, that relates the force $\vec{f}$ exerted on a spherical particle to the speed $\vec{v}$ at which this particle is moving:

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

being R the particle's radius and μ the shear viscosity of the medium. Hence, the calibration procedure consists of trapping a microsphere with the optical trap inside the micro-fluidics chamber and moving the chamber at different controlled speeds. If R and η are known, the drag force acting on the bead can be extracted and force measurements calibrated.

Finally, the Brownian motion of the trapped bead in an optical trap can be used in force calibration. As sketched in Fig. 2.8, the potential generated by the optical trap is well approximated by an harmonic potential. In this condition, the Langevin equation that properly describes the movement of the microsphere in the optical trap is:

$$\gamma\frac{dx}{dt} = -k_{\mathrm{b}}x + \eta(t), \tag{2.4}$$

where the inertia term md^2x/dt^2 is neglected (over-damped limit); x is a one-dimensional coordinate for the relative position of the bead to the center of the optical trap (where the force is zero); k_{b} is the stiffness of the optical trap; γ is the viscosity coefficient, equal to $6\pi\mu R$ (see Eq. 2.3); and $\eta(t)$ is the stochastic force acting on the bead due to thermal fluctuations. Equation (2.4) can be solved and an analytical expression for the power spectrum of the force $S_f(\nu)$ (defined as the Fourier transform of the force correlation function) can be obtained:

$$S_f(\nu) = 2\frac{k_{\mathrm{b}}^2 k_{\mathrm{B}}T}{\gamma}\frac{1}{4\pi^2\nu^2 + \left(\frac{k_{\mathrm{b}}}{\gamma}\right)^2}. \tag{2.5}$$

The methodology here is to measure the displacements of the bead captured in the optical trap at high bandwidth (50 kHz) and fit the resulting power spectrum to Eq. (2.5) in order to get the trap stiffness k_{b}. In Fig. 2.13 the power spectrum of a bead captured in the optical trap and the fit to the theoretical expression given by Eq. (2.5) are shown. In this case, $k_{\mathrm{b}} = 0.066 \pm 0.006$ pN/nm and $\gamma = (3.0 \pm 0.1) \times 10^{-7}$ pN·s/nm.

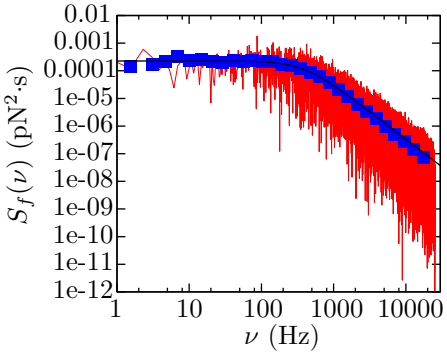

Figure 2.13: Power spectrum of a bead captured in the optical trap. Raw data is shown in red, blue squares are the average, and the fit to Eq. (2.5) is plotted in black.

2.3.2 Experimental protocols

As mentioned before, with the mini-tweezers instrument it is possible to macroscopically control the relative distance λ between the center of the optical trap and the tip of the micro-pipette (Fig. 2.12). For this reason, λ is the so-called "control parameter" in most of the experiments performed. At the same time, the instrument allows to measure forces. Therefore, λ can change at our will in order to perform different experimental protocols to manipulate the molecular system under study. The most typical experimental protocols are summarized in what follows:

Pulling experiments

In a pulling experiment the optical trap is moved at a constant pulling speed v, and consequently $\lambda(t) = \lambda_0 + vt$. When λ increases (decreases), the force applied to the molecular system also increases (decreases) at the loading rate r, therefore satisfying $f = f_0 + rt$. r and v are mathematically related through the effective stiffness of the experimental system, k_{eff}, and $r = v \cdot k_{\text{eff}}$.

This protocol is used to carry out non-equilibrium single-molecule experiments.

Passive-mode experiments

In passive-mode experiments, the position of the trap is kept constant and the captured bead passively moves in the optical trap. These experiments are often performed under equilibrium conditions, where the accessible states at the chosen value of λ are sampled according to the Gibbs probability density function. Transitions between different states are observed through changes in the values of the measured force.

Constant-force experiments

Despite the fact that the distance λ is the natural control parameter of the mini-tweezers instrument, it is possible to implement a feedback control operating at 1 kHz that keeps

the force almost constant by correcting the position of the bead relatively to the center of the trap by changing λ. Therefore, using the feedback mechanism the molecular construct can be kept approximately at a constant value of force and transitions between accessible states are detected through changes in λ.

2.4 Summary

In this chapter a brief description of the biomolecules studied in this thesis is presented. Hence, the chemical properties and biological tasks of nucleic acids (both DNA and RNA) and proteins are presented. In addition, qualitative explanations of how the maturation line of the humoral immune system works and the main properties of antigens and antibodies are provided.

On the other hand, the discovery and physical principles of optical trapping are explained. Different optical trap setups are introduced, and the mini-tweezers instrument, which is the one used to perform single-molecule experiments throughout this thesis, is presented in detail. In this direction, the different elements of the setup are shown and the steps required to calibrate and perform single-molecule experiments are summarized. Finally, the main experimental protocols (pulling, passive mode and constant force) are briefly explained by characterizing how the control parameter, which in this case is the relative distance between the optical trap and the tip of the micro-pipette, changes with time.

Part II

Dynamic force spectroscopy

Chapter 3

Elastic properties of single-stranded DNA

The elastic properties of biomolecules (such as proteins and nucleic acids) play a relevant role in several processes occurring at the molecular scale, like the DNA packing inside the cell nucleus, important to store genetic information; the formation of the actin filament network inside the cytoplasm, which is crucial for nutrient transport; or the transcription and translation of nucleic acids performed by molecular motors to read and process the genetic information. Moreover, elastic properties determine the molecular structure and therefore are highly related to the function of molecules.

The elastic properties of the DNA double helix, also referred to as double-stranded DNA (dsDNA) have been deeply investigated using OT by pulling a long molecule from its ends [Bust 94, Smit 96, Bouc 99]. At forces below 65 pN, dsDNA behaves according to the worm-like chain (WLC) model with a persistence length equal to $\sim$50 nm (Fig. 3.1a). In contrast, at higher forces dsDNA undergoes a highly cooperative transition known as the overstretching, where it dramatically increases its extension with apparently no force resistance [Mame 09].

In the case of single-stranded DNA (ssDNA) the elastic properties are more difficult to access. Even though at large forces ($>$ 10 pN) the stretched molecule is well-fitted by the ideal freely-jointed chain (FJC) or WLC models [Smit 96, Hugu 10b, Bosc 13, Dess 02], at low forces ($<$ 10 pN) the molecule self-interacts and forms secondary structure through hydrogen bonds occurring between different nucleotides. This effect is revealed by the presence of a smooth plateau in the force-extension curve at low forces which deviates from ideal elastic behaviors (Fig. 3.1b).

A first attempt to investigate the elastic properties of single-stranded nucleic acid molecules unable to form secondary structure was to stretch a poly-U molecule using OT [Seol 04] or to perform Förster resonance energy transfer (FRET) experiments on a poly-T molecule [Murp 04]. However, in case of a sequence-dependent elastic behavior, such molecules can lead to systematically deviated parameters. Furthermore, both studies have focused on long ssDNA, and the question remains whether the elastic properties of long ssDNA can be extrapolated to short ssDNA (a few tens of bases), which is much

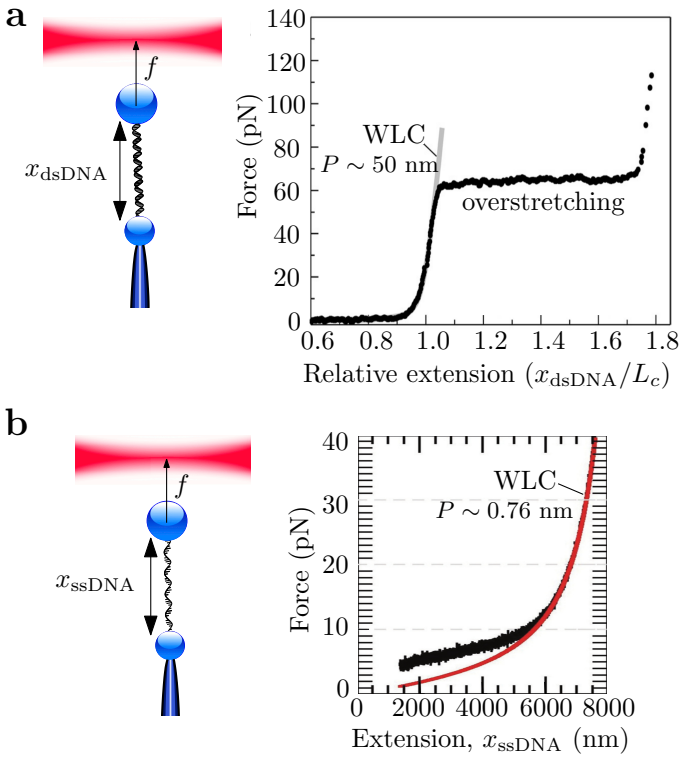

Figure 3.1: Direct measurement of the elastic properties of DNA. a. A dsDNA molecule is attached between two polystyrene beads and the response of the molecular extension to applied force can be measured (black line). At low forces it is well-described by the WLC model (gray line), and at large forces the overstretching transition is observed. Figure taken from [Mame 09]. **b.** A ssDNA molecule can be tethered between two polystyrene beads, as described in [Bosc 13]. At large forces the extension versus force response is in agreement with the WLC model (red line). At low forces, a plateau suggests the formation of secondary structure. Figure taken from [Bosc 13].

more relevant for genomic regulation.

In this chapter two new methods are introduced to extract the ideal elastic response of short biopolymers that fold into compact structures at low forces. The key idea is to measure and compare the elastic response of the polymer in the folded and unfolded states using dynamic-force spectroscopy (DFS) experiments [Alem 14]. First, the force-dependent molecular extension released upon unfolding and absorbed upon folding is extracted. Second, the molecular stiffness of the polymer is measured by determining the difference in elastic compliances of the polymer along the folded and unfolded force branches. The two methods are illustrated for the case of the elastic response of ssDNA by using OT to pull short DNA hairpins that fold into well-known double helix secondary structures. The methods can also be used to unravel the elastic properties of the polypeptide chain in proteins (Chapter 7) and, by suitably engineering molecular constructs, biopolymers in general.

3.1 Pulling experiments with short DNA hairpins

In order to obtain the ideal elastic response of short ssDNA molecules, pulling experiments are performed with the mini-tweezers on short DNA hairpins with different contour lengths. To this end, a hairpin is tethered between two identical dsDNA handles, and the free end of each handle is attached through different antibody-antigen bonds (digoxigenin-antidigoxigenin in one end; biotin-streptavidin in the other) to two micron-sized polystyrene microspheres, as described in Appendix A.1 [Forn 11]. One bead is immobilized by air suction in the tip of a micropipette whereas the other is captured in an optical trap produced by a highly stable dual-beam optical trap (Section 2.3 and Fig. 2.12a).

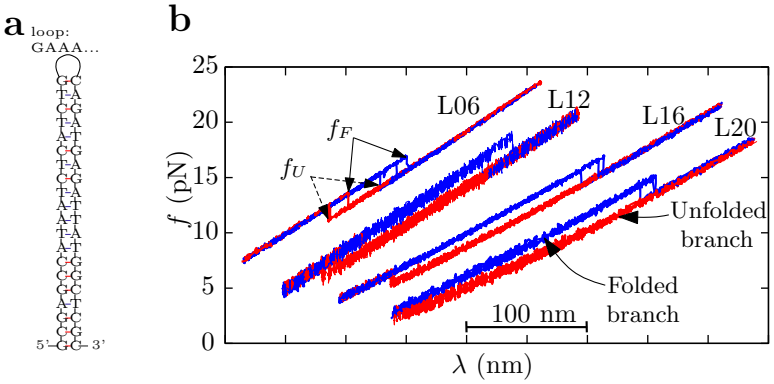

Figure 3.2: Pulling experiments with short DNA hairpins. **a.** Sequence of hairpins LX: the four hairpins have the same stem and different loop sizes. All loops consist of a G followed by five, eleven, fifteen or nineteen A (hairpin L06, L12, L16 and L20, respectively). **b.** Unfolding (blue) and folding (red) FDC measured in pulling experiments performed at 60 nm/s for hairpins L06, L12, L16 and L20. The two force-branches (unfolded and folded) are observed in the FDC of each of the four molecules. Solid (dashed) arrows indicate forces f_F (f_U) in the folded (unfolded)-force branch along unfolding and folding transitions.

A DNA hairpin is made of a ssDNA molecule with a palindromic sequence that in the absence of force forms a double helix, as sketched in Fig. 2.3a. All the hairpins studied in this chapter have an identical stem made of 20 base pairs and different loop sizes. The name of each hairpin is LX, being X the number of bases in the loop (L06, L12, L16 and L20). Under applied force, these short DNA hairpins show a two-state behavior: they are found either at their native/folded state (N), where the stem forms dsDNA, or at their unfolded stretched conformation (U), where the stem is found as ssDNA [Wood 06a, Mano 06]. This can be seen in Fig. 3.2b, where the force-distance curve (FDC) obtained by pulling at 60 nm/s in a buffer containing 1M NaCl concentration for the four different short DNA hairpins are shown. Two force-branches are observed inthe FDC: the upper branch shows the elastic response of the whole molecular construct (handles plus molecule) when the hairpin is folded, whereas the lower branch shows it when the hairpin is unfolded. They are referred to as the folded and unfolded branch, respectively. Transitions between states N and U are viewed as a sudden jump in force

where the molecule transits from one branch to the other. Such jump is a consequence of the displacements of the bead towards/away from the center of the optical trap due to the releasing/absorption of ssDNA that takes place when the molecule unfolds/folds. Forces at which such unfolding and folding transitions take place change upon repetition of the same experiment due to thermal fluctuations. It can be also observed that the larger the loop size, the larger the hysteresis between unfolding and folding curves [Wood 06a].

3.1.1 Model of the molecular system

As already mentioned in Section 2.3, the mini-tweezers instrument allows us to control the relative distance λ between the tip of the micropipette and the center of the optical trap. This distance is equal to the sum of contributions coming from the position of the captured bead relative to the optical trap and the molecular extension of the handles and the hairpin projected along the force axis. Therefore, according to the scheme provided in Fig. 3.3:

$$\lambda(f) = x_{\mathrm{h}}(f) + x_{\mathrm{b}}(f) + x_{\mathrm{DNA}}(f), \tag{3.1}$$

where $x_{\mathrm{h}}(f)$ is the extension of the handles at force f, $x_{\mathrm{b}}(f)$ is the position of the bead relative to the center of the optical trap, and $x_{\mathrm{DNA}}(f)$ is the molecular extension of hairpin that depends on its state. If the hairpin is folded (Fig. 3.3a), $x_{\mathrm{DNA}}(f)$ is equal to the projection of the double helix diameter, $x_{\mathrm{d}}(f)$, into the force axis. If the hairpin is unfolded (Fig. 3.3b), $x_{\mathrm{DNA}}(f)$ is equal to the equilibrium end-to-end distance of the ssDNA at the given force f, x_{ssDNA}.

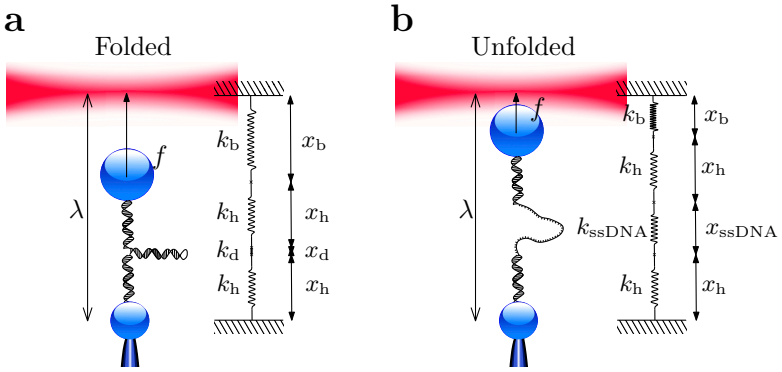

Figure 3.3: Contributions to the total trap-pipette distance λ. Schematic illustration of the elements involved in single-molecule pulling experiments performed on a DNA hairpin, from the point of view of distance and stiffness, when the molecule is folded (**a**) or unfolded (**b**).

For a small change in the trap-pipette position, $\Delta\lambda$, there is a small change in force, Δf. Division between $\Delta\lambda$ and Δf gives:

$$\frac{\Delta\lambda}{\Delta f} = \frac{1}{k_{\mathrm{eff}}} \tag{3.2a}$$

$$= \frac{1}{k_{\mathrm{h}}} + \frac{1}{k_{\mathrm{b}}} + \frac{1}{k_{\mathrm{DNA}}}, \tag{3.2b}$$

where the stiffness k_i (i = h, b or DNA) is equal to $\Delta f/\Delta x_i$. The stiffness of the DNA, k_{DNA}, depends on whether the molecule is folded, k_{d}, or unfolded, k_{ssDNA}. The effective stiffness of the system, $k_{\mathrm{eff}} = \Delta f/\Delta\lambda$, is equal to the slope of experimental FDC, and it is the combination of different serially-connected springs, each one related to a different part of the experimental setup (Fig. 3.3).

3.2 Elastic properties from force-jump measurements

The FDC shows that when a two-state DNA hairpin changes conformation (from U to N and vice-versa) a sudden force-jump $\Delta f = f_F - f_U$ is observed, whereas λ remains constant ($\Delta\lambda = 0$). f_F and f_U correspond to the forces in the folded and unfolded branches at the transition, respectively (Fig. 3.2b). Therefore, from Eq. (3.1):

$$\begin{aligned} \lambda(f_F) &= \lambda(f_U) \\ x_{\mathrm{h}}(f_F) + x_{\mathrm{b}}(f_F) + x_{\mathrm{d}}(f_F) &= x_{\mathrm{h}}(f_U) + x_{\mathrm{b}}(f_U) + x_{\mathrm{ssDNA}}(f_U). \end{aligned} \tag{3.3}$$

If the value of Δf is divided by the effective stiffness of the system when the hairpin is folded, k^F_{eff}, if follows from Eq. (3.2b) that:

$$\frac{\Delta f}{k^F_{\mathrm{eff}}} = \Delta f \left(\frac{1}{k_{\mathrm{h}}} + \frac{1}{k_{\mathrm{b}}} + \frac{1}{k_{\mathrm{d}}} \right) \tag{3.4a}$$

$$= \Delta x_{\mathrm{h}} + \Delta x_{\mathrm{b}} + \Delta x_{\mathrm{d}} \tag{3.4b}$$

$$= [x_{\mathrm{h}}(f_F) - x_{\mathrm{h}}(f_U)] + [x_{\mathrm{b}}(f_F) - x_{\mathrm{b}}(f_U)] + [x_{\mathrm{d}}(f_F) - x_{\mathrm{d}}(f_U)] \tag{3.4c}$$

Inserting Eq. (3.3) into (3.4) it can be shown that:

$$x_{\mathrm{ssDNA}}(f_U) = \frac{\Delta f}{k^F_{\mathrm{eff}}} + x_{\mathrm{d}}(f_U), \tag{3.5}$$

Therefore, the elastic response of ssDNA, $x_{\mathrm{ssDNA}}(f)$, is accessible from the unfolding and folding force-jump values measured when pulling a two-state DNA hairpin. The same method is used by X. Zhang and colleagues in reference [Zhan 09] to determine the elastic properties of the peptide chain, but the contribution of the folded molecule is there neglected. It will be seen that this provides slightly stiffer elastic properties for ssDNA.

To get the elastic response of the ssDNA released or absorbed by each DNA hairpin during unfolding and folding, $x_{\mathrm{ssDNA}}(f)$, the pair of values $\Delta f = f_F - f_U$ and k^F_{eff} are determined and divided by each other for each unfolding and folding transition. Next, according to Eq. (3.5), the extension of the hairpin double-helix diameter at force f_U, $x_{\mathrm{d}}(f_U)$, is added. Such contribution is modeled as a single bond of lenght d (hereafter taken equal to the helix diameter, $d = 2$ nm) that is oriented due to the action of a force [Wood 06a, Forn 11]. The energetic contribution that describes this effect is given by $E = -fd\cos\theta$, being $\theta \in [0, \pi]$ the relative angle between the bond and the force axes.

The corresponding partition function is:

$$Z \propto \int_{-1}^{1} d\,(\cos\theta)\, e^{\frac{fd\cos\theta}{k_{\mathrm{B}}T}} = \frac{k_{\mathrm{B}}T}{fd} \sinh\left(\frac{fd}{k_{\mathrm{B}}T}\right), \tag{3.6}$$

where k_{B} is the Boltzmann constant and T is the absolute temperature taken equal to 298 K. Hence, the average distance x_{d} projected along the force axis is:

$$x_{\mathrm{d}}(f) = d\langle\cos\theta\rangle = k_{\mathrm{B}}T\frac{\partial \log Z}{\partial f} \tag{3.7a}$$

$$= d\left[\coth\left(\frac{fd}{k_{\mathrm{B}}T}\right) - \frac{k_{\mathrm{B}}T}{fd}\right], \tag{3.7b}$$

which is equal to the FJC model describing the elastic behavior of a polymer with identical Kuhn and contour lengths.

Figure 3.4a shows the 2D-contour plot of the histogram of forces f_U plotted against the histogram of $x_{\mathrm{ssDNA}}(f_U)$ values derived for each of the four DNA hairpins.

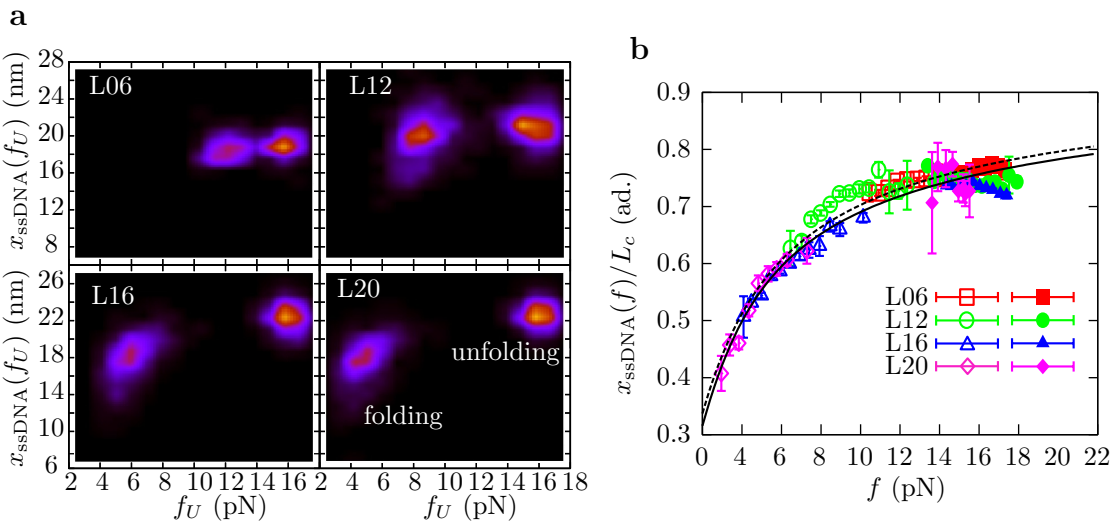

Figure 3.4: Elastic response of ssDNA by mechanically unfolding DNA hairpins. a. 2D-contour plot of the histogram of x_{ssDNA} measured values plotted against the histogram of f_U measured in unfolding and folding events for DNA hairpins L06, L12, L16 and L20. **b.** Normalized elastic response of ssDNA. Solid(empty) symbols refer to values measured using data from unfolding(folding) curves. The black-solid line shows the fit to the WLC model using the interpolation formula proposed by Bustamante and collaborators ($a_i = 0$ in Eq. 3.8, [Bust 94]), and the black-dashed line shows the fit to the WLC model using the extended version proposed by Bouchiat *et. al.* [Bouc 99]. Error bars are standard errors from averaging over different molecules and force bins.

The elastic response of biomolecules such as ssDNA, RNA or peptide chains, is usually assumed to behave according to the WLC model. The interpolation formula for this elastic model is generally taken as [Bust 94, Bouc 99]:

$$f = \frac{k_{\mathrm{B}}T}{P}\left[\frac{1}{4(1 - x_{\mathrm{ssDNA}}/L_c)^2} - \frac{1}{4} + \frac{x_{\mathrm{ssDNA}}}{L_c} + \sum_{n=2}^{7} a_n \left(\frac{x_{\mathrm{ssDNA}}}{L_c}\right)^2\right], \tag{3.8}$$

being P the persistence length and L_c the molecular contour length of the molecule. For the DNA hairpins under study $L_c = (40+X)d_b$, where 40 and X are the number of bases in the stem and the loop, respectively (Fig. 3.2a), and d_b is the interphosphate distance. The coefficients a_i ($i = 1, \ldots, 7$) can be taken equal to zero according to Bustamante and collaborators in [Bust 94], or to the numerical values proposed by in Bouchiat *et. al.* in reference [Bouc 99] ($a_2 = -0.5164228$, $a_3 = -2.737418$, $a_4 = 16.07497$, $a_5 = -38.87607$, $a_6 = 39.49944$, $a_7 = -14.17718$). Hence, by fitting the experimental values of $x_{\mathrm{ssDNA}}(f)$ determined from pulling experiments to Eq. (3.8), numerical values for the persistence length P and the interphosphate distance d_b will be obtained.

According to ideal elastic models for semi-flexible polymers, the molecular extension at a given force is proportional to the contour length and hence $f = f(x_{\mathrm{ssDNA}}/L_c)$. Consequently, data from the four molecules (L06, L12, L16 and L20) can be merged into a single curve by dividing each experimental $x_{\mathrm{ssDNA}}(f)$ by the corresponding value of the contour length L_c. In order to obtain the points represented in Fig. 3.4b, data from Fig. 3.4a is averaged over different force bins and results for the different $x_{\mathrm{ssDNA}}(f)$ values for each DNA hairpin are divided by the corresponding L_c. Since d_b is an unknown parameter, the merging procedure is repeated for different values of d_b ranging from 0.2 to 0.7 nm/base, and each merged data set is independently fitted to Eq. (3.8). The pair of values for the persistence length and the distance between bases that best fit our experimental results are $P = 1.35 \pm 0.05$ nm, $d_b = 0.58 \pm 0.02$ nm/base and $P = 1.10 \pm 0.05$ nm, $d_b = 0.59 \pm 0.02$ nm/base, depending on whether the interpolation formula of [Bust 94] or the extension proposed in [Bouc 99] are used (Table 3.1).

WLC model	P (nm)	d_b (nm/base)
[Bust 94], $a_i = 0$	1.35±0.05	0.58±0.02
[Bouc 99], $a_i \neq 0$	1.10±0.05	0.59±0.02

Table 3.1: Elastic parameters of ssDNA obtained by force-jump measurements in pulling experiments. The elastic response of short ssDNA molecules can be fitted to the WLC model taking the persistence length P and the distance between bases d_b as free parameters. Numerical values have been obtained by fitting data shown in Fig. 3.4b to two different theoretical expressions of the WLC model [Bust 94, Bouc 99].

The quality of the force-jump measurement to get the elastic properties of ssDNA has been tested by analyzing simulated pulling experiments. As it can be seen in Appendix B, results validate the method presented in this section.

If the presence of the hairpin diameter is neglected the values P, d_b that best fit experimental data are $P = 1.45 \pm 0.05$ nm, $d_b = 0.53 \pm 0.02$ nm/base or $P = 1.20 \pm 0.05$ nm, $d_b = 0.54 \pm 0.02$ nm/base. Therefore, persistence lengths are slightly overestimated whereas interphosphate distances tend to be shorter. This trend is also confirmed with simulated data (Appendix B).

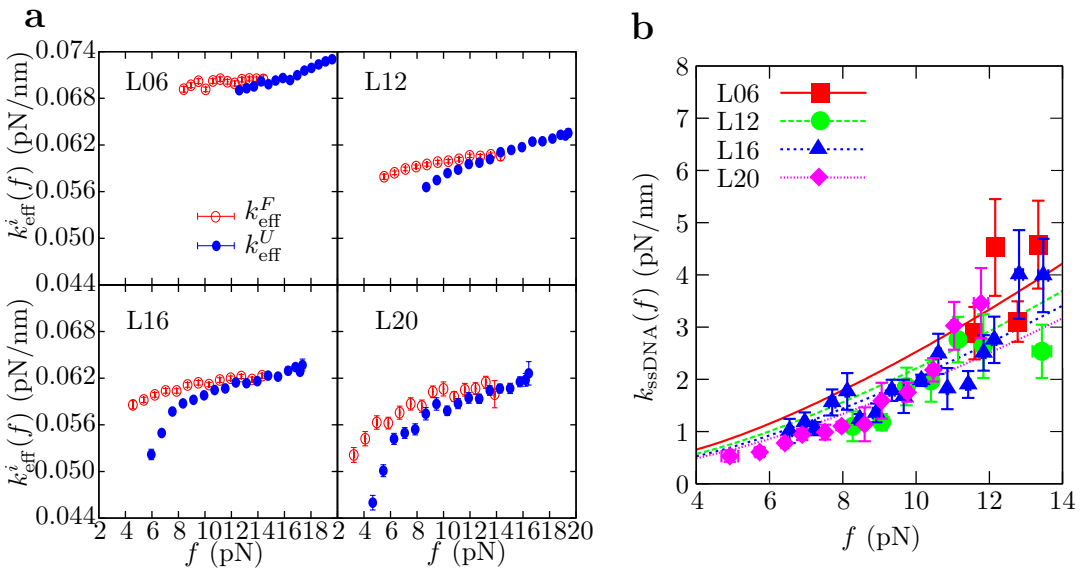

Figure 3.5: Force-dependent stiffness of ssDNA measured by mechanically unfolding DNA hairpins. a. Effective stiffnesses of the whole molecular system when the hairpin is folded ($k^F_{\rm eff}$, open-red circles) and unfolded ($k^U_{\rm eff}$, closed-blue circles) measured for the four DNA hairpins. **b.** Experimental data and fit to the WLC model using the expression from Bouchiat and collaborators [Bouc 99] for molecule L06 (squares and solid line), L12 (circles and dashed line), L16 (triangles and dotted line) and L20 (diamonds and dashed-dotted line). Error bars are standard errors from averages taken over different molecules of a given type.

3.3 Elastic properties from effective stiffnesses

The elastic response of ssDNA can also be extracted from the measurement of the effective stiffness of the molecular construct when the hairpin is folded and unfolded. When hysteresis effects between unfolding and folding processes are large enough, both the folded and the unfolded force-branches can be measured from pulling experiments for a large range of forces. First, the effective stiffnesses $k^F_{\rm eff}$ and $k^U_{\rm eff}$ can be extracted at different forces from the slopes of each force-branch. Results obtained by averaging the experimental FDC for the different molecules are shown in Fig. 3.5a. It can be seen that, as expected, $k^F_{\rm eff} > k^U_{\rm eff}$ and that the range of forces where both stiffnesses can be measured increases with hysteresis effects (*i. e.*, with loop size). Second, the inverse values of $k^U_{\rm eff}$ and $k^F_{\rm eff}$ at a given force f can be subtracted, and it can be shown from Eq. (3.2b) that it is satisfied that:

$$\frac{1}{k^U_{\rm eff}(f)} - \frac{1}{k^F_{\rm eff}(f)} = \frac{1}{k_{\rm ssDNA}(f)} - \frac{1}{k_{\rm d}(f)}. \tag{3.9}$$

Therefore, the stiffness of ssDNA, $k_{\rm ssDNA}(f)$ can be measured as a function of force from the slopes of the FDC.

Again, the helix diameter is modeled as a single bond under an applied force, with

a total length of $d = 2.0$ nm. From Eq. (3.7b) it can be derived that:

$$\frac{1}{k_{\mathrm{d}}(f)} = \frac{dx_{\mathrm{d}}}{df} = \frac{d^2}{k_{\mathrm{B}}T}\left[-\frac{1}{\sinh^2\left(\frac{fd}{k_{\mathrm{B}}T}\right)} + \left(\frac{fd}{k_{\mathrm{B}}T}\right)^2\right]. \tag{3.10}$$

Results of the force-dependence of $k_{\mathrm{ssDNA}}(f)$ obtained for the four different molecules are shown together in Fig. 3.5b. All data can be simultaneously fitted to the interpolation formula of the WLC model:

$$k_{\mathrm{ssDNA}} = \frac{k_{\mathrm{B}}T}{PL_c}\left[\frac{1}{2(1 - x_{\mathrm{ssDNA}}/L_c)^3} + 1 + \sum_{n=2}^{7} n a_n \left(\frac{x_{\mathrm{ssDNA}}}{L_c}\right)^{n-1}\right], \tag{3.11}$$

where, as before, the coefficients a_n ($n = 1, 7$) can be taken equal to zero [Bust 94] or to the numerical values proposed in reference [Bouc 99]. In each case, the pair of values for the persistence length and the distance between bases that best fit our experimental data are $P = 2.1 \pm 0.1$ nm, $d_b = 0.75 \pm 0.2$ nm/base or $P = 1.00 \pm 0.05$ nm, $d_b = 0.56 \pm 0.02$ nm/base (Table 3.2).

Despite the good agreement found between the two fitting approaches for both P and d_b when using the interpolation formula proposed by Bouchiat and coworkers in [Bouc 99], a remarkable discrepancy is found when using the WLC expression proposed in [Bust 94]. However, if we carry out this latter fit by setting d_b equal to 0.58 nm/base, we obtain $P = 1.35 \pm 0.5$ nm, which again is in very good agreement with the previous fitting method.

Again, the quality of the measurement of the stiffness to extract the elastic properties of short ssDNA molecules has been tested by analyzing simulated data. Results validate the method presented here (see Appendix B).

	P (nm)	d_b (nm/base)
[Bust 94], $a_i = 0$	2.1±0.1	0.75±0.02
	1.35±0.05	0.58
[Bouc 99], $a_i \neq 0$	1.00±0.05	0.56±0.02

Table 3.2: Elastic parameters of ssDNA obtained by stiffness measurements. Numerical results of the fit to the WLC model of the stiffness of short ssDNA molecules. The theoretical expression proposed by Bustamante and collaborators [Bust 94] has been used by taking both the persistence length P and the distance between base pairs d_b as free parameters (first raw), or by fixing the value of d_b (second raw). The fit can also be performed using the theoretical expression found in reference [Bouc 99] by taking both the persistence length P and the distance between base pairs d_b as free parameters (third raw).

3.4 Conclusions

Elastic properties of nucleic acids and proteins are determinant factors for the molecular structure and play an important role in genomic regulation. Single-molecule and DFS experiments pave the way to unravel the mechanical properties of such biomolecules with unprecedented detail. The measurement of the ideal elastic properties of biopolymers at low forces is usually hindered by the formation of high-order structures, such as secondary and tertiary structures. Most studies to date have focused on the elastic response of long polymers whereas measurements in case of short polymers appear technically more challenging, requiring methods such as FRET that nevertheless are strongly affected by intra-chain diffusive kinetics and secondary structure formation. In this regard, the development of methods that directly probe the ideal elastic response of short polymers capable of forming high-order structures is essential.

In this chapter it is shown how to extract the ideal elastic properties of short ssDNA molecules by mechanically unfolding short DNA hairpins of different contour lengths [Alem 14]. Two different methods are proposed: in the first method (Section 3.2), the released/absorbed molecular extension upon hairpin unfolding/folding is related to the end-to-end distance of ssDNA, $x_{\mathrm{ssDNA}}(f)$. In the second method (Section 3.3), the slope of the experimentally measured force-distance curves is related to the rigidity of ssDNA, $k_{\mathrm{ssDNA}}(f)$, further characterizing its elastic response. For a given molecule, $x_{\mathrm{ssDNA}}(f)$ can only be measured in the range of forces at which the hairpin unfolds and folds, which is usually a small force interval (see hairpins L06 and L12 in Fig. 3.2a-b). The same occurs with $k_{\mathrm{ssDNA}}(f)$ that should be measured along a force range where both the folded and unfolded states are kinetically stable and hence the respective force-branches are experimentally accessible. In order to increase the force interval explored in the two methods, pulling experiments are performed on DNA hairpins with identical stem sequence and different loop sizes. The second method is particularly useful for biomolecules displaying strong hysteresis effects in pulling experiments (Fig. 3.5a). However, error bars for the measured rigidities tend to be large due to experimental artifacts, such as tether misalignment and drift effects.

The results obtained by the two methods were fit to theoretical expressions of the WLC model [Bust 94, Bouc 99] to extract values for the interphosphate distance d_b and the persistence length P of ssDNA. By averaging over the best fits obtained with the two methods (Tables 3.1 and 3.2), the values $d_b = 0.58 \pm 0.02$ nm/base and $P = 1.3 \pm 0.2$ nm are obtained. Both numbers were obtained at room temperature (298 K) and standard 1M NaCl ionic conditions. The value found for d_b is in very good agreement with the one generally reported in the literature ($d_b = 0.59$ nm/base). In contrast, previous measurements of P for ssDNA molecules obtained at 1M NaCl by fitting the WLC model to the force-extension curves predict lower values nearly half the ones reported here ($P = 0.76 \pm 0.05$ nm) [Smit 96, Hugu 10b, Bosc 13].

Results here reported show that the elastic properties of ssDNA molecules of a few tens of base pairs differ from those reported for long ssDNA molecules (of a few kilobases). This fact has also been observed in previous works, for example FRET studies

of a series of oligodeoxythymidylates performed over a wide range of salt concentrations and chain lengths (between 10 and 70 nucleotides) have reported comparably large values for P (between 1.5 and 5 nm, depending on salt concentration) [Murp 04]. This is much larger than the reported values for long ssDNA (in the range 0.7-1nm depending on ionic conditions). The origin of such discrepancy is presently unknown and might be related to finite-size effects. Future studies could address this point by checking the systematic effect of sequence and finite contour length corrections on the elastic properties of nucleic acids at different ranges of forces.

A remarkable property of the two methods here presented to extract elastic properties is that neither of them requires a direct measurement of the absolute molecular extension, x_m. This becomes an advantage in OT setups that are only able to provide relative measurements of the molecular extension, that is, that take an undefined origin of coordinates [Bosc 13]. In many OT setups where video imaging can be used to measure bead-to-bead distances, it is often very difficult to extract the absolute value of the molecular extension in case of tethers that are misaligned due to geometrical constraints. In such cases, an extra free parameter x_0 related to the zero in molecular extension must be introduced in the theoretical force-extension curve $f = f(x_m - x_0)$. The fit of the experimental data to such expressions can be inaccurate and lead to systematic errors in the derivation of the elastic parameters of the biomolecules under study.

Another aspect that requires careful examination is related to the treatment given to the folded hairpin. It is considered that the folded hairpin behaves as a single bond oriented under the action of an external force. Previous studies that use the force-jump measurement to extract elastic properties decided not to include in the analysis the elastic effect of the folded molecule. However, it is shown here from both simulation and experiments that neglecting the presence of the hairpin diameter leads to stiffer elastic properties for ssDNA. Even though results are compatible within error bars, the relative error made in the estimation of the persistence length is $\sim$10%, whereas for the interphosphate distance it is $\sim$2%. Hence, it is important to take into account the presence of the DNA double helix along the folded force-branch to determine the persistence length of short ssDNA by pulling DNA hairpins.

The main conclusion of this chapter is that it is possible to measure the ideal elastic properties of ssDNA between 1 and 20 pN, a force range where the determination of the ideal elastic response for large ssDNA molecules is particularly difficult due to the formation of secondary structures. The methodology presented here can be applied to any molecule displaying two-state behavior under pulling experiments and exhibiting a cooperative transition mediating folded and unfolded states. In particular, it would be very interesting to apply this method to extract the elastic parameters of ssDNA, RNA, and polypeptides (Chapter 7) by varying ionic strength and temperature.

Chapter 4

Unraveling the free-energy landscape of biomolecules

One of the main challenges in molecular biophysics is to understand how biomolecules explore their configurational space in the short timescales at which most biological phenomenon occur. A random search of a particular molecular conformation (such as the most thermodynamically stable) can take an average time larger than the age of the Universe, even for small proteins [Levi 69]. This paradox –known as the Levinthal paradox– motivated the use of the concept of the free-energy landscape (FEL) to model the molecular folding problem. In this context, the search of a given state is influenced by both thermodynamic and kinetic factors [Dinn 00].

Single-molecule experiments stimulated a new perspective in the study of molecular folding. The observation and manipulation of the behavior of a single molecule under different conditions grants access to invaluable information about its kinetic and thermodynamic properties. Using key concepts from the transition state theory (which is still under development), this information can be used to characterize the molecular FEL and to study folding and unfolding reaction pathways.

Nucleic acid hairpins are model systems that allow us to test the connection between the molecular FEL and the unfolding and folding kinetic rates. In this chapter, it is explained in great detail how to calculate the theoretical profile of the FEL of nucleic acid hairpins using the nearest-neighbor model. A strong emphasis is made on the effect of applied mechanical force, since this plays a crucial role in the interpretation of single-molecule experiments performed with optical tweezers. Additionally, non-equilibrium DFS experiments are performed on a two-state DNA hairpin and it is shown how to extract the force-dependent unfolding and folding kinetic rates by defining survival probabilities and using Markov models.

Several theoretical frameworks are proposed in the literature to extract important features of the molecular FEL from experimentally measured force-dependent unfolding and folding kinetic rates. In this chapter, advantages and limitations of the very well-accepted Bell-Evans model are discussed, and a Kramers-based approach is presented that allows us to investigate how the position and the height of the transition state

depends on force. The benefit of working with nucleic acid hairpins is that a direct comparison between experimental results and theoretical predictions is possible. In Chapter 5, it will be shown how the methodology hereafter presented can be extended to analyze kinetic intermediate states, to extract elastic properties under different ionic conditions, or to characterize the ligand binding process.

4.1 The free-energy landscape

The free-energy landscape (FEL) is the mapping between all possible configurations of a system and their corresponding free energy. Each configuration is labeled by one or more reaction coordinates, which are also used to characterize the degree of evolution of a reaction.

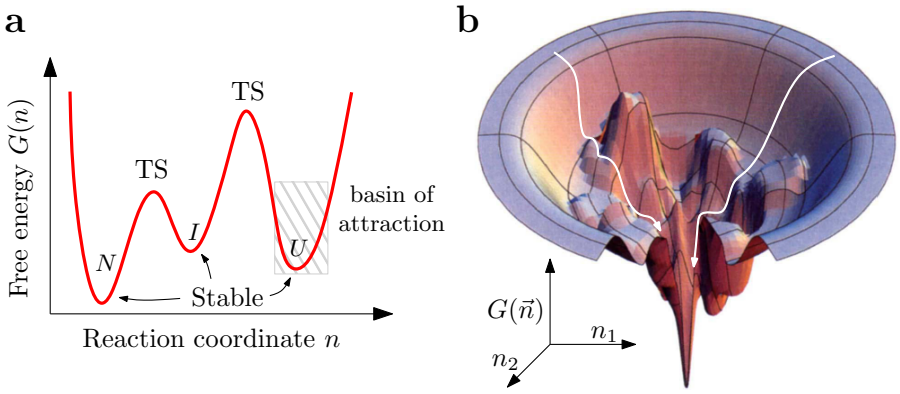

Figure 4.1: Free-energy landscape a. Representation of a one-dimensional FEL. Different stable states (N, I, U) and transition states (TS) are sketched. **b.** Three-dimensional representation of a rugged funnel-like FEL (picture reproduced from [Dill 97]). White arrows indicate two different reaction pathways connecting two different states with highest free energy to the most stable state.

Figure 4.1a shows a schematic representation of a one-dimensional FEL. In this case, the different accessible configurations are characterized by using a single reaction coordinate, indicated by n. States with a local minimum for the free-energy value are stable states (labeled here by N, I and U), whereas local maxima are transition states (TS). A system initially equilibrated at state N can evolve towards configuration I and U by overcoming the different TS. The set of configurations surrounding a stable state is usually referred to as the basin of attraction.

For some systems a single reaction coordinate is not enough to characterize the different configurations. For example, two reaction coordinates, n_1 and n_2, have been employed to sketch the FEL depicted in Fig. 4.1b. Therein several stable and transition states are represented. In this case, one may think of many alternative reaction pathways connecting highly unstable states with the state with lowest free energy, that is, the most stable state.

Measurable quantities, such as bond angles or molecular lengths, are usually assumed to be good reaction coordinates, and TS are identified as states with a local maximum

in their free-energy values. However, in a purely theoretical framework, the TS is a state from which the system has an equal probability to reach neighboring stable states. This introduces a restriction in the choice of reaction coordinates. For example, in the study of the dissociation dynamics of sodium chloride in water the inter-ionic distance $r_{\text{ion}} = |\vec{r}_{\text{Cl}} - \vec{r}_{\text{Na}}|$ was usually considered to be the reaction coordinate until molecular dynamic simulations showed that the number of water molecules in the sodium-coordination zone is more adequate [Mart 00].

To study the FEL of proteins, the fraction of native contacts (also known as degree of nativeness) is used to characterize molecular configurations [Borg 08]. In the case of nucleic acids one uses the number of open base pairs, which is proportional to the equilibrium molecular end-to-end distance. Recent studies show that the molecular extension is a good reaction coordinate to monitor molecular unfolding under applied force [Best 08, Morr 11, Lin 12].

4.1.1 Nucleic acids under mechanical stress

The understanding of how external control parameters, such as temperature, mechanical force or ionic strength, influence the FEL is crucial to predict the fate of molecular reactions under different conditions. Here the action of a mechanical force applied to the ends of a nucleic acid hairpin is theoretically described.

As already explained in Chapters 2 and 3, a hairpin consists of a single-stranded nucleic acid chain that has a palindromic sequence (see Fig. 2.3). Hence, the two complementary sections of the molecule interact through Watson-Crick base pairs and a stable double helix (hereafter referred to as stem) ended by a loop is formed (Fig. 4.2a). A theoretical description of the FEL of nucleic acid hairpins is possible by taking into account the sequential configurations that appear by progressively opening the molecule [Cocc 03, Mano 07b, Moss 09b]. Each configuration is labeled by the number of open base pairs n: n=0 corresponds to the folded state (all base pairs are formed) and $n = \text{N}$ to the fully unzipped unfolded state (Fig. 4.2b).

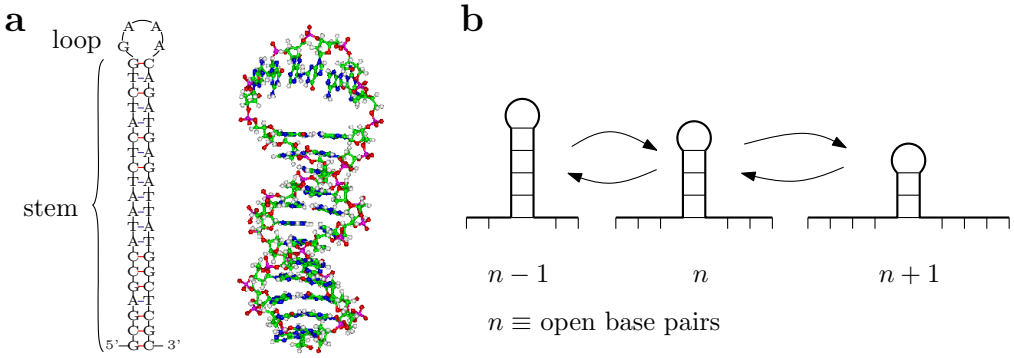

Figure 4.2: Nucleic acids hairpins. a. Sequence and structure of a DNA hairpin, named CD4, made of a 20 base pairs-long stem ended by a tetraloop. **b.** Sketch of the sequential configurations $n-1 \rightleftarrows n \rightleftarrows n+1$ that appear when progressively unzipping the base pairs of the stem, being n the total number of open base pairs.

When mechanical tension is applied to the ends of a hairpin, states with larger extension (or, equivalently, larger values of n) are energetically favored [Tino 02]. For a given force f, the free energy of a given configuration n is given by:

$$\Delta G_n(f) = \Delta G_n^0 + \Delta G_n^{\mathrm{st}}(f) + \Delta G_n^{\mathrm{d}}(f). \tag{4.1}$$

ΔG_n^0 is the free energy of formation at zero force, $\Delta G_n^{\mathrm{st}}(f)$ is the reversible work needed to stretch the released single-stranded nucleic acid and $\Delta G_n^{\mathrm{d}}(f)$ is the reversible work of orientation of the hairpin double helix. The different contributions are hereafter explained.

Free energy of formation at zero force

The first term in Eq. (4.1), ΔG_n^0, accounts for the free energy of formation at zero force of the configuration $n \in [0, \mathrm{N}]$, being N the total number of base pairs in the hairpin stem. It is a sequence-dependent magnitude and can be estimated using the nearest-neighbor (NN) model, by considering that the free energy of one base pair depends on the base itself and on its first neighbor in the 3'-5' direction. Each base-pair free energy, $g_{i,i+1}$, contains the energy of formation of the hydrogen bonds between two complementary bases, the stacking interactions between two consecutive base pairs and the loss of entropy due to bond formation.

If only Watson-Crick base pairs are considered, there are 10 different values of base-pair free energies $g_{i,i+1}$ for both DNA and RNA (Table 4.1). In 1998, John Santalucia and Douglas H. Turner unified different measurements carried out with UV absorbance and calorimetry experiments and provided the unified-oligonucleotide (UO) set of NN parameters for DNA and RNA at the standard 1M NaCl ionic conditions [Sant 98, Xia 98, Zuke 03]. These parameters have also been measured in the *Small Biosystems Lab* for DNA from single-molecule unzipping (UNZ) experiments [Hugu 10a, Hugu 10b].

Therefore, according to the NN model, ΔG_n^0 can be calculated as:

$$\Delta G_n^0 = \sum_{i=n+1}^{\mathrm{N}-1} g_{i,i+1} + g_{\mathrm{loop}}(1 - \delta_{n\mathrm{N}}), \tag{4.2}$$

where the first term $\sum_{i=n+1}^{\mathrm{N}-1} g_{i,i+1}$ is the free energy of the hairpin stem and depends on the sequence, g_{loop} is the free energy of formation of the loop, and δ_{nN} is the Kronecker delta (equal to 1 if $n = \mathrm{N}$ and 0 otherwise). The loop is an entropic unfavorable structure and consequently g_{loop} is positive. Values of g_{loop} used in this thesis are taken from the Mfold web server [Zuke 03], who in turn take values derived in previous experimental work [Hilb 85, Blom 89, Anta 91, Anta 92, Seni 88].

The stretching free energy

At each configuration n of a hairpin, $2n$ nucleotides are released from the stem. Consequently, configuration n has a fraction of single-stranded nucleic acid whose contour

DNA	$g_{i,i+1}$ (kcal/mol)		RNA	$g_{i,i+1}$ (kcal/mol)
NN base pair	UO	UNZ	NN base pair	UO
AA/TT	-1.28	-1.23	AA/UU	-1.12
TA/AT	-0.85	-0.84	UA/AU	-1.37
GA/CT	-1.58	-1.47	GA/CU	-2.73
CA/GT	-1.74	-1.66	CA/GU	-2.14
AT/TA	-1.12	-1.17	AU/UA	-1.09
TT/AA	-1.28	-1.23	UU/AA	-1.12
GT/CA	-1.72	-1.49	GU/CA	-2.41
CT/GA	-1.54	-1.36	CU/GA	-1.93
AG/TC	-1.54	-1.36	AG/UC	-1.93
TG/AC	-1.73	-1.66	UG/AC	-2.14
GG/CC	-2.07	-1.93	GG/CC	-3.26
CG/GC	-2.49	-2.37	CG/GC	-2.23
AC/TG	-1.72	-1.49	AC/UG	-2.41
TC/AG	-1.58	-1.47	UC/AG	-2.73
GC/CG	-2.53	-2.36	GC/CG	-3.82
CC/GG	-2.07	-1.93	CC/GG	-2.36

Table 4.1: Nearest-neighbor base-pair free energies. Unified-oligonucleotide (UO) and unzipping (UNZ) set of nearest-neighbor (NN) base-pair free energies $g_{i,i+1}$ for DNA, and UO set for RNA provided by the UO set derived at 1M NaCl ionic conditions at 25℃ [Sant 98, Xia 98, Zuke 03, Hugu 10b].

length equals:

$$L_c^n = (2n + n_{\text{loop}}\delta_{n\text{N}})\, d_b, \tag{4.3}$$

where n_{loop} is the number of base pairs in the loop and d_b equals the average interphosphate distance between consecutive nucleotides. For ssDNA d_b is ∼0.59 nm/base, as measured in Chapter 3 and other studies [Olso 75, Saen 84]. For ssRNA, d_b is usually taken equal to 0.655 nm/base [Murp 04, Biza 12].

In Chapter 3 it was already discussed that the elastic properties of ssDNA are usually modeled according to the WLC model. This model is also useful to extract the equilibrium end-to-end distance $x_n(f)$ of the released single-stranded nucleic acid (both ssDNA and ssRNA) in any of the hairpin's configurations n at force f . Hence:

$$f = \frac{k_{\text{B}}T}{P}\left[\frac{1}{4(1-x_n/L_c^n)^2} - \frac{1}{4} + \frac{x_n}{L_c^n} + \sum_{n=2}^{7} a_n \left(\frac{x_n}{L_c^n}\right)^2\right], \tag{4.4}$$

where k_{B} is the Boltzmann constant, T is the absolute temperature, P is the persistence length, and a_n $(n = 2, \ldots, 7)$ are coefficients that can be taken equal to zero [Bust 94] or to the numerical values proposed in reference [Bouc 99] (Chapter 3).

Finally, the term $\Delta G_n^{\text{st}}(f)$ in Eq. (4.1) is equal to the reversible work needed to stretch the released single-stranded nucleic acid of the configuration n from a random coiled state to its equilibrium end-to-end distance $x_n(f)$ that depends on the applied

force f. Thus:

$$\Delta G_n^{\mathrm{st}}(f) = -\int_0^f x_n(f')df', \tag{4.5}$$

where $x_n(f)$ is the inverse of Eq. (4.4).

Free energy of orientating the double helix diameter

The last term in Eq. (4.1) accounts for the energetic contribution related to the orientation of the hairpin diameter along the force axis. As already explained in Chapter 3, the double helix diameter is modeled as a single bond of length $d = 2$ nm that is oriented due to the action of an external force f [Wood 06b, Forn 11]. According to the partition function Z calculated in the previous chapter (Eq. 3.6), the free energy of orientation is equal to:

$$\Delta G_n^{\mathrm{d}}(f) = -k_{\mathrm{B}}T \log Z \cdot (1 - \delta_{n\mathrm{N}}) \tag{4.6}$$

$$= -k_{\mathrm{B}}T \log \left[\frac{k_{\mathrm{B}}T}{fd} \sinh\left(\frac{fd}{k_{\mathrm{B}}T}\right) \right] \cdot (1 - \delta_{n\mathrm{N}}), \tag{4.7}$$

where the term $(1 - \delta_{n\mathrm{N}})$ is equal to 0 if $n = \mathrm{N}$ and 1 otherwise. Hence, when the hairpin is in the unfolded conformation ($n = \mathrm{N}$) there is no double helix and $\Delta G_n^{\mathrm{d}}(f)$ is zero.

The equilibrium distance projected along the force axis is $x_{\mathrm{d}}(f)$, and it can be evaluated using the FJC model with identical Kuhn and contour lengths (Eq. 3.7b).

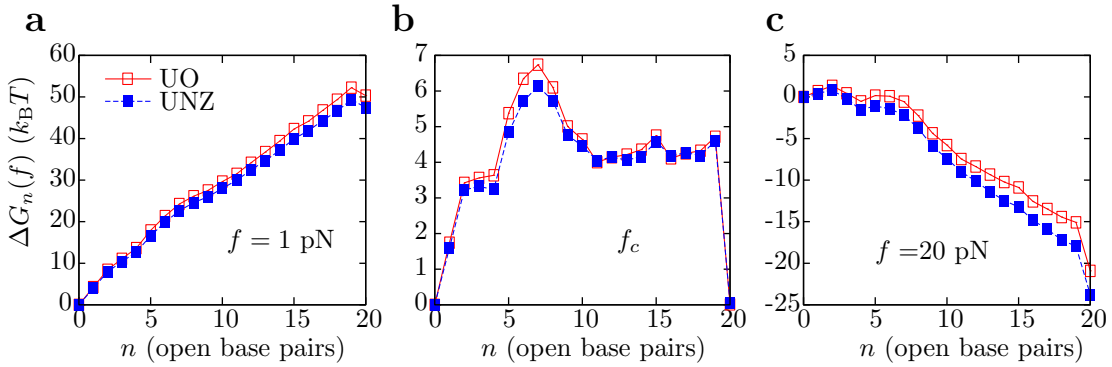

Figure 4.3: Free-energy landscape of a DNA hairpin. FEL evaluated using Eq. (4.1) for the CD4 hairpin (Fig. 4.2a) calculated at 1 pN (**a**), at the coexistence force f_c (**b**), and at 20 pN (**c**). The UO set of base-pair free energies derived in bulk experiments [Sant 98] (open-red squares and solid line, $f_c = 15.3$ pN) and single-molecule UNZ experiments [Hugu 10b] (solid-blue squares and dashed line, $f_c = 14.6$ pN) have been used.

Figure 4.3 represents theoretical profiles of the FEL calculated at different forces using Eq. (4.1) for the DNA hairpin shown in Fig. 4.2a and using the two different sets of base-pair free energies for DNA [Sant 98, Hugu 10b] (Table 4.1). Differences between the two estimations are negligible. The value of the persistence length P used to calculate

the energetic contribution from stretching has been taken equal to 1.3 nm, in accordance to the range of values measured in the previous chapter (Tables 3.1 and 3.2). In panel a the force is set equal to 1 pN, and the most stable state is located at $n = 0$. In panel b the force is chosen so that the two states $n = 0$ and $n = \mathrm{N}$ have the same free energy and therefore are equally populated. This force is usually referred to as the coexistence force, f_c, and depends on the hairpin sequence. The value of f_c slightly depends on the set of base-pair free energies used to estimate the FEL. In the example provided in Fig. 4.3b, $f_c = 15.3$ pN when using the UO set of NN parameters, and $f_c = 14.6$ pN when using the NN parameters derived from UNZ experiments. Finally, in panel c the force is set equal to 20 pN and the most stable state is $n = N$. Here it can be observed how force favors states with larger extensions.

4.1.2 Transition state theory: a brief reminder

Nowadays it is possible to measure the kinetic rates of several biological processes, such as the molecular folding/unfolding or protein–protein and protein–DNA association/dissociation, under different conditions. A proper knowledge of the relation between kinetic rates and the microscopic behavior of the system is therefore crucial for the interpretation of experimental observations. This is precisely the main goal of transition state theory (TST) [Hang 90, Zwan 01, Zhou 10].

The first attempts to understand chemical reactions were made by Jacobus H. van't Hoff and Swante Arrhenius in the late 1880s. Given a chemical reaction where the conversion of reactants (R) to products (P) and vice-versa are in equilibrium:

$$R \rightleftarrows P, \tag{4.8}$$

the equilibrium constant K_{eq} is defined as the ratio between the concentration of products and reactants:

$$K_{\mathrm{eq}} = \frac{[P]_{\mathrm{eq}}}{[R]_{\mathrm{eq}}} \tag{4.9}$$

First, van't Hoff derived in 1884 his well-known equation to explain the temperature dependence of the equilibrium constant K_{eq} of chemical reactions [Vant 84]:

$$\frac{d \log K_{\mathrm{eq}}}{dT} = \frac{\Delta H}{k_{\mathrm{B}} T^2}, \tag{4.10}$$

where ΔH is the change in enthalpy when one reactant molecule is converted into one product molecule. Five years later [Arrh 89], Arrhenius made a step further and considered that, in any chemical reaction, reactant molecules that successfully become products are in an active form R^* that must be in equilibrium with the rest of reactant molecules,

$$R \overset{K_{\mathrm{eq}}}{\rightleftarrows} R^* \longrightarrow P \tag{4.11}$$

Under this scenario, the observed rate constant $k_{\rightarrow}$ of the chemical reaction $R \rightarrow P$ must be proportional to the normal-to-active equilibrium constant, $k_{\rightarrow} \propto K_{\mathrm{eq}}$. Using

the van't Hoff equation (4.10), Arrhenius derived the following equation for the rate constant:

$$\frac{d \log k_{\rightarrow}}{dT} = \frac{\Delta E^{\ddagger}}{k_{\mathrm{B}} T^2}, \tag{4.12}$$

which can be solved to:

$$k_{\rightarrow} = k_0 \exp\left(-\frac{\Delta E^{\ddagger}}{k_{\mathrm{B}} T}\right). \tag{4.13}$$

Here k_0 is a constant of integration usually referred to as a pre-exponential factor, and $\Delta E^{\ddagger}$ is the temperature-independent change in energy between a reactant molecule in its active conformation (R^*) and a reactant molecule in its "passive" state (R). This change in energy is known as the activation energy of a chemical reaction, and intuitively implies that reactants must overcome an energetic barrier in order to become products (Fig. 4.4a). The activated state is viewed as a TS along the course of a chemical reaction.

The Arrhenius equation is widely used in the study of chemical reactions, despite the three facts that (1) it is mainly derived based on empirical observations, (2) it does not explain the physical origin of the pre-exponential factor k_0 neither of the activation energy $\Delta E^{\ddagger}$, and (3) it does not consider the presence of multiple transition states between reactants and products.

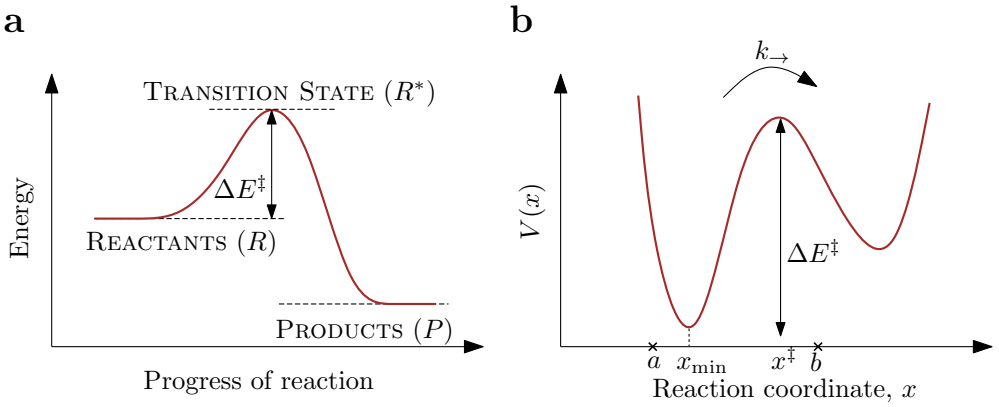

Figure 4.4: Profile of a chemical reaction. a. Arrhenius point of view. Molecules in a chemical reaction exist in the form of reactants (either in the stable or activated state, R and R^*, respectively) or as products (P). **b.** Eyring point of view. Potential of mean force.

One of the most notable contributions to TST was made by Henry Eyring [Eyri 35], and was almost simultaneously derived by Merdith G. Evans and Michael Polanyi [Evan 35]. They used the concept of the potential of mean force $V(x)$ to identify the energetic state of a molecule undergoing a chemical reaction, where x is a reaction coordinate that indicates the degree of advance of the reaction. Figure 4.4b shows a sketch of a potential of mean force, where $x^{\ddagger}$ denotes the position of the TS. Reactants that are located in the left-hand well of $V(x)$ are in thermal equilibrium with reactants located at the top of the energy barrier at $x^{\ddagger}$, according to Arrhenius. Products of the chemical reaction are located in the right-hand well of $V(x)$. Eyring assumed that molecules obey

the Newton's equation of motion and that molecules that cross the energy barrier at $x^{\ddagger}$ immediately become products. Using statistical mechanics, he derived that the kinetic rate $k_{\rightarrow}$ of a chemical reaction satisfies the following relation:

$$k_{\rightarrow} = \frac{k_{\mathrm{B}}T}{hZ_R} \exp\left(-\frac{\Delta E^{\ddagger}}{k_{\mathrm{B}}T}\right), \tag{4.14}$$

where h is the Planck constant and Z_R is the partition function of reactant molecules. A detailed derivation of Eq. (4.14) is found in Appendix C.1. The ideas introduced by Eyring were very important because they endowed the prefactor k_0 with a physical explanation for the first time ($k_0 \sim k_{\mathrm{B}}T/hZ_R$). However, one assumption of Eyring's theory, which is that molecules follow deterministic trajectories according to the Newton's law, is not correct: the movement of molecules in a thermal bath is Brownian.

In 1940, Hendrik A. Kramers considered that the motion of molecules in a heat bath along the reaction coordinate is governed by the Langevin equation [Kram 40]. As trajectories are stochastic, the molecular motion is better described using the probability density function $p(x,t)$, whose time evolution satisfies the Fokker-Planck differential equation:

$$\frac{\partial p(x,t)}{\partial t} = D\frac{\partial}{\partial x}\left[\frac{\partial}{\partial x} + \frac{1}{k_{\mathrm{B}}T}\frac{dV(x)}{dx}\right] p(x,t), \tag{4.15}$$

where $D = k_{\mathrm{B}}T/\gamma$ is the diffusion coefficient and γ is the friction coefficient of a molecule in the thermal bath. It is worth mentioning that Eq. (4.15) is derived for high frictions ($\gamma \rightarrow \infty$), where the inertial term of the Langevin equation can be ignored. Equation (4.15) can be solved and the rate of a chemical reaction satisfies:

$$k_{\rightarrow} = \frac{D}{\int_a^b dy\ \exp\left(\frac{V(y)}{k_{\mathrm{B}}T}\right)\int_a^y dz \exp\left(-\frac{V(z)}{k_{\mathrm{B}}T}\right)}, \tag{4.16}$$

where a is a position along the reaction coordinate found in the potential well of reactants, whereas b is a position found in the potential well of products (Fig. 4.4b). A full derivation of Eq. (4.16) is provided in Appendix C.2. By applying the saddle point approximation to Eq. (4.16) one obtains:

$$k_{\rightarrow} \sim \frac{\sqrt{-V''(x^{\ddagger})V''(x_{\min})}}{2\pi\gamma} e^{-\frac{V(x^{\ddagger})-V(x_{\min})}{k_{\mathrm{B}}T}}. \tag{4.17}$$

Therefore, by identifying the term $V(x^{\ddagger}) - V(x_{\min})$ and $\sqrt{-V''(x^{\ddagger})V''(x_{\min})}/2\pi\gamma$ with $\Delta E^{\ddagger}$ and k_0 in Eq. (4.13), respectively, the Kramers theory provides a new theoretical explanation of the Arrhenius equation derived from a microscopic description of the system.

Equation (4.16) is very powerful. From the sole knowledge of the profile of the potential of mean force, $V(x)$, it makes it possible to derive the behavior of the kinetic rates of chemical reactions for one-dimensional systems. Unfortunately, $V(x)$ is not always known and most of the times several assumptions must be done.

Finally, from the combination of Eqs. (4.13) and (4.16), an effective barrier B_{eff} for the system can be defined as:

$$B_{\text{eff}} = k_{\text{B}}T \log\left(\frac{k_0}{D}\int_a^b dy\; e^{\frac{V(y)}{k_{\text{B}}T}}\int_a^y dz e^{-\frac{V(z)}{k_{\text{B}}T}}\right). \tag{4.18}$$

This barrier provides a measurement of the energy of the TS along a reaction pathway from the profile of the potential energy $V(x)$. It is also interpreted as the size of the kinetic barrier that reactants R must overcome in order to become products P. Moreover, it can be shown that (Appendix C.3):

$$\frac{\partial B_{\text{eff}}}{\partial f} \simeq -\left(x^{\ddagger} - x_{\text{min}}\right) = -\Delta x_{\text{eff}}, \tag{4.19}$$

where $f = -\partial V(x)/\partial x$. $\Delta x_{\text{eff}} = x^{\ddagger} - x_{\text{min}}$ is the distance between the position of the transition state, $x^{\ddagger}$, where reactants are active (R^*), and the position of the minimum in the potential well for passive reactants, x_{min}, along the reaction coordinate.

It is important to emphasize that the last results of this section were derived using the potential of mean force $V(x)$. This potential gives the average energy over all the configurations of a system that are compatible to a given value of the reaction coordinate x. Therefore, for multi-dimensional systems Eqs. (4.13), (4.14) and (4.16) are only valid when the FEL can be projected along a single reaction coordinate. At the present time, TST still is a subject under development. In fact, the relation between the kinetic rates and multi-dimensional systems is nowadays a complex problem being studied by theoretical physicists [Hang 90]. In the scope of this thesis mostly one-dimensional FEL are studied, and therefore, Kramers theory is extremely helpful to describe the kinetic behavior of the molecules under consideration.

4.2 Force-spectroscopy experiments

In single-molecule experiments carried out with OT a mechanical force is applied to the ends of a molecule and its response is measured through changes in molecular extension. Pulling experiments, briefly explained in Section 2.3.2 and in Chapter 3, are widely used to investigate the mechanical behavior of molecules under non-equilibrium conditions. In such experiments, the force applied varies with time at a constant loading rate and the molecule stochastically undergoes transitions over different conformations at different forces. The study of breakage forces with these kind of experiments is known as dynamic force spectroscopy (DFS).

Figure 4.5a shows some FDC measured when pulling CD4 (Fig. 4.2a) at two different loading rates (3.8 and 12 pN/s). Unfolding trajectories are in red (slow rate) and purple (fast rate); folding trajectories are in blue (slow) and cyan (fast). CD4 behaves as a two-state system: the hairpin is either folded in its native conformation (state N, $n = 0$ in terms of open base pairs) or in the unfolded state (state U, n = N). Transitions between N and U are detected through a sudden jump in force, whose size is related to

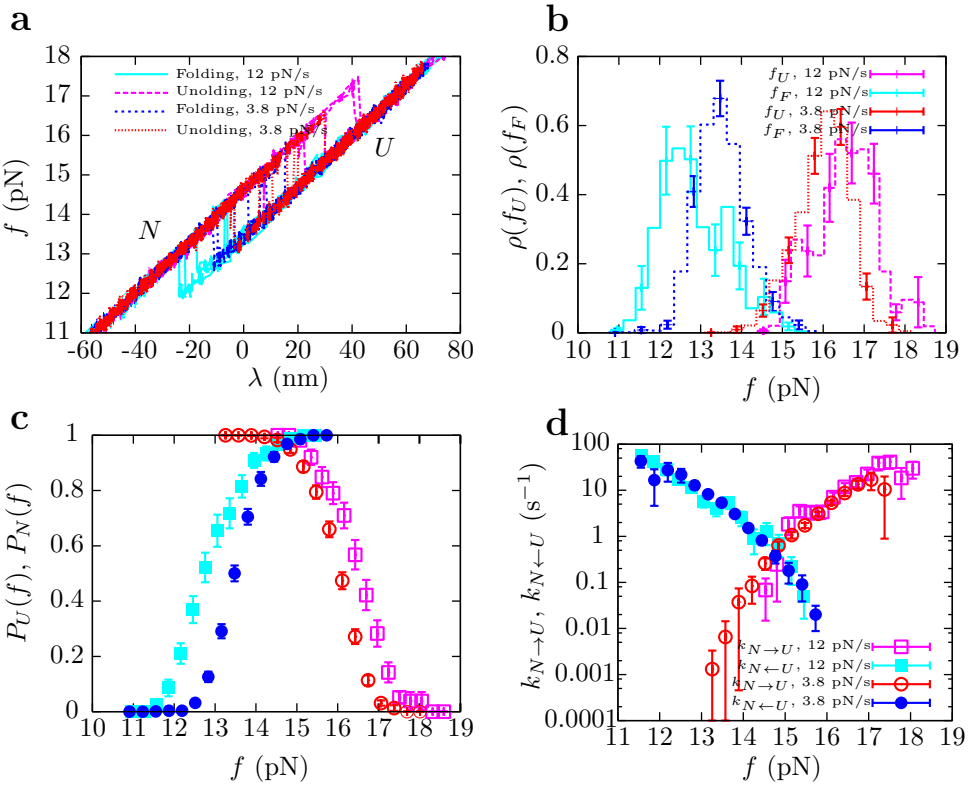

Figure 4.5: Pulling experiments and force spectroscopy analysis for two-state DNA hairpin. a. Force-distance curves (FDC) measured when pulling CD4 (Fig. 4.2a) at 3.8 and 12 pN/s (or, equivalently, 60 and 180 nm/s). Unfolding processes are plotted in red (slow) and purple (fast), and folding processes are plotted in blue (slow) and cyan (fast). Unfolding and folding force histograms (**b**), survival probabilities (**c**) and kinetic rates (**d**) measured for CD4. Error bars are standard errors and have been measured using the Bootstrap method. Color code for panels b,c and d as in a.

the extension of the released/absorbed ssDNA due to the sudden disruption/formation of the base pairs in the stem. This makes the bead to fall towards/move away from the center of the optical trap (Chapter 3). The two-state behavior was also predicted by the FEL represented in Fig. 4.3, as no relevant intermediate states are there observed.

In this section it is explained how rupture forces measured in pulling experiments can be analyzed in order to extract the maximum amount of information related to the kinetic and thermodynamic stability of nucleic acid hairpins. Key concepts such as survival probabilities, unfolding and folding kinetic rates, or the definition of the kinetic barrier for nucleic acid hairpins, are explained from a theoretical point of view and are experimentally measured.

Hereafter, forces at which the hairpin unfolds and folds are referred to as unfolding (f_U) and folding (f_F) forces, respectively[1]. The unfolding force f_U is the value of the force in the folded-force branch when the hairpin unfolds, while the folding force f_F is the value

[1]Note that the nomenclature is different from Chapter 3.

of the force in the unfolded-force branch when the hairpin folds. As mentioned before, they are stochastic and vary in each independent repetition of an unfolding or a folding process. Figure 4.5b shows the unfolding and folding force histograms, $\rho(f_U)$ and $\rho(f_F)$, measured when pulling CD4 at two different loading rates. It can be easily observed that hysteresis effects increase with pulling speed. Noticeably, the force at which unfolding and folding force histograms cross each other does not significantly change. It is worth emphasizing that such force is approximately equal to the theoretical coexistence force of CD4, $f_c \sim 14.6 - 15.3$ pN (Fig. 4.2b), determined using the NN model and different sets of base-pair free energies (Eq. 4.1 and Table 4.1). This result hints the power of knowing the profile of the FEL: it is possible to predict from a theoretical background that CD4 behaves as a two-state system whose coexistence force is around 15 pN.

The survival probability $P_{N(U)}(f)$ of states N (U) along the unfolding (folding) process is defined as the probability of the hairpin to remain at state N (U) at force f. These probabilities can be computed from a collection of unfolding (folding) forces values f_U (f_F), measured over repeated pulling cycles by taking:

$$P_N(f) = 1 - \frac{N(f_U < f)}{N_U} \tag{4.20a}$$

$$P_U(f) = 1 - \frac{N(f_F > f)}{N_F}, \tag{4.20b}$$

being N_U (N_F) the total number of unfolding (folding) trajectories and $N(f_U < f)$ ($N(f_F > f)$) the number of f_U (f_F) values that are smaller (larger) than the given force f. Alternatively, survival probabilities can be computed from the unfolding and folding rupture force histograms $\rho(f_U)$ and $\rho(f_F)$, according to:

$$P_N(f) = 1 - \int_0^f \rho(f_U) df_U \tag{4.21a}$$

$$P_U(f) = + \int_0^f \rho(f_F) df_F. \tag{4.21b}$$

The experimental profile of these survival probabilities for CD4 extracted from pulling experiments carried out at two different loading rates are plotted in Fig. 4.5c.

Under a pulling protocol, both survival probabilities $P_N(f)$ and $P_U(f)$ are described by a first-order Markov process,

$$\frac{dP_N(f)}{df} = -\frac{k_{N\to U}(f)}{r} P_N(f) \quad \text{(unfolding)} \tag{4.22a}$$

$$\frac{dP_U(f)}{df} = +\frac{k_{N\leftarrow U}(f)}{r} P_U(f) \quad \text{(folding)}, \tag{4.22b}$$

where r is the loading rate and $k_{N\to U}(f)$ and $k_{N\leftarrow U}(f)$ are the force-dependent kinetic rates of unfolding and folding of the hairpin. Therefore, the kinetic rates $k_{N\rightleftarrows U}(f)$ can

be extracted from the survival probabilities according to:

$$k_{N\to U}(f) = -r\frac{1}{P_N(f)}\frac{dP_N(f)}{df} \quad \text{(unfolding)} \tag{4.23a}$$

$$k_{N\leftarrow U}(f) = +r\frac{1}{P_U(f)}\frac{dP_U(f)}{df} \quad \text{(folding)}. \tag{4.23b}$$

Figure 4.5d shows results obtained from pulling experiments for CD4. As expected, $k_{N\to U}(f)$ increases with force whereas $k_{N\leftarrow U}(f)$ decreases. Note that results measured for the two different loading rates, $r = 3.8$ and 12 pN/s, collapse into a single master curve. This is a very general result: different experimental protocols also collapse into the same master curve.

4.3 Experimental measurement of the free-energy landscape

In Section 4.1.2 it was shown that there is an intimate and univocal relation between the kinetic rates and the FEL (or potential of mean force). Once the kinetic rates are experimentally measured, TST provides the tools to understand the microscopic behavior of the system and to obtain key features of the FEL. Moreover, the kinetic rates $k_{N\rightleftarrows U}(f)$, the survival probabilities $P_{N(U)}(f)$ and the rupture force histograms $\rho(f_{U(F)})$ are mathematically related through Eqs. (4.21), (4.22) and (4.23). Therefore, identical information can be extracted from any of the three different quantities measured from unfolding and folding rupture forces.

In this section, the state of the art to interpret results from DFS studies is presented. Widely accepted theories are summarized, and the standard procedure used in this thesis to analyze experimental data is explained in detail.

4.3.1 Bell-Evans approach

To interpret the results from pulling experiments, the effect of force must be included to model the kinetic rates. A first approach by Evan Evans, based on the Bell model [Bell 78], considers that the molecular FEL for two-state systems is made of a single kinetic barrier that linearly decreases with the force as $B_{\mathrm{BE}} - f\Delta x_{\mathrm{BE}}$. The term B_{BE} is the height of the kinetic barrier at zero force and Δx_{BE} is the distance between the TS and the state prior to the transition [Evan 97, Merk 99, Evan 01] (Fig. 4.6). According to this so-called Bell-Evans (BE) model, the term $\Delta E^{\ddagger}$ in Eq. (4.13) is directly identified with this force-dependent barrier, $B_{\mathrm{BE}} - f\Delta x_{\mathrm{BE}}$, and the kinetic rate can be written as:

$$k_{\to}(f) = k_0 \exp\left(-\frac{B_{\mathrm{BE}} - f\Delta x_{\mathrm{BE}}}{k_{\mathrm{B}}T}\right), \tag{4.24}$$

where k_0 is the pre-exponential factor.

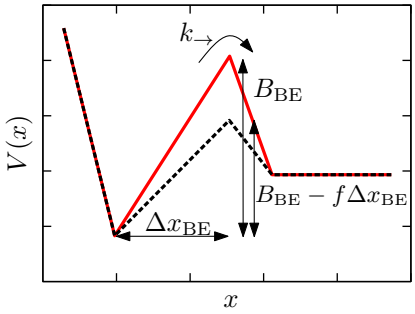

Figure 4.6: Bell-Evans model. The molecular FEL for two-state systems at zero force is made of a single kinetic barrier whose size does not depend on force, B_{BE} (continuous red line). The distance between the stable state and the transition state is labeled by Δx_{BE}. Under the action of a force f, the profile of the FEL is modified according to $V(x) - f\Delta x_{\mathrm{BE}}$ (dashed black line).

The BE model was initially used to unravel the energy landscape of receptor–ligand bonds [Merk 99] (Appendix C.4), but it has also been applied to characterize the mechanical folding and unfolding of DNA hairpins [Wood 06b, Moss 09b] (Appendix D). In this latter context, the unfolding and folding kinetic rates are written as:

$$\begin{aligned} k_{N\to U}(f) &= k_0 \exp\left(-\frac{B_{NU} - f x_{N-\mathrm{TS}}}{k_{\mathrm{B}}T}\right) \\ &= k_m \exp\left(\frac{f x_{N-\mathrm{TS}}}{k_{\mathrm{B}}T}\right) \end{aligned} \tag{4.25a}$$

$$\begin{aligned} k_{N\leftarrow U}(f) &= k_0 \exp\left(-\frac{B_{NU} - \Delta G_{NU} + f x_{\mathrm{TS}-U}}{k_{\mathrm{B}}T}\right) \\ &= k_m \exp\left(\frac{\Delta G_{NU} - f x_{\mathrm{TS}-U}}{k_{\mathrm{B}}T}\right), \end{aligned} \tag{4.25b}$$

where the kinetic barrier B_{BE} is here written as B_{NU} to emphasize that it is located between states N and U; $k_m = k_0 \exp\left(-B_{NU}/k_{\mathrm{B}}T\right)$ is the unfolding kinetic rate at zero force; $x_{N-\mathrm{TS}} = x_{\mathrm{TS}} - x_N$ and $x_{\mathrm{TS}-U} = x_U - x_{\mathrm{TS}}$ are the relative distances between state N and the TS and state U and the TS at the coexistence force f_c, respectively, and their sum should be equal to the molecular extension $x_{NU} = x_{N-\mathrm{TS}} + x_{\mathrm{TS}-U} = x_U - x_N$; and $\Delta G_{NU} = f_c x_{NU}$ is related to the free-energy difference between states N and U. Note that with these definitions, it is satisfied that $k_{N\to U}(f_c) = k_{N\leftarrow U}(f_c)$ (see Appendix D for more information about the BE model applied to nucleic acid hairpins).

According to Eqs. (4.25a) and (4.25b) the logarithm of the unfolding and folding kinetic rates versus force, $\log k_{N\to U}(f)$ and $\log k_{N\leftarrow U}(f)$, can be fitted to straight lines:

$$\log k_{N\to U}(f) = a_U + f \cdot b_U, \qquad \log k_{N\leftarrow U}(f) = a_F - f \cdot b_F, \tag{4.26}$$

where:

$$k_m = \exp(a_U), \quad \Delta G_{NU} = (a_F - a_U) k_{\mathrm{B}}T, \quad x_{N-\mathrm{TS}} = b_U k_{\mathrm{B}}T, \quad x_{\mathrm{TS}-U} = b_F k_{\mathrm{B}}T. \tag{4.27}$$

Figure 4.7a shows the results of the linear fits to the logarithm of the kinetic rates shown in Fig. 4.5d (continuous-black lines). Numerical values are summarized in Table 4.2. It can be seen that reasonable agreement is found between the estimations of $x_{N-\mathrm{TS}}$ and $f_c = \Delta G_{NU}/(x_{N-\mathrm{TS}} + x_{\mathrm{TS}-U})$ with their theoretical predictions. However, the

values obtained for $x_{\mathrm{TS}-U}$ and ΔG_{NU} are lower than expected. The reason for these discrepancies are explained below.

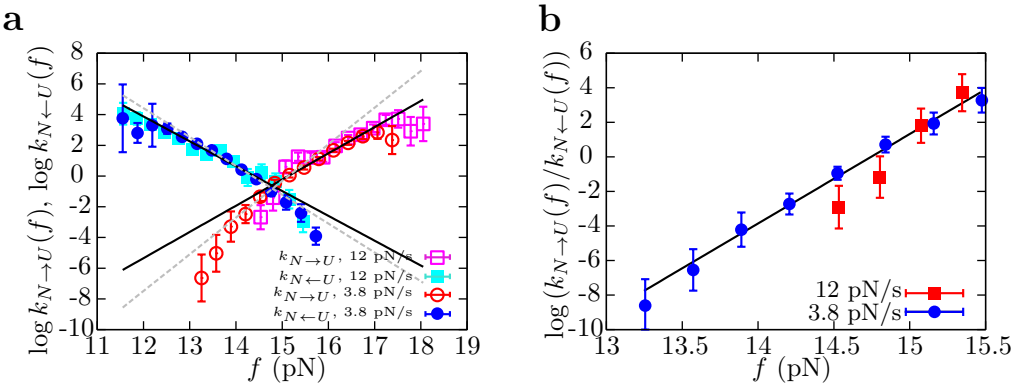

Figure 4.7: Kinetic rates and BE model. **a.** Linear fit of the logarithm of the experimental unfolding and folding kinetic rates to the BE model. Continuous-black lines show results from the fits performed on the whole interval of forces. Dashed-gray lines are the fits performed using only in the force interval $f \in [13-15.5]$ pN, where detailed balance can be experimentally verified. **b.** Experimental verification of detailed balance.

	$x_{N-\mathrm{TS}}$ (nm)	$x_{\mathrm{TS}-U}$ (nm)	ΔG_{NU} ($k_{\mathrm{B}}T$)	f_c (pN)
Experimental	7.0±0.4	6.7±0.3	49±2	15±1
Theoretical	7.2±0.2	10.8±0.2	65±2	15.0±0.3

Table 4.2: Fit of the kinetic rates for CD4 to the Bell-Evans model. The logarithm of the experimentally measured kinetic rates were fitted to straight lines (Eq. 4.26). Results are: $a_F = 23 \pm 1$, $a_U = -26 \pm 1$, $b_F = 1.65 \pm 0.06$ pN^{-1}, $b_U = 1.7 \pm 0.1$ pN^{-1}. Numerical values for $x_{N-\mathrm{TS}}$, $x_{\mathrm{TS}-U}$ and ΔG_{NU} are found using definitions given in Eq. (4.27). f_c is defined as the ratio $\Delta G_{NU}/(x_{N-\mathrm{TS}}+x_{\mathrm{TS}-U})$. Error bars for the experimental estimations are calculated by propagation of standard errors from the fit. Theoretical values are the average between the two theoretical predictions obtained using UO or the UNZ set of base-pair free energies [Sant 98, Hugu 10b] (Table 4.1). Error bars are statistical errors.

Limitations of the Bell-Evans model

According to Eqs. (4.25a) and (4.25b) the logarithm of the unfolding and folding kinetic rates, $\log k_{N\to U}(f)$ and $\log k_{N\leftarrow U}(f)$, should be linear with force (Eqs. 4.26 and 4.27). However, a curvature (and therefore a deviation from the linear fit) is observed for both $\log k_{N\to U}(f)$ and $\log k_{N\leftarrow U}(f)$ at extreme values of force in Fig. 4.7a. This suggests that the BE model is not accurate enough.

In Appendix D it is shown that the simplified kinetic rates provided by the BE model (Eq. 4.25) are only valid in a close vicinity of the coexistence force f_c, where it is satisfied that:

$$\log\left(\frac{k_{N\to U}(f)}{k_{N\leftarrow U}(f)}\right) = (f - f_c)\frac{x_{NU}}{k_{\mathrm{B}}T}, \tag{4.28}$$

since $\Delta G_{NU} = f_c x_{NU}$ to a first order. An experimental verification of Eq. (4.28) is shown for CD4 in Fig. 4.7b, where a linear dependence is recovered for $\log\left(k_{N\to U}(f)/k_{N\leftarrow U}(f)\right)$ plotted against f with slope $x_{NU} = 20 \pm 1$ nm. This value of x_{NU} is in reasonable agreement with theoretical predictions, which give $x_{NU} \sim 18$ nm. The verification can only be performed in the force range $f \in [13-15.5]$ pN, where the ratio $\log\left(k_{N\to U}(f)/k_{N\leftarrow U}(f)\right)$ can be computed from experimental data. For a wider force range it cannot be assumed that the simplified version of the BE kinetic rates are valid. Results summarized in Table 4.2 do not take into account this limitation and are obtained by fitting the experimental kinetic rates along the entire range of forces, from 11 to 19 pN.

Therefore, there are two main reasons that prevent the use of the simplified version of the BE kinetic rates in a large force interval (as it was done to obtain the results summarized in Table 4.2). First, the first order Taylor series is no longer valid for large force intervals (Appendix D). Second and most important, the Leffler-Hammond postulate predicts that as the force applied to the ends of a molecule increases, the TS approaches state U [Leff 53, Hamm 55]. As a consequence, the distances $x_{N-\mathrm{TS}}$ and $x_{\mathrm{TS}-U}$ do actually depend on force, not only because of the elastic contributions but also because the TS changes its position along the reaction coordinate. The use of the BE model to analyze the unfolding and folding kinetic rates of nucleic acid hairpins provides, to a first degree, estimators for $x_{N-\mathrm{TS}}$ and $x_{\mathrm{TS}-U}$ evaluated close to the coexistence force f_c. However, due to the Leffler-Hammond effect both distances tend to be underestimated, as seen in Table 4.2.

To improve the estimations provided by the BE model, the linear fit of the logarithm of the ratio of the kinetic rates versus force (Eq. 4.26) can be performed in the range of forces between 13.0 and 15.5 pN, where Eq. (4.28) is experimentally validated (Fig. 4.7b). Results are shown in Fig. 4.7a (gray-dashed lines) and summarized in Table 4.3. It can be appreciated that the agreement between experimental and theoretical values is much better than before.

	$x_{N-\mathrm{TS}}$ (nm)	$x_{\mathrm{TS}-U}$ (nm)	ΔG_{NU} ($k_{\mathrm{B}}T$)	f_c (pN)
Experimental	10±1	8±1	65±6	15±2
Theoretical	7.2±0.2	10.8±0.2	65±2	15.0±0.3

Table 4.3: Fit of the kinetic rates for CD4 to the Bell-Evans model in the force interval $f \in [13, 15.5]$. The logarithm of the experimentally measured kinetic rates were fitted to straight lines (Eq. 4.26). Results are: $a_F = 28 \pm 2$, $a_U = -36 \pm 4$, $b_F = 2.0 \pm 0.1$ pN^{-1}, $b_U = 2.4 \pm 0.3$ pN^{-1}. Numerical values for $x_{N-\mathrm{TS}}$, $x_{\mathrm{TS}-U}$ and ΔG_{NU} are found using definitions given in Eq. (4.27). f_c is defined as the ratio $\Delta G_{NU}/(x_{N-\mathrm{TS}} + x_{\mathrm{TS}-U})$. Error bars for the experimental estimations are calculated by propagation of standard errors from the fit. Theoretical values are the same as in Table 4.2.

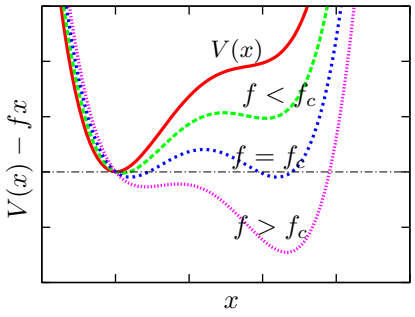

Figure 4.8: Generalization of the BE model. Under the action of a mechanical force, the molecular potential of mean force (or FEL) at zero force $V(x)$ is modified the quantity $-fx$. As a consequence, configurations with larger values of x become favored by force [Tino 02].

4.3.2 Kramers-based models

A more rigorous procedure is to directly introduce the effect of applied force into the Kramers result (Eq. 4.16):

$$k_{\rightarrow}(f) = \frac{D}{\int_a^b dy \ \exp\left(\frac{V(y,f)}{k_B T}\right) \int_a^y dz \exp\left(-\frac{V(z,f)}{k_B T}\right)}, \tag{4.29}$$

where $V(x, f)$ is the molecular potential of mean force as a function of the reaction coordinate x and the applied force f.

In order to model $V(x, f)$ the same lines followed by Evans in the late 90's are usually kept [Evan 97, Merk 99, Evan 01]. The simplest generalization to the BE model is to consider a force-independent potential $V(x)$ that under an external force is deformed with the linear potential $-fx$ (Fig. 4.8):

$$V(x, f) = V(x) - fx \tag{4.30}$$

Using deconvolution methods it is possible to find the profile of $V(x)$ that, according to Eqs. (4.29) and (4.30), best fits the experimentally measured rupture force histograms [Schl 06]. Alternatively, analytical expressions for $V(x)$ such as:

$$V(x) = \Delta G^{\ddagger} \left(\frac{x}{x^{\ddagger}}\right)^2, \qquad V(x) = \frac{3}{2}\Delta G^{\ddagger} \left(\frac{x}{x^{\ddagger}}\right) - 2\Delta G^{\ddagger} \left(\frac{x}{x^{\ddagger}}\right)^3, \tag{4.31a,b}$$

being $\Delta G^{\ddagger}$ and $x^{\ddagger}$ free parameters related to the energy and position of the TS along the FEL, are used to derive analytical formulas for the rupture force histograms and the kinetic rates that can be subsequently fitted to experimental data in order to get a picture of the profile of the FEL [Dudk 06, Dudk 08] (see Appendix C.5 for more information). Both procedures are very useful to obtain key features of the FEL in the absence of theoretical approaches, as is the case for proteins or intermolecular bonds (Chapters 6 and 7).

In the case of nucleic acid hairpins, the term $V(x, f)$ in Eq. (4.29) can be identified with the hairpin energy landscape, $\Delta G_n(f)$ (Eq. 4.1). Because the reaction coordinate in this case is discrete (number of open base pairs), the double integral that appears in the Kramers result for the kinetic rate (Eq. 4.16) must become a double sum. Consequently,

a discretized version of the effective kinetic barrier (Eq. 4.18) is introduced and hereafter denoted by $B_{NU}(f)$:

$$B_{NU}(f) = k_{\mathrm{B}}T \log\left(\sum_{n=0}^{\mathrm{N}}\sum_{n'=0}^{n} e^{\frac{\Delta G_n(f)-\Delta G_{n'}(f)}{k_{\mathrm{B}}T}}\right), \tag{4.32}$$

where the double sum runs through all the different configurations of the hairpin, labeled by n and being N the number of total base pairs along the stem. Due to the discretization procedure, an extra constant term approximately equal to $\log\left(4k_0 d_b^2/D\right)$ should be added in Eq. (4.32). Nevertheless, it will be seen in future sections that $k_0 \sim 10^4$ s^{-1} for DNA hairpins. In addition it has been measured in reference [Sola 14] that $D \sim 10^{-14}$ m^2/s for α-Synuclein, which is a small intrinsically disordered protein. With this numerical values, it can be observed that this extra term goes to zero and therefore does not play any relevant role in Eq. (4.32).

In Fig. 4.9a the theoretical profile of $B_{NU}(f)$ for CD4 (hairpin sketched in Fig. 4.2a) is plotted, predicted by using Eq. (4.32) and the NN model with the two different sets of base-pair free energies for DNA hairpins (Table 4.1). Differences between the two estimations are very small.

Remarkably, $B_{NU}(f)$ can be also extracted from unfolding and folding kinetic rates. By writing $k_{N\rightleftarrows U}(f)$ in terms of the force-dependent effective barrier $B_{NU}(f)$ that mediates the transition between states N and U, we get:

$$k_{N\to U}(f) = k_0^{NU} \exp\left(-\frac{B_{NU}(f)}{k_{\mathrm{B}}T}\right) \tag{4.33a}$$

$$k_{N\leftarrow U}(f) = k_0^{NU} \exp\left(-\frac{B_{NU}(f) - \Delta G_{NU}(f)}{k_{\mathrm{B}}T}\right). \tag{4.33b}$$

Here, $\Delta G_{NU}(f) = G_U(f) - G_N(f) = \Delta G_{n=\mathrm{N}}(f) - \Delta G_{n=0}(f)$ is equal to the difference in free energy between states U and N ($n = \mathrm{N}$ and $n = 0$ respectively) at force f, and k_0^{NU} is the attempt frequency at zero force. By applying logarithms to both sides of Eq. (4.33) one gets:

$$\frac{B_{NU}(f)}{k_{\mathrm{B}}T} = \log k_0^{NU} - \log k_{N\to U}(f) \tag{4.34a}$$

$$\begin{aligned}\frac{B_{NU}(f)}{k_{\mathrm{B}}T} &= \log k_0^{NU} - \log k_{N\leftarrow U}(f) + \frac{\Delta G_{NU}(f)}{k_{\mathrm{B}}T} \\ &= \log k_0^{NU} - \log k_{N\leftarrow U}(f) + \frac{\Delta G^0_{NU} - \int_0^f x_{\mathrm{N}}(f')df' + \int_0^f x_{\mathrm{d}}(f')df'}{k_{\mathrm{B}}T},\end{aligned} \tag{4.34b}$$

where Eqs. (4.1)–(4.6) are used, $\Delta G^0_{NU} = G^0_U - G^0_N = \Delta G^0_{n=\mathrm{N}} - \Delta G^0_{n=0}$ is the free energy of formation at zero force, $x_{\mathrm{N}}(f)$ is the equilibrium end-to-end distance of the hairpin in the U state at force f, and $x_{\mathrm{d}}(f)$ is the extension of the hairpin diameter projected along the direction of applied force. Both $x_{\mathrm{N}}(f)$ and $x_{\mathrm{d}}(f)$ can be computed using the corresponding elastic models with the appropriate elastic parameters (Chapter 3).

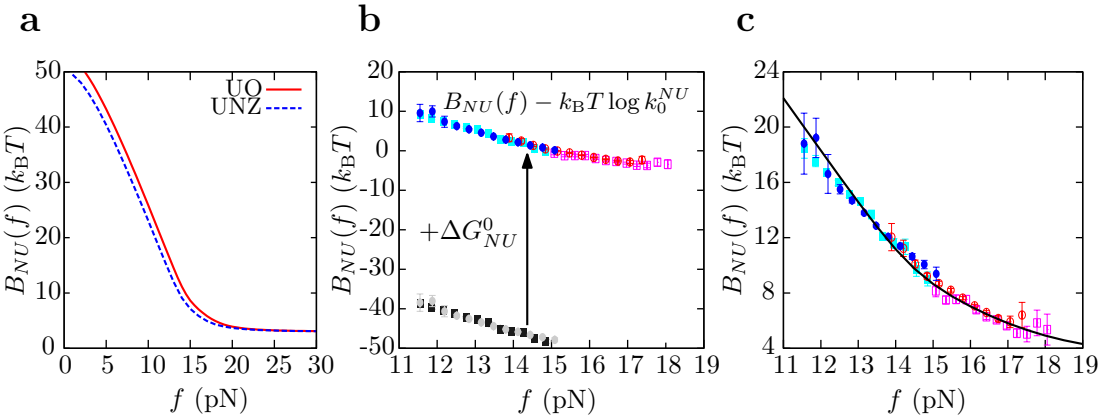

Figure 4.9: Force-dependent kinetic barrier $B_{NU}(f)$. a. Theoretical prediction of $B_{NU}(f)$ for hairpin CD4 using the NN model with the UO or the UNZ set of base-pair free energies (Table 4.1) [Sant 98, Hugu 10b]. **b.** Measurement of ΔG^0_{NU}. From a collection of f_U values $B_{NU}(f)/k_BT - \log k_0^{NU} = -\log k_{N\to U}(f)$ is extracted (red-empty circles and purple-empty squares) and from a collection of f_F values $B_{NU}(f)/k_BT - \log k_0^{NU} - \Delta G^0_{NU}/k_BT = -\log k_{N\leftarrow U}(f) - \int_0^f x_N(f')df'/k_BT + \int_0^f x_0^d(f')df'/k_BT$ is obtained (black-solid squares and gray-solid circles). An experimental estimation of the profile of $B_{NU}(f) - \log k_0^{NU}$ is obtained by adding $\Delta G^0_{NU}/k_BT = 49 \pm 1$ to the latter estimation (cyan-solid squares and blue-solid circles). **c.** From the matching between experimental and theoretical estimations of $B_{NU}(f)$ an estimation of $\log k_0^{NU}$ (equal to 9.0 ± 0.4 in this case) is obtained.

Therefore, from unfolding and folding rupture forces two experimental estimations of $B_{NU}(f)$ are obtained. On the one hand, from unfolding rupture forces it is straightforward to obtain $-\log k_{N\to U}(f)$. From Eq. (4.34a) it can be seen that this is equal to $B_{NU}(f)$ except for the logarithm of the attempt frequency at zero force, $\log k_0^{NU}$, which remains as an unknown constant. On the other hand, from folding rupture forces it is straightforward to measure $-\log k_{N\leftarrow U}(f)$, to which the elastic contributions from stretching the released single-stranded nucleic acid of state U and from orienting the hairpin double helix diameter at force f are subtracted and added respectively. This provides the second estimation of $B_{NU}(f)$ except for the logarithm of the attempt frequency at zero force, $\log k_0^{NU}$, and the free energy of formation of the hairpin, ΔG^0_{NU} (Eq. 4.34b) [Mano 06].

ΔG^0_{NU} can be measured by imposing analytical continuity of the two estimators of $B_{NU}(f)$, *i. e.*, by making the two kinetic barriers extracted from unfolding and folding rupture forces, Eqs. (4.34a) and (4.34b) respectively, to follow the same force curve. The case of CD4 is depicted in Fig. 4.9b, where $\Delta G^0_{NU} = 49 \pm 1\ k_BT$ is obtained. This value is in very good agreement with the free energy of formation predicted by the NN model using the UO set of parameters derived in bulk experiments (51 k_BT) [Sant 98, Zuke 03] or derived from UNZ experiments (48 k_BT) [Hugu 10b].

In order to obtain an estimation for $\log k_0^{NU}$, the experimentally measured $B_{NU}(f)$ (Fig. 4.9b) is fitted to the theoretical prediction obtained using Eq. (4.32) (Fig. 4.9a). The result is shown in Fig. 4.9c and a remarkable good agreement is obtained between the theoretical and experimental profiles of $B_{NU}(f)$. The value of $\log k_0^{NU}$ that best fits

our data is 9.0 ± 0.4, which corresponds to $k_0^{NU} \simeq (8 \pm 3) \times 10^4$ s^{-1}. This is in the same order of magnitude than previous estimations of the attempt frequency at zero force obtained for nucleic acid hairpins [Mano 06, Enge 11, Biza 12].

Still one question remains: which is the effect of the whole molecular construct into the behavior of the kinetic barrier? It is known from previous studies that the variation in the kinetics due to the size of the DNA handles is small compared to the intrinsic properties of the nucleic acid under study [Wood 06b, Wood 06a, Wen 07, Mano 07b, Forn 11]. The size of the DNA handles and the stiffness of the optical trap may modify the value of the kinetic rates up to 40%. However, coexistence forces have been shown to be independent of changes in the experimental setup. Therefore, the estimated value for the attempt frequency k_0^{NU} obtained using the kinetic analysis presented in this section may change with trap stiffness or handles length, whereas ΔG^0_{NU} is expected to remain constant.

4.4 Conclusions

In this chapter, it is addressed how the concept of the molecular FEL is essential to interpret single molecule DFS experiments. A short DNA hairpin (called CD4, Fig. 4.2a) is used to test the main ideas. The advantage of working with such molecule is twofold: (1) it is possible to calculate the profile of the theoretical FEL of nucleic acid hairpins at any value of the force using the NN model, (2) in DFS experiments, short hairpins behave as two-state systems and transitions between the two states are easily identified. Hence, the appeal of using nucleic acid hairpins as model systems to study the connection between experimental measurements of kinetic rates and the molecular FEL is clear.

In this direction, important concepts from TST have been reviewed, such as the notion of the reaction coordinate, the transition state (or activated state), the kinetic rate, or the kinetic barrier (or activation energy). Kramers theory becomes an excellent framework to model the behavior of molecules when a force is applied, since it establishes straightforward relations between the FEL of unidimensional systems and kinetic rates or kinetic barriers.

On the other hand, it is explained in great detail how non-equilibrium DFS experiments with two-state molecules can be analyzed to extract, from the unfolding and folding forces, the survival probabilities of states N and U along the experimental protocol. The relationship between such survival probabilities and unfolding and folding kinetic rates is immediate by using Markov models.

In the literature there are many approaches to relate experimental kinetic rates with key features of the molecular FEL. One of the most commonly used to study dissociation of molecular bonds is the BE model, which provides estimations for the position of the kinetic barrier along the reaction coordinate (Chapter 6). When properly adapted to study nucleic acid hairpins, the BE model can also be used to estimate not only the position of the kinetic barrier relative to states N and U, but also the free-energy difference between the two states. In this chapter, limitations of the use of the BE model in the study of molecular unfolding and folding and the importance of the

Leffler-Hammond postulate have been discussed in detail. Next, the effective barrier of a discrete molecular FEL is defined and studied using a Kramers-based approach. The key idea is to compare experimental and theoretical estimations of the profile of force-dependent kinetic barrier, $B_{NU}(f)$, to get free-energy differences and estimations of the attempt frequency at zero force. Most of the results presented in this subject were previously introduced by M. Manosas, D. Collin and F. Ritort and can be found in reference [Mano 06]. This method is powerful, since it allows to follow the dependence with force of the height and position and the molecular FEL for two-state systems and to directly observe the Leffler-Hammond effect.

This chapter serves as pedagogical introduction to study molecular systems using DFS experiments. In Chapter 5, it will be shown how to extend the Kramers-based analysis of DFS experiments in order to measure the kinetic and thermodynamic behavior of several nucleic acid hairpins under different conditions and with different structural properties. For instance, the effect of the position of the kinetic barrier along the hairpin stem, the presence of intermediate states and even the binding properties of an intercalating peptide will be characterized. In Chapter 7 this analysis will be applied to study the unfolding kinetics of Barnase, a globular protein. In this case a theoretical prediction for $B_{NU}(f)$ is absent but the method is still useful to estimate the free energy of formation of the protein and a lower bound of its attempt frequency.

Chapter 5

Thermodynamics and kinetics of nucleic acid hairpins

One of the most surprising discoveries of the Human Genome Project was that the number of human protein-coding genes is significantly lower than previously estimated [HGP 14]. From initial predictions of 100×10^3 or more genes, the number went down to 20×10^3-25×10^3 with the improvement of sequencing and gene finding tools. Nowadays, the fraction of non-coding DNA is known to be different among species, and in the specific case of humans it is believed to be around 80%.

Non-coding DNA was first presumed to contain information to synthesize proteins required by ancient organisms, and hence to be useless in the present day. However, since the early 1980s a growing number of biochemical functions (such as coding, decoding, regulation and expression of genes) have been associated to small non-coding RNA molecules. The mechanisms of action of these small RNA molecules are very diverse. To mention a couple, in the case of riboswitches both ligand binding and structural changes play an important role to confer the molecule with the ability to regulate gene transcription or translation. A deeper understanding of the dynamics of such mechanisms involving nucleic acid molecules is a critical step towards the elucidation of many biological processes and the interpretation of the full genome. Therefore, it is of high interest to have accurate tools to investigate the thermodynamics and kinetics of small nucleic acid hairpins, which is a goal in this chapter.

In Chapter 4 it was shown that DFS experiments provide an excellent platform to investigate the kinetic properties of two-state nucleic acid molecules. Combining Markov models and a Kramers-based tools, force-dependent unfolding and folding kinetic rates can be measured and used to extract information about the molecular folding and unfolding, such as the position and height of the kinetic barrier mediating each transition, the free-energy difference between the native and the unfolded states, and the attempt frequency at zero force.

More complex situations, all involving the use of nucleic acid molecules, will be addressed in this chapter using DFS experiments. First, the kinetic properties of a variety of two-state DNA molecules with different positions of the TS along the reaction

coordinate (known as "fragility") will be investigated. It will be shown that, even at high irreversible conditions, a careful interpretation of DFS experimental data provides accurate information of the molecular FEL. Second, the behavior of an RNA hairpin at different ionic conditions will be studied. In this case, the effect of both the monovalent and divalent ions is mainly non-specific. The use of DFS techniques will grant access not only to the dependence of the kinetic properties of RNA on the ionic strength, but also to the ionic-dependent elastic properties of short ssRNA molecules. The novelty of these results is noticeable: RNA molecules are exposed to many different ionic conditions inside cells, which determines their elasticity and may have an effect on the efficiency in their biological tasks. Third, Markov models will be extended to characterize kinetic states with DFS. Kinetic states are generated dynamically and play important roles in many biological processes, such as molecular folding and ligand binding. With these two examples in mind, DFS experiments are performed on DNA hairpins with one or two intermediate states on-pathway along the molecular unfolding and folding reactions. It will be shown that an extended DFS analysis to unravel the kinetic properties of intermediate states becomes particularly useful in the study of nucleic acid molecules with high kinetic barriers in the FEL. Under these circumstances, non-equilibrium DFS experiments allow us to accurately determine the existence and the stability of molecular states that are difficult to sample under equilibrium conditions. To end, the specific binding of a peptide in the DNA double helix will be investigated. The ligand binding position, the association and dissociation kinetic rates, and the binding free energies will be quantitatively characterized using a DFS approach. With all these results, the power of DFS to quantitatively understand multiple molecular mechanisms present in many biological processes happens to be way more than proven.

5.1 DNA hairpins with different fragilities

The deformability of nucleic acid molecules plays a role in biochemical processes involving the transient opening of the molecules or the recognition by ligands. The position of the TS along a reaction coordinateregulates the degree of molecular deformability. Hence, when a force is applied to the ends of a nucleic acid hairpin, the molecule unfolds in a compliant or brittle way depending on the position of the TS. The opposite behavior is expected in the reversed process: compliance (brittleness) along unfolding is accompanied by brittleness (compliance) along folding.

The concept of molecular fragility μ is a quantitative measurement of the degree of deformability of the molecule until it reaches the TS under applied force. A mathematical definition for nucleic acid hairpins is given by [Mano 06, Moss 09b]:

$$\mu = \frac{x_{N-\mathrm{TS}} - x_{\mathrm{TS}-U}}{x_{NU}}, \tag{5.1}$$

where, as already defined in Chapter 4, $x_{N-\mathrm{TS}}$ and $x_{\mathrm{TS}-U}$ are the relative distances between the TS and states N or U, respectively, and $x_{NU} = x_{N-\mathrm{TS}} + x_{\mathrm{TS}-U} = x_U - x_N$ is the molecular extension (*i. e.*, the distance between U and N). When $\mu \simeq 1$ the TS

is located close to U and the hairpin is expected to show a compliant behavior along unfolding, whereas when $\mu \simeq -1$ the TS is located close to N and the hairpin is expected to show a brittle behavior along unfolding.

Fragility does not always have the same value for a given nucleic acid hairpin. From the Leffler-Hammond postulate it is known that, as force is applied to the ends of a hairpin, the TS approaches state N and consequently μ tends to -1. In addition, it has been observed that the binding of a ligand to a nucleic acid hairpin has an effect on its ability to deform, which implies a displacement of the position of the TS [Peck 11]. In this section it will be shown how this information can be extracted with DFS experiments.

In order to study how the position of the TS affects the kinetics of a two-state system pulled under mechanical force, three different DNA hairpins were designed with different values of the fragility at the coexistence force, hereafter referred as μ_c. All hairpins are made of a 30-base pairs stem and are ended by the same GAAA-tetraloop. μ_c for the hereafter referred to as hairpin "Neutral" (HN) is approximately equal to zero, whereas hairpins "Compliant" (HC) and "Britte" (HB) have the TS close to the U and the N state, respectively. The sequences and theoretical profiles of the FEL for each haipin are shown in Fig. 5.1. The GC-content of each sequence, and the theoretical predictions for the free energy of formation obtained using the NN model (Section 4.1.1), the coexistence force f_c defined as the force at which states N and U are equally probable, and the fragility at coexistence μ_c are given in Table 5.1.

Hairpin	GC-content (%)	ΔG^0_{NU} (k_BT)	f_c (pN)	μ_c
Neutral (HN)	50	80±2	15.7±0.3	0.00
Compliant (HC)	67	86±3	16.3±0.5	0.76
Brittle (HB)	27	67±1	13.4±0.2	-0.47

Table 5.1: Design of hairpins HN, HC and HB. The GC-content of the sequence, and the theoretical predictions for the free energy of formation ΔG^0_{NU} obtained using the NN model (Section 4.1.1), the coexistence force f_c defined as the force at which states N and U are equally probable, and the fragility μ_c at coexistence. Theoretical values for ΔG^0_{NU} and f_c are the average over the two predictions obtained using the UO or the UNZ set of base-pair free energies (Table 4.1) [Sant 98, Hugu 10b]. Error bars are standard errors. The error bars for the values of fragility are small and hence they are not shown. All the quantities are derived at the standard ionic conditions of 1 M NaCl and at 298 K.

The main properties of each hairpin are:

- **Hairpin HN.** This hairpin is made of 15 GC-base pairs (50% GC-content along the stem). At f_c the FEL predicts a two-state behavior (Fig. 5.1b), where N and U are the most stable states and the TS is located at the middle of the hairpin stem (at $n \simeq 15$). The predicted height of the kinetic barrier at f_c is 13-14 k_BT, which is much larger than the one obtained for hairpin CD4 (Fig. 4.3b). This suggests that hysteresis effects for N will be much larger than the ones measured for CD4.

- **Hairpin HC.** This hairpin is made of 20 GC-base pairs (67% GC-content along the stem). The FEL profile computed at f_c is rugged but a two-states behavior is still expected, with a TS located close to U ($n \simeq 26 - 27$). The position of the

TS in the FEL profile obtained at f_c suggests that the molecule will mechanically unfold after significant deformation, *i. e.* it will have a compliant behavior.

- **Hairpin HB.** This hairpin is made of 8 GC-base pairs (27% GC-content along the stem). Again, the FEL calculated at f_c suggests a two-state behavior (Fig. 5.1b). In this case, the TS is located close to N ($n \simeq 8$). In contrast to HC, HB is expected to only slightly deform before unfolding, *i. e.* it will have a brittle behavior . Since the height of the kinetic barrier is close to 12 $k_{\mathrm{B}}T$, pulling experiments performed on this molecule should show hysteresis effects larger than the ones exhibited by hairpin CD4.

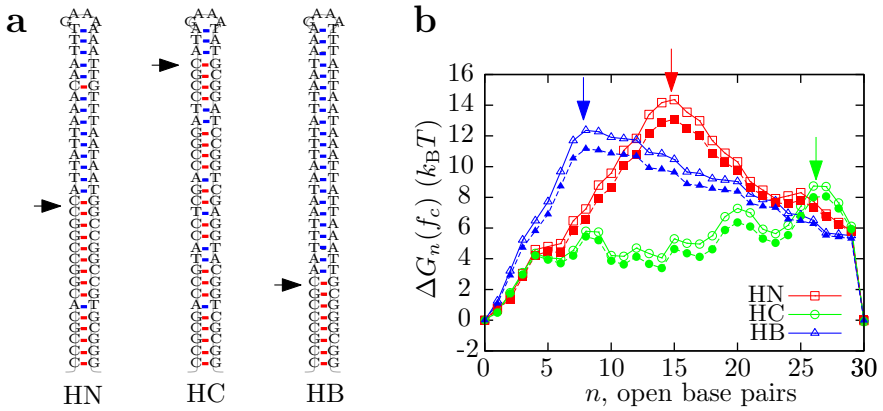

Figure 5.1: DNA hairpins with different fragilities. a. Sequence of hairpins HN, HB and HC. **b.** Predicted FEL for HN, HB and HC computed at the coexistence force f_c using the UO [Sant 98] (empty symbols, f_c =16.8, 16.0, 13.6 pN, respectively) or the UNZ [Hugu 10b] (solid symbols, f_c =15.8, 15.4 and 13.2 pN, respectively) set of base-pair free energies (Table 4.1). Arrows in both panels indicate the position of the TS along the DNA hairpin at force f_c.

DFS experiments were carried out for the three molecules. As in previous chapters, the hairpins were attached between two dsDNA handles made of 29-base pairs. The end of each handle is attached to a polystyrene bead through streptavidin-biotin and digoxigenin-antidigoxigenin bonds (Fig. 5.2a and Section 2.3.2). Details of the synthesis are provided in Appendix A.1. Two distinct experimental protocols were carried out for these hairpins: pulling and force-jump experiments.

Pulling experiments were performed at four different pulling speeds (30, 60, 120 and 180 nm/s, or equivalently, 1.8, 3.6, 7.2 and 10.8 pN/s). Examples of FDC collected for the three molecules at 3.6 pN/s are plotted in Fig. 5.2a. As expected, a two-state behavior and strong hysteresis effects between unfolding and folding curves are observed in the three cases (Fig. 5.1b). In agreement with the height of the kinetic barrier predicted at f_c, these hysteresis effects are larger for HN and HB than for HC. The unfolding and folding rupture force histograms $\rho(f_U)$ and $\rho(f_F)$ are shown in Fig. 5.2c–e. In the case of HC (panel d), it can be observed at first sight that the variance of $\rho(f_U)$ is much lower than the variance of $\rho(f_F)$, indicating that the hairpin is compliant under unfolding and therefore the unfolding rupture force tends to be the same for independent repetitions

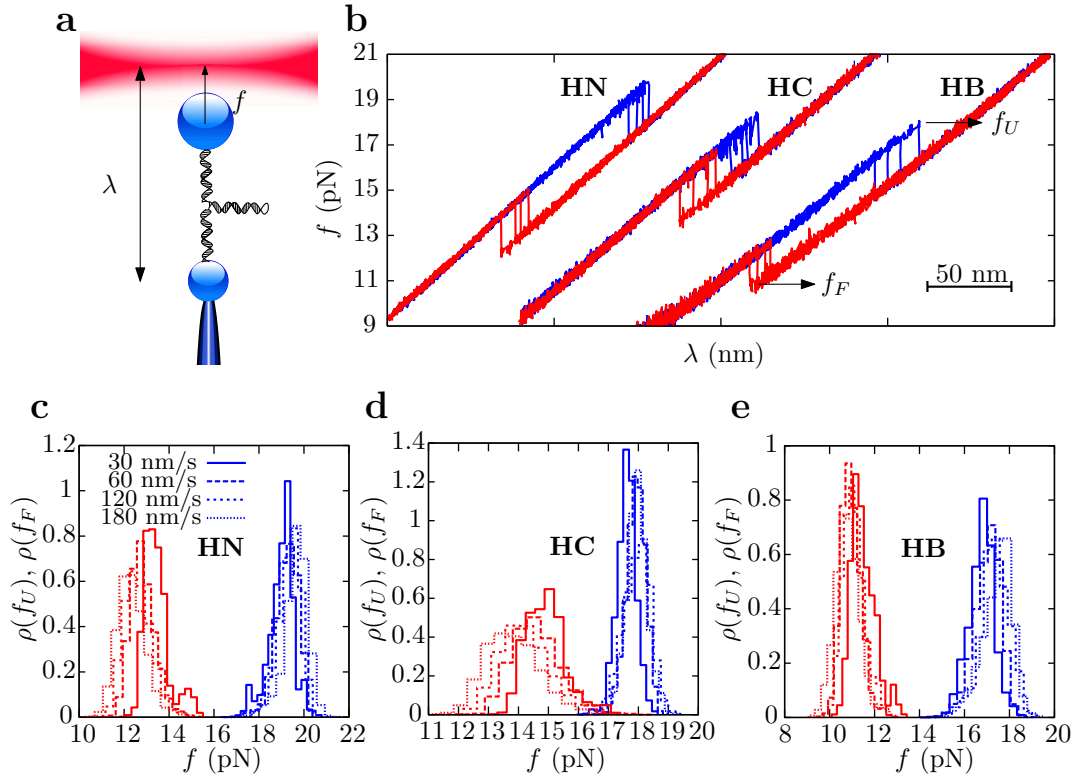

Figure 5.2: Pulling experiments performed with hairpins HN, HC and HB. a. Experimental setup (image not to scale). **b.** Collection of unfolding (blue) and folding (red) curves measured in pulling experiments carried out at 60 nm/s (or, equivalently, 3.6 pN/s). Unfolding (blue) and folding (red) rupture force histograms $\rho(f_U)$ and $\rho(f_F)$ for molecules HN (**c**), HC (**d**) and HB (**e**) obtained from pulling experiments performed at 30, 60, 120 and 180 nm/s (or 1.8, 3.6, 7.2 and 10.8 pN/s).

of unfolding experiments. In contrast, folding is abrupt (the molecule hardly deforms) and hence a large range of folding forces is explored.

In force-jump experiments the molecule is initially held at a low(high) force value, where it is found at $N(U)$ (below 10 pN or over 20 pN, respectively). The force is suddenly set to a new value (in the range 10-20 pN) where the hairpin changes state after a stochastic time interval $\Delta t_{N(U)}$, as revealed by a sudden jump in the relative distance λ between the center of the optical trap and the micropipette (Fig. 5.3). This protocol is repeated many times in order to measure the equilibrium lifetime distributions of N and U at different forces and the corresponding mean lifetimes $\langle \Delta t_{N(U)} \rangle$, whose inverses are equal to the kinetic rates $k_{N \to U}$ and $k_{N \leftarrow U}$, respectively. Force-jump experiments are performed instead of equilibrium hopping experiments (Section 2.3) because hopping between N and U was rarely simultaneously observed at any value of the force for any of the three hairpins.

Kinetic rates measured under pulling and force-jump experiments are compared in Fig. 5.4. As expected, measurements from different experimental protocols collapse into a single master curve. Here, two advantages in the use of pulling experiments instead

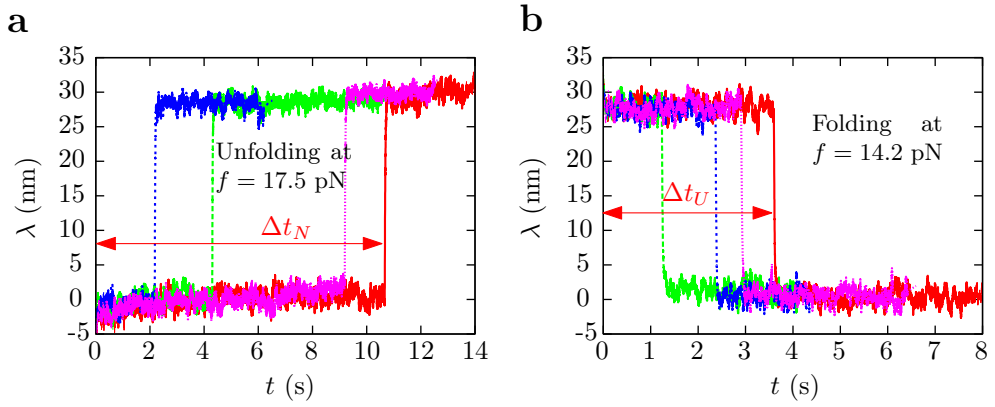

Figure 5.3: Force-jump experiments. Results for hairpin HN. **a.** Unfolding force-jump experiments carried out at 17.5 pN. **b.** Folding force-jump experiments carried out at 14.2 pN. Different repetitions of the same experiment are shown.

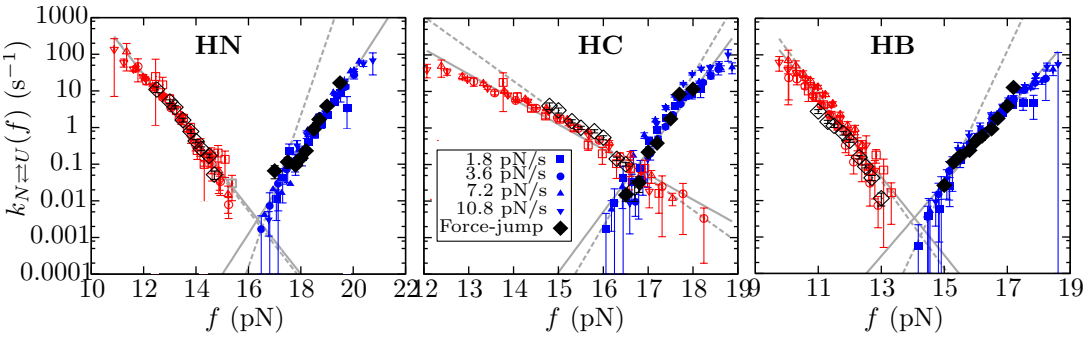

Figure 5.4: Kinetic rates for hairpins HN, HC and HB. Kinetic rates of unfolding $k_{N\rightarrow U}(f)$ (blue-solid symbols) and folding $k_{N\leftarrow U}(f)$ (red-empty symbols) measured from pulling experiments performed at different loading rates and force-jump experiments (black symbols), and fits to the BE model performed along the whole range of measured force (solid-gray line) or performed in the vicinity of f_c (dashed-gray lines) for the three hairpins HN, HC and HB.

of force-jump measurements to estimate the kinetic rates are found. First, in pulling experiments a larger range of forces is explored for both $k_{N\rightarrow U}(f)$ and $k_{N\leftarrow U}(f)$. This is because of hysteresis effects that allow the system to sample a larger range of folding and unfolding forces. Second, pulling experiments allow us to collect data faster and on a wider range of forces. Moreover, drift effects –typically attributed to uncontrolled movements of the microfluidics chamber– are easier to correct in non-equilibrium than in force-jump experiments.

5.1.1 The Bell-Evans model

As a first approximation, the kinetic rates can be analyzed using the BE model (Section 4.3.1). In this direction, the experimentally measured $k_{N\rightarrow U}(f)$ and $k_{N\leftarrow U}(f)$ are first fitted to Eqs. (4.25a) and (4.25b) in the entire experimentally available range of

forces in order to extract k_m, ΔG_{NU}, $x_{\mathrm{TS}-U}$, $x_{N-\mathrm{TS}}$ and f_c. Results of such fits for the three hairpins are shown in Fig. 5.4b (solid-gray line) and numerical estimations are summarized in Table 5.2. Comparison with theoretical values for ΔG_{NU}, $x_{\mathrm{TS}-U}$ and $x_{N-\mathrm{TS}}$ (provided in Table 5.3) reveals discrepancies. In the best case the relative error of ΔG_{NU} is of the 20% for molecule HC.

Hairpin	k_m (s^{-1})	ΔG_{NU} ($k_{\mathrm{B}}T$)	$x_{N-\mathrm{TS}}$ (nm)	$x_{\mathrm{TS}-U}$ (nm)	f_c (pN)	μ_c (ad.)
HN	$(2.6 \pm 0.5)\times 10^{-21}$	76±2	10.5±0.3	8.6±0.3	16.3±0.7	0.10±0.06
HC	$(1.8 \pm 0.5)\times 10^{-29}$	90±3	15.6±0.7	6.5±0.2	16.7±0.7	0.41±0.07
HB	$(2.8 \pm 0.6)\times 10^{-17}$	68±2	9.5±0.3	10.2±0.5	14.0±0.5	-0.04±0.05

Table 5.2: Fit of the experimental kinetic rates for HN, HC and HB to the BE model. The logarithm of the experimental kinetic rates plotted in Fig. 5.4b were fitted to Eq. (4.25) to get estimations for k_m, ΔG_{NU}, $x_{N-\mathrm{TS}}$, $x_{U-\mathrm{TS}}$ and $f_c = \Delta G_{NU}/(x_{N-\mathrm{TS}} + x_{U-\mathrm{TS}})$. Error bars are calculated using error propagation of standard errors from the fit.

Hairpin	ΔG_{NU} ($k_{\mathrm{B}}T$)	$x_{N-\mathrm{TS}}$ (nm)	$x_{\mathrm{TS}-U}$ (nm)	f_c (pN)	μ_c (ad.)
HN	111±3	13.55±0.05	15.4±0.1	15.7±0.3	0.00
HC	118±4	24.05±0.05	5.5±0.1	16.3±0.5	0.76
HB	94±2	7.20±0.05	21.5±0.1	13.4±0.2	-0.47

Table 5.3: Predicted parameters for BE model of HN, HC and HB. Theoretical values for hairpins HN, HC and HB were obtained by averaging over the two predictions obtained using the NN model with the UO or the UNZ set of base-pair free energies (Table 4.1) [Sant 98, Hugu 10b]. Error bars are standard errors. Error bars for fragilities are small and hence they are not shown.

As already discussed in Section 4.3.1, there are two reasons to explain such large discrepancy. The first is that the simplified expressions of the BE kinetic rates (Eqs. 4.25a and 4.25b) are only valid for a short force interval centered around f_c. The second is related to the Leffler-Hammond postulate [Leff 53, Hamm 55], that predicts that the TS approaches N as force applied increases. This second situation is not considered by the BE model.

By restricting the BE fit to a force interval close to the coexistence force ($f \in [f_c - \Delta f, f_c + \Delta f]$, $\Delta f \simeq 1$ pN) one gets the dashed-gray line in Fig. 5.4b and the numerical values summarized in Table 5.4. Although the agreement with theoretical predictions is a little bit better, strong discrepancies are still evident. Hence, for the case of hairpins HN, HC and HB the the Leffler-Hammond effect is relevant for forces values close to f_c. This fact will be discussed in the next section.

5.1.2 The force-dependent kinetic barrier

As it has been previously shown for CD4 (Section 4.3.2), it is possible to extract the force-dependence of the kinetic barrier $B_{NU}(f)$ for two-state molecules from the experimental kinetic rates measured with DFS experiments (Eqs. 4.34a and 4.34b). Moreover, in the

Hairpin	k_m (s^{-1})	ΔG_{NU} (k_BT)	x_{N-TS} (nm)	x_{TS-U} (nm)	f_c (pN)	μ_c (ad.)
HN	$(1.4 \pm 1)\times 10^{-37}$	115±20	20±5	9±1	16±3	0.4±0.1
HC	$(1.5 \pm 1)\times 10^{-36}$	115±10	20±1	8±1	17±4	0.43±0.06
HB	$(6.0 \pm 2.0)\times 10^{-30}$	100±10	17±2	11±3	14±3	0.21±0.08

Table 5.4: Fit of the experimental kinetic rates for HN, HC and HB to the BE model restricted to a narrow range of forces close to f_c. The logarithm of the experimental kinetic rates shown in Fig. 5.4b were fitted to Eq. (4.25) in the range of forces $f \in [f_c \pm \Delta f]$, where $\Delta f = 1$ pN. Estimations for k_m, ΔG_{NU}, x_{N-TS}, x_{U-TS} and $f_c = \Delta G_{NU}/(x_{N-TS} + x_{U-TS})$ are obtained. Error bars are calculated using error propagation of standard errors from the fit.

case of nucleic acid hairpins it is possible to theoretically predict the profile of $B_{NU}(f)$ using a Kramers-based approach (Eq. 4.32). The comparison between theoretical and experimental profiles of $B_{NU}(f)$ provides estimates for the free energy of formation of the hairpin, ΔG^0_{NU}, and the attempt frequency at zero force, k_0.

In Fig. 5.5 the match between the theoretical and the experimental profile of $B_{NU}(f)$ is shown for the three hairpins HN, HC and HB. A χ^2-function was minimized in order to find the best estimates for both ΔG^0_{NU} and k_0 (Appendix E). Numerical results are given in Table 5.5, where it can be observed that the values of ΔG^0_{NU} are in excellent agreement with their corresponding theoretical predictions for the three hairpins (Table 5.1). Additionally, the minimization of the χ^2-function also allows us to obtain the elastic properties of ssDNA for each hairpin (Appendix E). As it can be appreciated in Table 5.5, values found for the persistence length are in very good agreement with results found in Chapter 3.

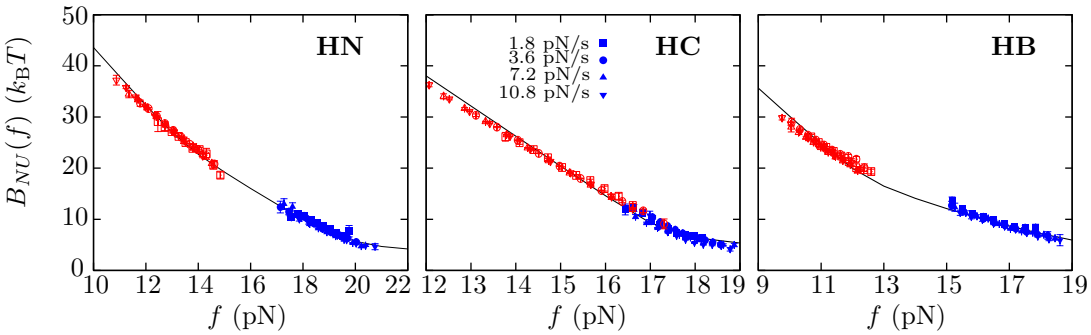

Figure 5.5: Force-dependent kinetic barrier. Comparison between the experimental profile of the kinetic barrier (blue symbols obtained with Eq. 4.34a and red symbols Eq. 4.34b) and the theoretical prediction obtained using Eq. (4.32) and the NN model with the UO set of parameters from [Sant 98] for hairpin HN (left), HC (middle) and HB (right). Similar predictions are obtained with the UNZ set.

It can be seen in Fig. 5.5 that the kinetic barrier monotonically decreases with force. By taking the derivative of $B_{NU}(f)$ with respect to force, the position of the TS relative to N, x_{N-TS}, can be extracted (Eq. 4.19). This has been done for both the theoretical profile of $B_{NU}(f)$ and experimental data, and results are shown in Fig. 5.6. The rea-

Hairpin	k_0 (s^{-1})	ΔG^0_{NU} (k_BT)	P (nm)
HN	$(3\pm1)\times10^4$	80±2	1.3±0.1
HC	$(9\pm1)\times10^4$	85±3	1.3±0.2
HB	$(2\pm1)\times10^4$	68±2	1.6±0.1

Table 5.5: Numerical results for attempt frequency k_0, free energy of formation ΔG^0_{NU}, and persistence length P obtained by matching the experimental and the theoretical profile of $B_{NU}(f)$. Results obtained by minimization of χ^2-function (Appendix E). Numerical values are the average over results obtained using the two theoretical estimations of $B_{NU}(f)$ calculated using the NN model with the UO and UNZ set of base-pair free energies [Sant 98, Hugu 10b]. Error bars are standard errors.

sonable agreement found corroborates that, as predicted by the Leffler-Hammond effect, the TS approaches N as force increases (that is, $x_{N-\mathrm{TS}}$ decreases) [Leff 53, Hamm 55]. Hence, for the three hairpins the kinetic barrier at low forces is dominated by the entropy of the loop located near the unfolded state. At intermediate forces the TS is found in the interphase separating high GC- from high AT-content regions. At forces above 17 pN the kinetic barrier is close to N. Finally, at large forces the unfolding process is barrier-less and reminiscent of the behavior observed in first order phase transitions beyond the spinodal line.

These results reveal that the unfolding and the folding of a DNA hairpin can be mediated by different TS. Here, most folding events take place at forces where the kinetic barrier is located either in the loop or in the interphase between high GC- and AT-content regions (large values of $x_{N-\mathrm{TS}}$); in contrast, most unfolding events take place at forces where the kinetic barrier is located close to N (low values of $x_{N-\mathrm{TS}}$). Thus, the DFS analysis based on the study of the kinetic barrier is very predictive for nucleic acid hairpins. In the next sections this method will be extended to unravel the kinetic behavior of nucleic acid hairpins under different conditions.

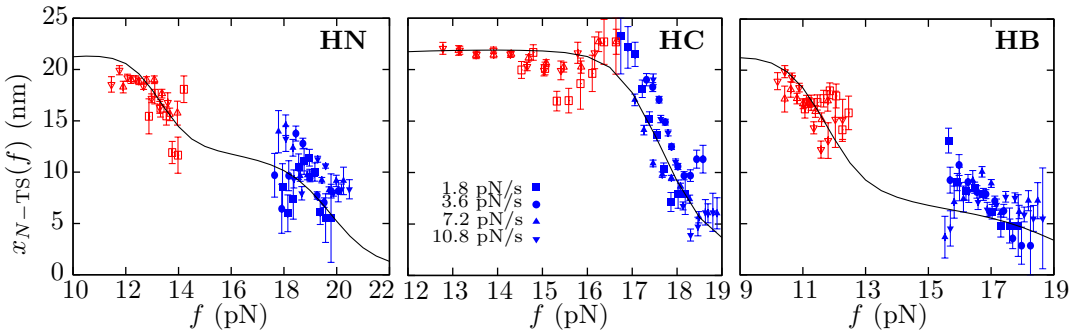

Figure 5.6: Force-dependent position of the kinetic barrier for hairpin HN, HC and HB. Relative distance between the TS and state N, $x_{N-\mathrm{TS}}$, calculated as the force derivative of $B_{NU}(f)$ (Eq. 4.19) for both theoretical (solid-black line) and experimental data, for three hairpins HN (left), HC (center) and HB (right). Blue (red) symbols obtained from experimental unfolding (folding) forces (Eqs. 4.34a and 4.34b, respectively).

5.2 Non-specific binding of Na^+ and Mg^{2+} to RNA

Inside cells, RNA molecules are exposed to a wide variety of monovalent and multivalent ionic conditions. At the same time, RNA is a highly charged polyanion whose stability strongly depends on ionic salt conditions [Reco 75, Reco 76, Will 96, Tan 05]. Despite its biological significance, a proper quantification of the ionic effects on the stability of the RNA double helix is still absent. In Section 4.1.1 is was mentioned that the base-pair free energies for RNA have only been derived at standard 1M NaCl ionic conditions using bulk measurements, such as UV absorbance or calorimetry experiments [Walt 94, Serr 94, Serr 97, Xia 98, Math 99, Znos 02, Math 04]. In addition, there is limited information about the elastic properties of ssRNA as compared with ssDNA. Despite the chemical similarity between both polymers, the RNA-specific 2'-OH group favors a conformationally restricted C3'-endo structure and hence the elastic properties cannot be assumed to be the same [Rich 03] (Fig. 2.1).

In this section the non-specific effect of both monovalent and divalent ions to the stability and the kinetics of a two-state RNA hairpin is studied. To this end, DFS experiments are performed on the RNA hairpin shown in Fig. 5.7a. This molecule, hereafter called CD4-RNA, has the exact same sequence as hairpin CD4 used in Chapter 4, with thymines replaced by uracils. CD4-RNA is inserted between two ∼500 bases-long hybrid RNA/DNA handles (instead of the short 29-bases long dsDNA handles used in the previous experiments reported). The end of each handle is attached to a polystyrene bead through SA-biotin and AD-digoxigenin bonds (Fig. 5.7b). Details of the synthesis of the molecular construct are provided in reference [Biza 12] and summarized in Appendix A.2.

A systematic study is performed by mechanically pulling with the mini-tweezers the RNA hairpin at different monovalent and mixed monovalent/divalent cation concentrations and at two different pulling speeds (1.8 pN/s and 12.5 pN/s) [Biza 12]. Experiments are carried out in buffers containing 100 mM Tris.HCl (pH 8.1), 1mM EDTA, and NaCl concentrations of 0, 100, 500, and 1000 mM, or in buffers containing 100 mM Tris.HCl (pH 8.1) and $MgCl_2$ concentrations of 0.01, 0.1, 0.5, 1, 4, and 10 mM. The monovalent cation concentration $[Mon^+]$ includes the contributions from $[Na^+]$ ions and dissociated $[Tris^+]$ ions. At 25°C and pH 8.1, about half of the Tris molecules are protonated, therefore 100 mM Tris buffer adds 50 mM to the total monovalent ion concentration [Owcz 08].

Examples of FDC measured at 1 M NaCl conditions are shown in Fig. 5.7c. The two-state behavior –the hairpin is either folded or unfolded– is preserved among all ionic conditions. Consequently, using the DFS analysis introduced in Chapter 4, the dynamics and kinetics of CD4-RNA will we extracted at different ionic strengths. Moreover, the method also allows us to extract the ionic-dependence of the elastic properties of short ssRNA molecules.

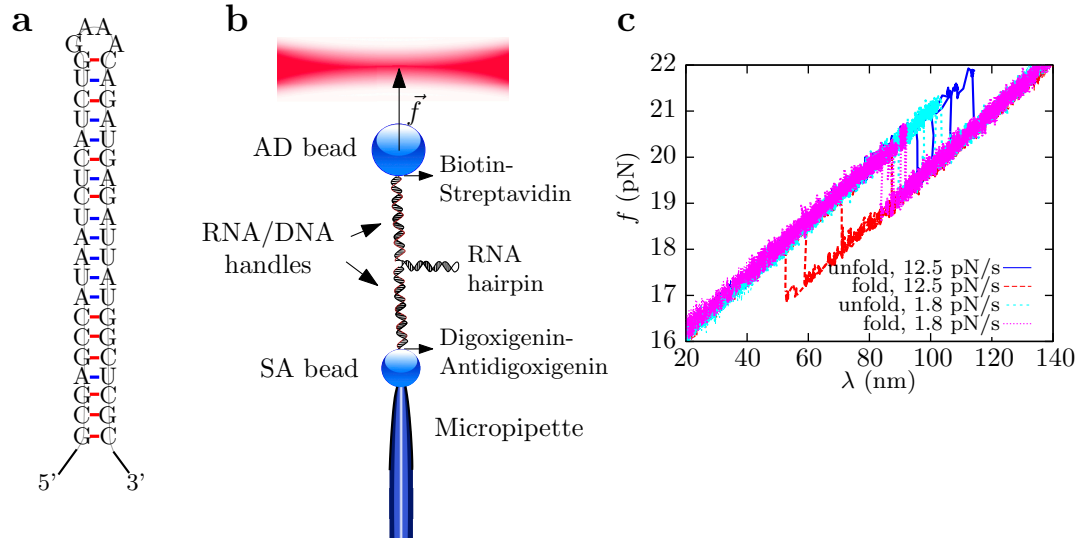

Figure 5.7: Pulling experiments on CD4-RNA hairpin. a. RNA hairpin sequence and structure. **b.** A single RNA hairpin is attached to two polystyrene microspheres through RNA/DNA heteroduplexes used as handles. The AD bead is optically trapped while the SA bead is positioned at the tip of a micropipette by air suction. **c.** FDC measured in pulling experiments carried out at the two different pulling speeds 1.8 pN/s (unfolding in cyan and folding in purple) and 12.5 pN/s (unfolding in blue and folding in red) in a buffer containing 1 M NaCl and 100 mM Tris.HCl at pH 8.1.

5.2.1 Monovalent cation conditions

Pulling experiments were performed at two different pulling speeds and in a buffer containing 100 mM Tris.HCl and four different NaCl concentrations (0, 0.1, 0.5 and 1 M). Therefore, four different concentrations of monovalent ions are explored, which are 50, 150, 550 and 1050 mM [Mon^+]. At any of these ionic conditions CD4-RNA behaves as a two-state system from which we can extract unfolding and folding rupture forces.

It is observed that RNA duplex stability increases at higher [Mon^+]. For instance, rupture force distributions are displaced to higher forces as the concentration of NaCl increases (Fig. 5.8a). The increment in the duplex stability at higher salt concentrations can also be observed in the salt-dependence of the mean unfolding and folding forces (Fig. 5.8b, left), which also increase with the concentration of NaCl. Interestingly, the standard deviation of rupture force distributions remains almost constant for the different ionic conditions explored (Fig. 5.8b, right). According to the BE model (Appendix C.4), such behavior denotes that the position of the TS mediating unfolding and folding transition does not depend on [Mon^+], despite the fact that both transitions occur at higher forces. In Fig. 5.8c it is observed that the unfolding and folding kinetic rates (obtained using Eq. 4.23) displace towards large forces along the force-axis with the salt concentration, which again indicates the stabilizing effect of salt on the RNA hairpin.

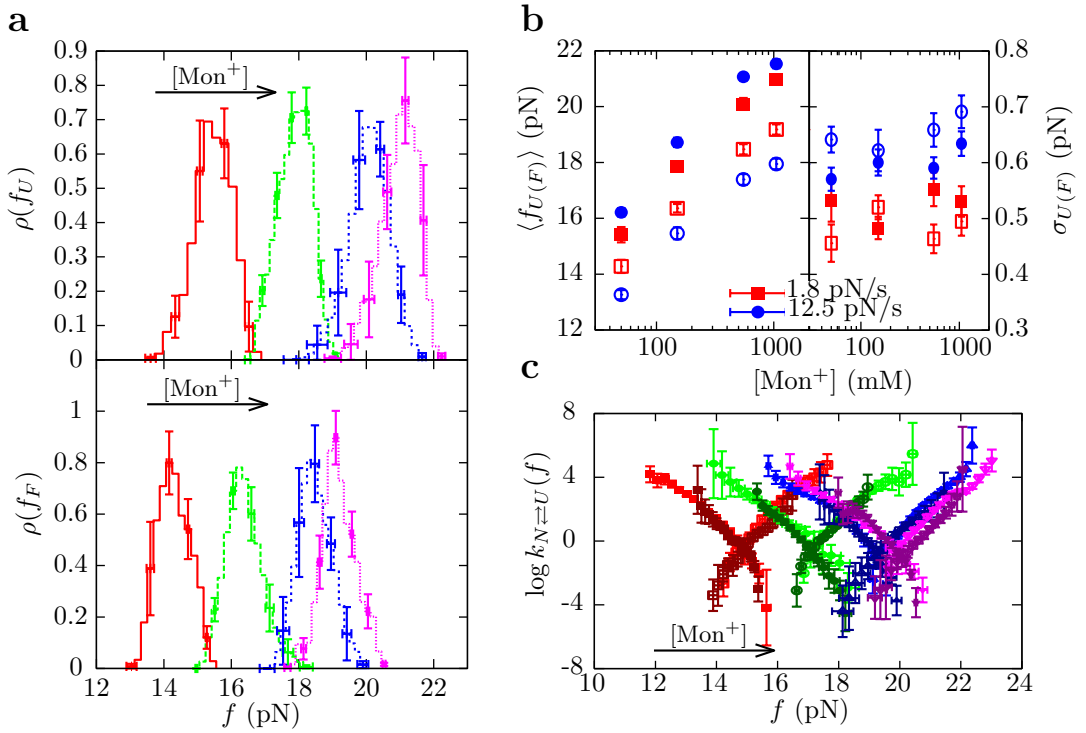

Figure 5.8: DFS results at varying [Mon$^+$]. **a.** Experimental distribution of the unfolding (top) and folding (bottom) rupture forces measured in pulling experiments performed at 1.8 pN/s in buffers containing 50 mM (red), 150 mM (green), 550 mM (blue) and 1050 mM (magenta) [Mon$^+$]. The arrow indicates the direction of increasing monovalent concentration. **b.** Average rupture forces $\langle f_{U(F)} \rangle$ and standard deviations $\sigma_{U(F)}$ as a function of monovalent cation concentration at loading rates of 1.8 pN/s (red) and 12.5 pN/s (blue). Filled symbols refer to unfolding and empty symbols to refolding. **c.** Kinetic rates versus force. Experiments were done at 50 mM [Mon$^+$] for loading rates of 1.8 pN/s (dark colors) and 12.5 pN/s (light colors), at different ionic conditions. Color code as in a. Error bars in the different panels are statistical errors and have been calculated using the Bootstrap method.

Standard 1M NaCl ionic conditions

From the current set of base-pair free energies for RNA, the free energy of formation ΔG^0_{NU} of secondary structures can be predicted at 1 M NaCl concentration (Table 4.1) [Xia 98, Math 04, Tino 71, Tino 73, Zuke 03]. For CD4-RNA at this particular condition using Eq. (4.2) one gets $\Delta G^0_{NU} = 63.0\ k_{\mathrm{B}}T$.

As already shown in Chapter 4 and Section 5.1, both the logarithm of the attempt frequency at zero force, $\log k_0^{NU}$, and the free energy of formation, ΔG^0_{NU}, of CD4-RNA are accessible from the direct comparison between theoretical and experimental estimations of the kinetic rate (obtained using Eq. 4.32 and 4.34, respectively). Here, experiments are performed at 1.05 M of monovalent salt, whereas theoretical predictions are available at 1 M. However, a negligible error is expected by comparing the two estimations.

Noteworthy, the elastic properties of ssRNA at 1 M NaCl have not been determined

before. In the absence of a value for the persistence length P at such ionic condition, the extended definition of the χ^2-function to find the best match between theoretical and experimental estimations of $B_{NU}(f)$ is used to determine P, $\log k^0_{NU}$ and ΔG^0_{NU} (Appendix E). The interphosphate distance is set to $d_b = 0.665$ nm/base, which is a standard value found in literature [Murp 04]. Finally, the best match obtained using the WLC model and minimizing the χ^2-function gives $P = 0.75 \pm 0.05$ nm, $k_0 = (5 \pm 2) \times 10^4$ s^{-1} and $\Delta G^0_{NU} = 65.1 \pm 0.5\ k_BT$, the former value being in reasonable agreement with theoretical predictions determined using the UO set of base-pair free energies ($\Delta G^0_{NU} = 63\ k_BT$).

Elastic parameters at different monovalent ionic conditions

The elastic behavior of a ssRNA molecule made of uracil nucleotides (poly-U) has been studied in single-molecule stretching and fluorescence experiments carried out at various [Na^+] concentrations [Seol 04, Toan 06, Chen 12]. In Fig. 5.9 the resulting values of the persistence lengths of poly-U are plotted as a function of monovalent salt concentration (black-solid squares) together with the value found for of CD4-RNA at 1.05 M [Mon^+] (black-open circle).

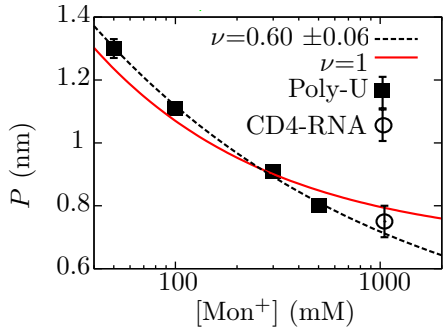

Figure 5.9: Persistence length P of ssRNA *vs* [Mon^+]. Experimental data for a poly-U RNA molecule taken from [Seol 04, Toan 06] (solid squares), and value obtained for CD4-RNA at 1.05 M [Mon^+] (empty circle). Two different fits are done to Eq. (5.2): ν set to 1 (solid red) and ν as free parameter $\nu = 0.60 \pm 0.06$ (dashed black).

[Mon^+] (M)	P (nm)
0.05	1.27±0.03
0.15	1.01±0.01
0.55	0.82±0.02
1.05	0.75±0.05

Table 5.6: Persistence length P of ssRNA *vs* [Mon^+]. Values of ssRNA persistence length P obtained by averaging results from the fits of data in Fig. 5.9 to Eq. (5.2) performed by fixing ν to 1 or by letting it be a free parameter.

It is reasonable to assume that the persistence length is related to the Debye screening length, which is considered as the fundamental length of ionic solutions. Hence, if a power dependence is considered, P depends on [Mon^+] according to:

$$P \sim \lambda_D^\nu \propto \frac{1}{[\mathrm{Mon}^+]^{\nu/2}}. \tag{5.2}$$

No unique scaling law can be derived, as the value of ν can be 1, 2 or <1 depending on the polymer properties [Mick 96]. Data in Fig. 5.9 is fitted to Eq. (5.2) using $\nu = 1$ (solid-red line in Fig. 5.9), according to the Poisson-Boltzmann theory (Appendix G.1).

In addition, the fit is also performed by letting ν be a free parameter, which results in a value of $\nu = 0.60 \pm 0.06$ and (dashed-black line in Fig. 5.9). In order to infer the values of P for ssRNA in our experimental conditions of monovalent ionic concentration (50, 150, 550 and 1050 mM [Mon^+]) results are averaged over both fits and shown in Table 5.6. These numbers have been obtained under two significant assumptions, which are:

i. Values obtained at different monovalent ionic conditions for two molecules with extremely different contour lengths have been merged together: CD4-RNA in its unfolded stretch conformation is made of 44 bases, whereas poly-U is a polynucleotide made of 1500-4000 bases [Toan 06]. However, its was shown in Chapter 3 that, the persistence length of ssDNA might have a dependence of the molecular contour length. Here, this dependence is neglected for ssRNA.

ii. It is also assumed that the elastic properties of ssRNA strands are independent of the sequence, which can lead to systematic errors in the values obtained in Table 5.6 in the case of sequence-dependent behavior. In fact, a sequence-dependent elastic behavior for ssDNA strands was previously considered as a possible explanation for the NN-specific salt corrections found for the base pair free-energy values obtained from UNZ experiments [Hugu 10b].

Non-specific monovalent salt correction for the base-pair free energies of the RNA double helix

It is not possible to determine the free energy of formation of CD4-RNA at different ionic conditions with UV absorbance experiments because the melting temperature of the molecule is too high to obtain reliable results (Appendix F). At 1 M NaCl, the estimation of ΔG^0_{NU} is obtained using the NN model and the base-pair free energies proposed in references [Walt 94, Math 99].

It has been previously shown that a sequence-dependent salt correction to the NN base-pair free energies of DNA improves the free-energy prediction of both unzipping and melting experiments [Hugu 10b]. In addition, it has been found that cation concentration affects RNA stability in a sequence-dependent manner [Vier 07]. In the absence of sequence-dependent salt correction parameters available for RNA, a sequence-independent correction $g_1([Mon^+])$ for the free energy of formation of one base pair is assumed [Sant 98, Sant 04, Shke 04]. The free energy of formation of an intermediate configuration n, being n the number of open base pairs, of an RNA hairpin at a given monovalent ionic concentration [Mon^+] is determined by the following expression:

$$\Delta G^0_n([Mon^+]) = \Delta G^0_n - (N - n)\, g_1([Mon^+]), \tag{5.3}$$

where ΔG^0_n corresponds to the free energy of formation at zero force of configuration n at 1 M [Mon^+] and N to the total number of base pairs in the stem. To determine the free energy of the configuration n at a given force f it is necessary to include the stretching contribution due to the released ssRNA chain and the reversible work of orientation of the RNA double helix, as explained in Section 4.1.1.

There are experimental and theoretical evidences that support a logarithmic effect of monovalent ions on the stability of nucleic acid hairpins. Therefore, the sequence-independent salt correction for a single base pair will be assumed to be:

$$g_1([\mathrm{Mon}^+]) = m \log[\mathrm{Mon}^+], \tag{5.4}$$

where m is a constant and $[\mathrm{Mon}^+]$ is expressed in molar (M) units. Using this correction, the variation of $\Delta G_n^0([\mathrm{Mon}^+])$ with monovalent salt concentration depends strictly on the value of the constant m (Eq. 5.3).

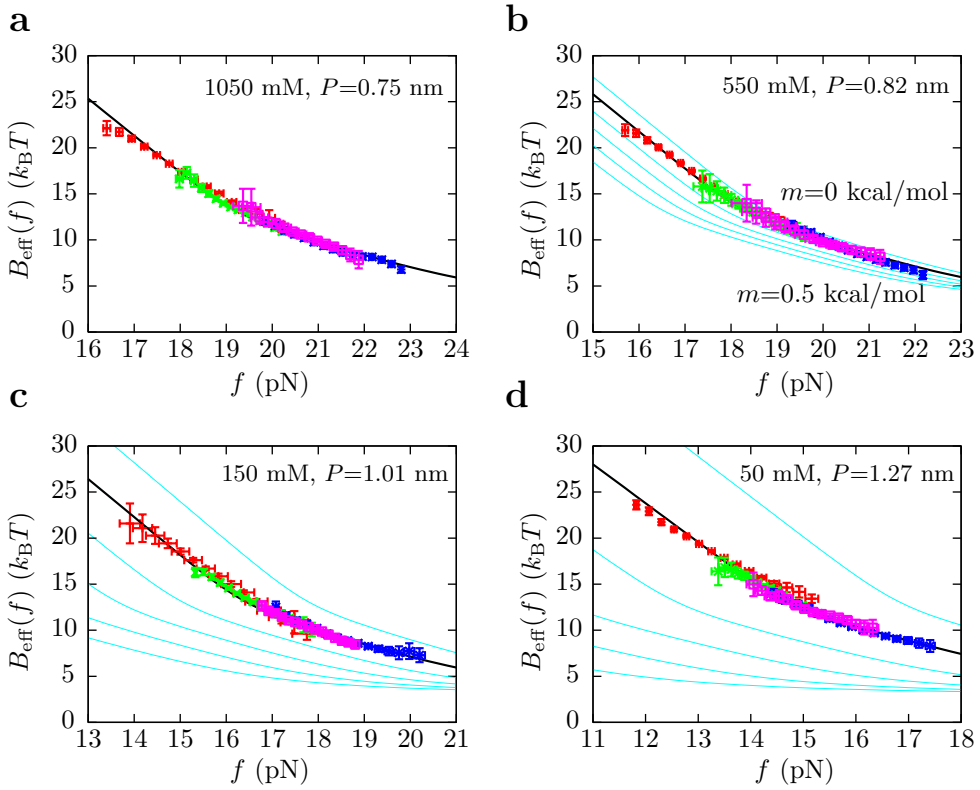

Figure 5.10: Kinetic barrier at different $[\mathrm{Mon}^+]$ and determination of the salt correction parameter m. The experimental profile of $B_{NU}(f, [\mathrm{Mon}^+])$ is compared to the theoretical prediction obtained for different values of m (0, 0.1, 0.2, 0.3, 0.4 and 0.5 kcal/mol, light blue lines from top to bottom) at 1050 mM $[\mathrm{Mon}^+]$ (**a**), 550 mM $[\mathrm{Mon}^+]$ (**b**), 150 mM $[\mathrm{Mon}^+]$ (**c**), and 50 mM $[\mathrm{Mon}^+]$ (**d**). Red (green)/blue (magenta) points are the experimental estimators $B_{NU}(f)$ determined from folding/unfolding forces at the pulling speeds 12.5 (1.8) pN/s. Black lines are the theoretical estimators of $B_{NU}(f)$ at $m = 0.1$ kcal/mol that match experimental results.

In order to determine m, the theoretical and the experimental profiles of $B_{NU}(f, [\mathrm{Mon}^+])$ obtained using Eqs. (4.32) and (4.34), respectively, with the persistence length values taken from Table 5.6 are compared at each monovalent ionic condition for different values of m until the best match is found. Again, this is equivalent to minimize a new χ^2 function (as the ones defined in Appendix E) where $\Delta G_{NU}^0([\mathrm{Mon}^+])$, $\log k_{NU}^0([\mathrm{Mon}^+])$

and m are the free parameters. In Fig. 5.10 the matching between theory and experiments at each concentration of monovalent ions is shown. For all cases, the best agreement is found at $m = 0.10 \pm 0.01$ kcal/mol. This value agrees with the sequence-independent salt correction reported for DNA duplex oligomers in melting experiments, $m = 0.110 \pm 0.033$ kcal/mol [Sant 98, Peyr 00], and in unzipping experiments of polymeric DNA, $m = 0.104 \pm 0.010$ kcal/mol [Hugu 10b]. For a given force, it can be seen in Fig. 5.10a-d that the height of the kinetic barrier increases with salt concentration. This is yet another indication that salt increases the stability of the RNA structure. In addition, the increment of ΔG^0_{NU} with salt also points towards that direction (Table 5.7).

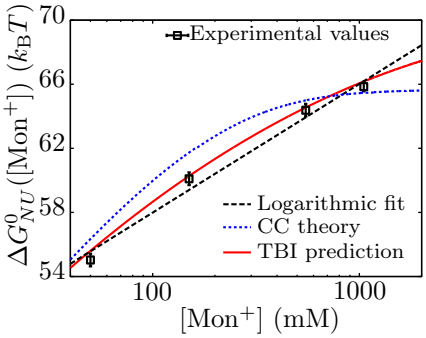

Figure 5.11: Free energy of formation as a function of [Mon$^+$] for CD4-RNA. $\Delta G^0_{NU}([\mathrm{Mon}^+])$ determined by matching experimental and theoretical profiles of $B_{NU}(f)$ for the different ionic conditions (black squares), fit to a logarithmic dependence (black-dashed line), and comparison to predictions provided by the Counterion condensation (CC) and the Tightly Bound Ion (TBI) theories (blue-dotted and red-solid lines, respectively).

[Mon$^+$] (M)	$\log k^0_{NU}$	ΔG^0_{NU} ($k_\mathrm{B}T$)
1.05	10.9±0.4	65.1±0.5
0.55	10.5±0.4	64.0±0.5
0.15	11.3±0.5	59.4±0.5
0.50	12.4±0.4	54.0±0.5

Table 5.7: Logarithm of the attempt frequency and free energy of formation of CD4-RNA as a function of [Mon$^+$]. Results obtained by matching the experimental profile of the kinetic barrier and the theoretical predictions at the different ionic conditions (Fig. 5.10).

In Fig. 5.11 the dependence of $\Delta G^0_{NU}([\mathrm{Mon}^+])$ on the monovalent ion concentration for CD4-RNA is plotted. As expected from earlier observations on DNA [Reco 75, Dove 62, Priv 69], an approximately linear dependence of the RNA duplex stability on the logarithm of monovalent salt concentration is observed (black-dashed line in Fig. 5.11). Nevertheless, experimental results are also well-described by different theories that account for the energetic interactions between ions in solution and nucleic acids (Appendixes G.1 and G.2). First, predictions from the Poisson-Boltzmann theory and the counterion condensation (CC) theory are in reasonable agreement with the experimental data (blue-dotted line). The CC theory was derived by Manning and developed to interpret polyelectrolyte effects on equilibrium involving highly charged, locally rod-like polyelectrolytes [Reco 76, Mann 72, Mann 78, Reco 78, Mann 02]. It is based on different mean field approaches and neglects correlations between the ions in the solution. Second, the experimental results are in better agreement with the empirical expressions derived in the context of the Tightly Bound Ion (TBI) model [Tan 05, Tan 07, Tan 08] (Appendix G.3), that predicts free energies of formation of RNA hairpins at different

ionic conditions (red-solid line in Fig. 5.11). Unlike the CC theory, the TBI model takes into account the different modes of correlations between counterions.

5.2.2 Mixed monovalent/divalent cation conditions

Pulling experiments were performed at two different pulling speeds and in a buffer containing 100 mM Tris.HCl and varying concentrations of $MgCl_2$ (0.01, 0.10, 0.50, 1.00, 4.00 and 10.0 mM). Because under working conditions half of the Tris molecules are protonated, DFS experiments are carried out at mixed monovalent/divalent ionic conditions: the concentration of monovalent ions is always 50 mM and the concentration of divalent ions is equal to the concentration of $MgCl_2$. Again, at any of these ionic conditions CD4-RNA behaves as a two-state system.

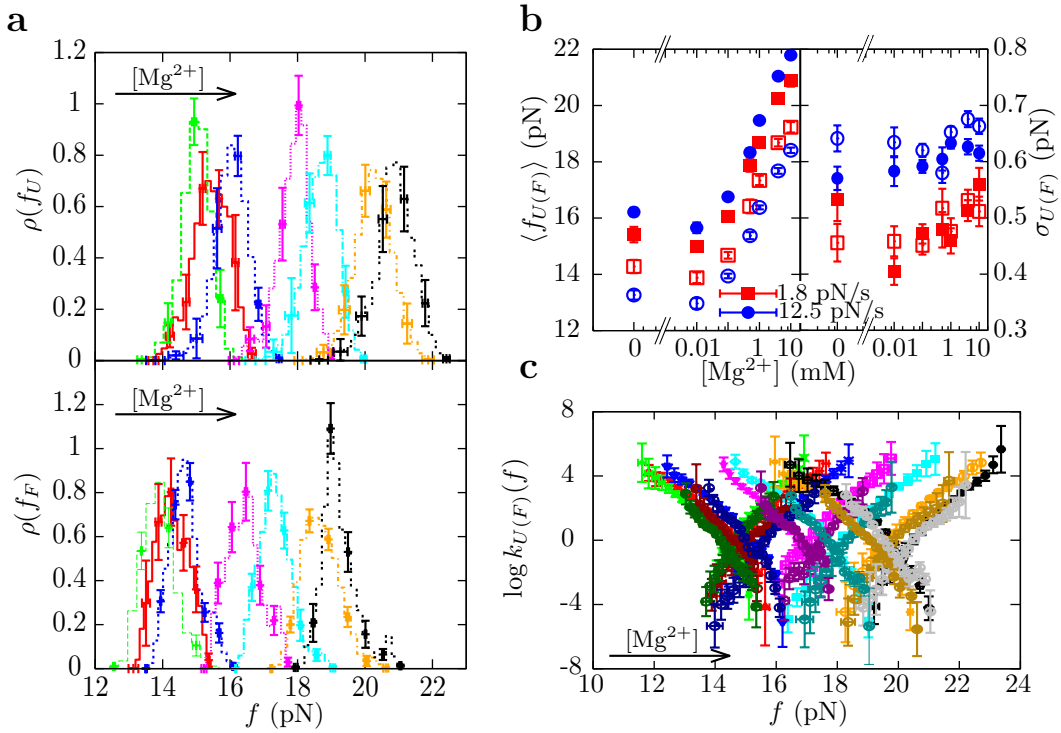

Figure 5.12: DFS results at varying $[Mg^{2+}]$. a. Experimental distribution of the unfolding (top) and folding (bottom) rupture forces measured in pulling experiments performed at 1.8 pN/s in buffers containing 50 mM (red), 150 mM (green), 550 mM (blue) and 1050 mM (magenta) $[Mg^{2+}]$ (and constant $[Mon^+]$=50 mM). The arrow indicates the direction of increasing concentration of Mg^{2+}. **b.** Average rupture forces $\langle f_{U(F)} \rangle$ and standard deviations $\sigma_{U(F)}$ as a function of divalent cation concentration at loading rates of 1.8 pN/s (red) and 12.5 pN/s (blue). Filled symbols refer to unfolding and empty symbols to refolding. **c.** Kinetic rates versus force. Experiments were done at fixed 50 mM $[Mon^+]$ for loading rates of 1.8 pN/s (dark colors) and 12.5 pN/s (light colors), at different ionic conditions. Color code as in a. Error bars in the different panels are statistical errors and have been calculated using the Bootstrap method.

From rupture force distributions and average unfolding and folding forces two regimes

are found in the behavior of the hairpin stability along the range of $[Mg^{2+}]$ experimentally explored (Fig. 5.12). Below 0.1 mM $[Mg^{2+}]$, there is no significant difference between control (no Mg^{2+} added) and magnesium-containing conditions. However, at higher magnesium concentrations, rupture force distributions are displaced to higher forces and the mean unfolding and folding forces linearly increase with the logarithm of $[Mg^{2+}]$ (Fig. 5.12a-b).

Interestingly, Owczarzy *et. al.* made similar observations in DNA melting experiments done in mixed monovalent/Mg^{2+} conditions [Owcz 08]. They found that the ratio R:

$$R = \frac{\sqrt{[Mg^{2+}]}}{[Mon^{+}]} \tag{5.5}$$

is an empirical parameter to determine whether divalent or monovalent ions are dominant on duplex stability (in this expression both salt concentrations are given in molar units). If R is less than 0.22, then monovalent ions are dominant and the presence of Mg^{2+} can be ignored. Accordingly, in our experiments the presence of magnesium ions becomes relevant when the concentration of $MgCl_2$ is larger than 0.1 mM.

As in the case of pure monovalent ion conditions, the standard deviation of rupture forces remains almost constant (Fig. 5.12b, right), which suggests that the position of the TS in the unfolding and folding transitions does not depend on the presence of magnesium ions. Finally, unfolding and folding kinetic rates displace towards larger forces along the force-axis with salt concentration, which again suggests the stabilizing effect of salt on the RNA hairpin.

Non-specific mixed monovalent/divalent salt correction for the base-pair free energies for the RNA double helix

In the literature there is no experimental characterization of the effect of mixed monovalent/divalent salt conditions on RNA double helix stability neither on the elastic properties of ssRNA. Nevertheless, there are theoretical models that study the interaction between counterions and nucleic acids, like the CC theory or the TBI model (briefly introduced in the previous section), and that provide estimators for the free energy of formation of nucleic acid hairpins (Appendix G).

In the previous section it was shown that the TBI model gives values for the free energy of formation of CD4-RNA at different pure monovalent ionic conditions that are in excellent agreement with experimental results (Fig. 5.11). Such results motivate the use of the TBI theory to determine the free energy of formation ΔG^0_{NU} of the RNA hairpin at mixed monovalent/Mg^{2+} ionic conditions. Another reason that suggests the use of the TBI model rather than the CC theory is given by the predicted salt-dependence profile of the free energy of formation of CD4-RNA. Using Eq. (5.5) and taking $[Mon^{+}]$=0.05 M, it can be seen in Fig. 5.13 that for magnesium concentrations $[Mg^{2+}]$ whose value of R falls below 0.22 both the CC and the TBI models show no dependence on the concentration. However, for values of R larger that 0.22 the TBI model predicts a stronger dependence on $[Mg^{2+}]$ than the CC theory, in agreement

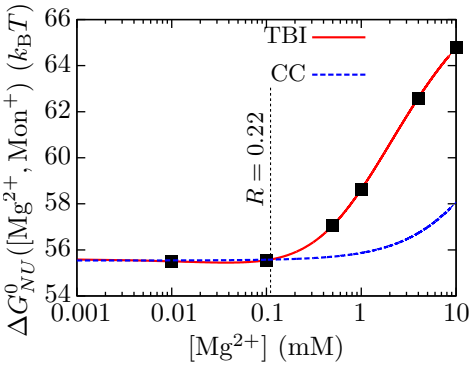

Figure 5.13: Salt-dependence of the free energy of formation of CD4-RNA predicted with the TBI and the CC models. Predictions obtained using the TBI model (straight red) and the CC model (dashed blue) are shown. Black squares are the values of the free energy of formation predicted by the TBI model that has been used to calculate the salt correction $g_2([Mg^{2+}])$. The vertical-dashed line indicates the magnesium concentration at which $R = 0.22$ (Eq. 5.5), taking $[Mon^+]$=0.05 M.

with the experimental results found with DFS experiments for the salt-dependence of unfolding and folding rupture forces (Fig. 5.12).

In the previous section, the effect of monovalent ions on the energetic stability of RNA hairpins was considered to be sequence-independent. The same will be assumed to model the effect of divalent ions. In this way, the sequence-independent correction for a given RNA base pair $g_2([Mg^{2+}])$ due to the presence of divalent ions can be written as:

$$g_2([Mg^{2+}]) = \frac{1}{N}\left(\Delta G^0_{NU} - \Delta G^{0,TBI}_{NU}\right) - g_1([Mon^+]). \tag{5.6}$$

where N is the number of base pairs in the hairpin stem (equal to 20 for CD4-RNA, as shown in Fig. 5.7a); ΔG^0_{NU} is the free energy of formation at 1 M NaCl estimated using the UO set of parameters (summarized in Table 4.1 and taken from [Walt 94, Math 99]); $\Delta G^{0,TBI}_{NU}$ is the free energy of formation obtained using the TBI model at the working mixed monovalent/Mg^{2+} conditions [Tan 07, Tan 08] (red-solid line in Fig. 5.13); and the term $g_1([Mon^+]) = m\log([Mon^+])$ is the sequence-independent correction to the free energy of a single base pair due the presence of monovalent ions (Eq. 5.4). In Table 5.8 the values of $\Delta G^{0,TBI}_{NU}$ and $g_2([Mg^{2+}])$ determined at the mixed monovalent/divalent ionic concentrations experimentally investigated are given.

To sum up, from the sequence-independent corrections to the free energy of formation of a single base pair due to presence of monovalent and divalent ions ($g_1([Mon^+])$ and $g_2([Mg^{2+}])$, respectively) it is possible to determine the free energy of formation of any intermediate configuration n of the RNA hairpin at mixed monovalent/Mg^{2+} ionic conditions,

$$\Delta G^0_n([Mon^+],[Mg^{2+}]) = \Delta G^0_n - (N-n)\left(g_1([Mon^+]) + g_2([Mg^{2+}])\right), \tag{5.7}$$

being ΔG^0_n the free energy of formation of the configuration n at standard 1 M NaCl conditions (Eq. 4.2). This result turns out to be very powerful, because it allows to calculate the theoretical profile of the kinetic barrier $B_{NU}(f)$ for RNA hairpins using the Kramers result (Eq. 4.32) once the elastic properties of ssRNA are known.

$[Mg^{2+}]$ (mM)	$\Delta G^{0,\mathrm{TBI}}_{NU}$ (k_BT)	g_2 (kcal/mol)
0.00	55.58	0.000±0.005
0.01	55.50	-0.007±0.005
0.10	55.55	-0.005±0.005
0.50	57.06	0.0393±0.005
1.00	58.63	0.0858±0.005
4.00	62.60	0.2033±0.005
10.0	64.77	0.2678±0.005

Table 5.8: Sequence-independent free-energy correction of an RNA base pair at mixed monovalent/divalent ionic conditions. The values provided for $g_2([Mg^{2+}])$ are derived using Eq. (5.6). A fixed concentration of 50 mM $[Mon^+]$ is always used and therefore $g_1(0.05) \simeq -0.3$ kcal/mol, according to Eq. (5.4) and taking $m = 0.1$ kcal/mol. Error bars for $g_2([Mg^{2+}])$ have been obtained by using the predicted value of ΔG^0_{NU} determined from the UO set of base-pair free energies or from the experimental determination at 1.05 M $[Mon^+]$ shown in the previous section.

Elastic properties of ssRNA at mixed ionic conditions

In order to determine the elastic properties of ssRNA at any mixed ionic conditions, the matching between the theoretical and experimental estimations of $B_{NU}(f)$ is performed by minimizing a χ^2-function that allows to estimate the persistence length P, together with k_0 (Appendix E). In this case, the value of ΔG^0_{NU} is fixed to the prediction given by the TBI model (Table 5.8).

In Table 5.9 numerical results for the $\log k_0^{NU}$ and P are shown for which the best match is found. The resulting salt-dependent persistence length is plotted in Fig. 5.14. The manifest screening effect of ions on the RNA double helix is observed in the decrease of the contour length with $[Mg^{2+}]$. Resulting profiles of the $B_{NU}(f)$ are shown in Fig. 5.15. Again, it is revealed in the different panels that the height of the kinetic barrier increases with salt concentration for any given force, which indicates the stabilizing effect of counterions on the RNA double helix.

$[Mg^{2+}]$ (mM)	P (nm)	$\log k_0$ (1/s)
0.00	1.27±0.03	12.4±0.4
0.01	1.50±0.15	12.0±0.3
0.10	1.25±0.10	11.4±0.3
0.50	0.90±0.15	11.5±0.3
1.00	0.80±0.10	11.4±0.5
4.00	0.75±0.10	11.1±0.5
10.0	0.75±0.10	10.4±0.5

Table 5.9: ssRNA elastic properties and logarithm of the attempt frequency at zero force for CD4-RNA obtained for experiments at different $[Mg^{2+}]$. Results obtained by matching the theoretical and experimental prediction of the force-dependent profile of the kinetic barrier at each ionic condition.

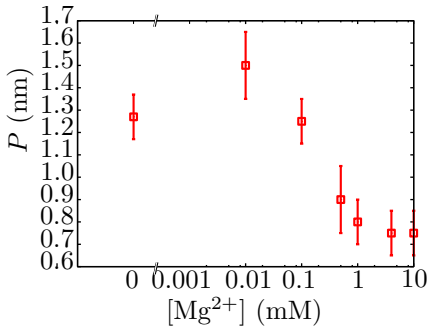

Figure 5.14: Dependence of the persistence length P of ssRNA on $[Mg^{2+}]$. Experimental persistence length versus $[Mg^{2+}]$. The point on the left of all is for pure monovalent salt conditions, *i. e.*, $[Mon^+]$=50 mM.

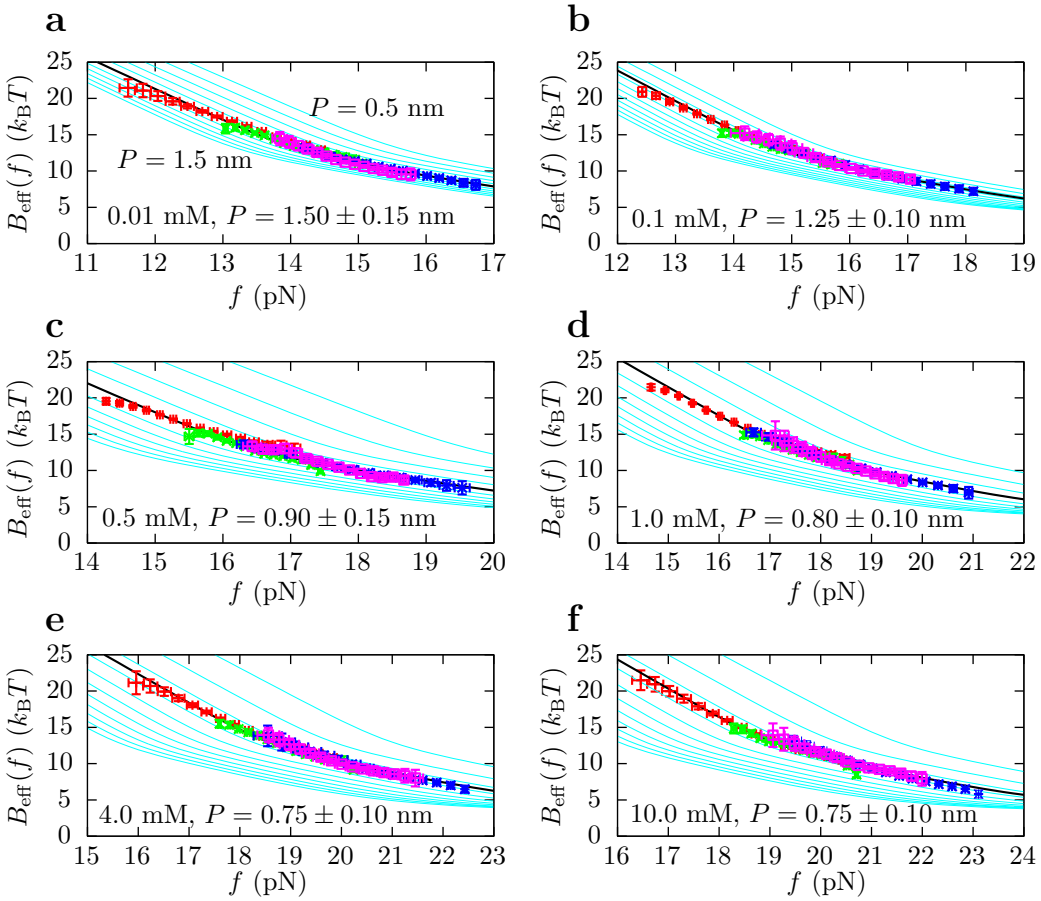

Figure 5.15: Kinetic barrier of CD4-RNA at varying $[Mg^{2+}]$. The experimental profiles of $B_{NU}(f)$ are matched with theoretical predictions obtained for different values of P (0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4 and 1.5 nm, light-blue lines from top to bottom) at 0.01 mM $[Mg^{2+}]$ (**a**), 0.1 mM $[Mg^{2+}]$ (**b**), 0.5 mM $[Mg^{2+}]$(**c**), 1 mM $[Mg^{2+}]$ (**d**), 4 mM $[Mg^{2+}]$ (**e**), and 10 mM $[Mg^{2+}]$ (**f**). For all panels $[Mon^+]$=50 mM. Red (green)/ blue (magenta) points are the experimental estimators $B_{NU}(f)$ determined from folding/unfolding forces at a pulling speeds of 12.5 (1.8) pN/s. Black lines are the theoretical estimators of $B_{NU}(f)$ that match experimental results.

5.2.3 Comparison between monovalent and divalent ionic conditions

The distance of the TS to state N, $x_{N-\rm TS}(f)$, can be calculated from the force derivative of the effective barrier $B_{NU}(f)$ (Eq. 4.19). Figure 5.16 shows the theoretical prediction for the two extreme cases with 50 and 1050 mM $[Mon^+]$, and 0.01 and 10 mM $[Mg^{2+}]$ (black-solid and black-dashed lines in panels a and b, respectively). A reasonable agreement is obtained for all the ionic conditions between theoretical predictions and experimental data. In all the cases, the location of the TS along the FEL changes as a function of force. In both panels of Fig. 5.16 the end-to-end distance of configurations $n = 19$ and $n = 6$ are plotted in gray (solid lines belong to predictions at low salt concentrations whereas dashed lines belong to predictions at high salt concentrations).

At low forces, the TS is located near the loop, at $n = 19 \pm 1$ (gray and black lines in all panels coincide). At intermediate forces that depend on salt concentration the TS moves to the stem region $n = 6 \pm 1$. Finally, at large forces the TS disappears. These results are in agreement with previous findings using the same hairpin sequence [Mano 06]. Noteworthy, it can be observed that the TS mediating unfolding and folding does not depend on salt concentration. That is, despite that rupture forces are shifted to higher values as salt increases for both monovalent and mixed ionic conditions, in all cases the TS is located at $n \simeq 19$, 6 or 0 for low, intermediate and high forces, respectively. These results agree with the Leffler-Hammond postulate [Leff 53, Hamm 55]: at increasing [Mon^+] state N is increasingly stabilized while the TS is shifted towards state U; simultaneously, as force increases the TS approaches state N.

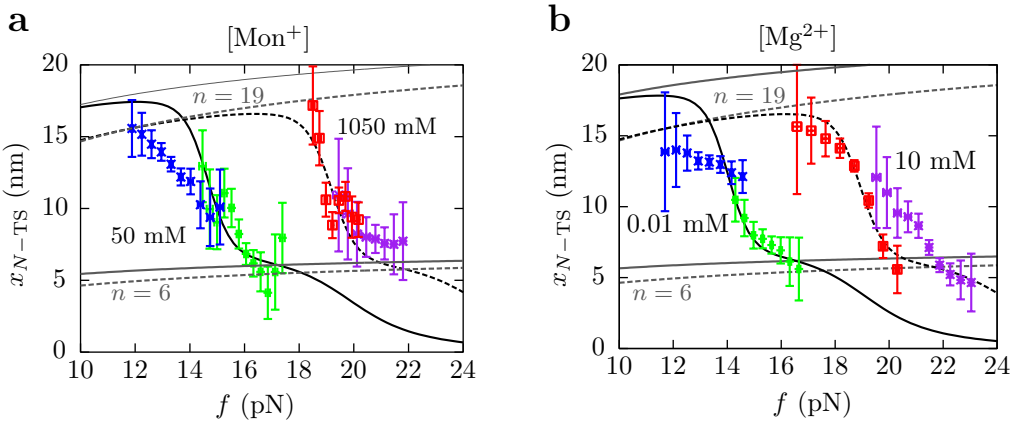

Figure 5.16: Position of the TS for CD4-RNA. The distance $x_{N-\mathrm{TS}}$ between the TS and state N is calculated using Eq. (4.19). Results for pure 50 mM and 1050 mM [Mon^+] (**a**) and mixed 0.01 and 10 mM [Mg^{2+}] plus 50 mM [Mon^+] (**b**) are shown. In bot panels, the solid(dashed)-black line is the theoretical prediction of $x_{N-\mathrm{TS}}$ at low(high) salt concentrations. Points with error bars are the average over estimations of $x_{N-\mathrm{TS}}$ obtained at two different pulling speeds (1.8 and 12.5 pN/s); error bars are standard errors. Solid(dashed)-gray lines are theoretical prediction of the end-to-end distance of hairpin intermediate configurations $n = 19$ and $n = 6$ at low(high) salt concentrations.

The ability of Mg^{2+} ions to stabilize RNA structures at much lower concentrations than monovalent ions was recognized almost 40 years ago [Cole 72]. In fact, the free energies of formation versus salt concentration obtained at monovalent and mixed monovalent/divalent salts collapse into a single master curve after multiplying [Mg^{2+}] by a factor 100 (Fig. 5.17a). This effect has been previously observed and can be explained using the CC theory [Heil 01, Tan 09, Schr 00]. A similar trend (but not a data collapse) is obtained for the persistence length values of ssRNA (Fig. 5.17b). It can be interpreted as the screening effect of the counterions present in solution: the shorter the screening length, the stiffer the molecule. This equivalence found between the non-specific binding of monovalent and divalent salts with nucleic acids might be useful to develop new biochemical assays in situations where divalent ion-specific interactions could be detrimental.

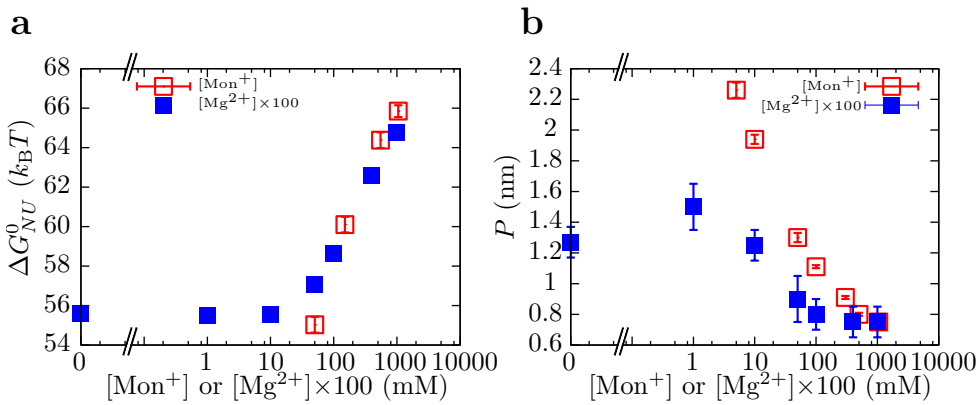

Figure 5.17: Comparison between [Mon$^+$] and [Mg^{2+}] results. a. Free energy of formation of the RNA hairpin at different salt conditions. **b.** Persistence length values for the ssRNA hairpin at different salt conditions. Magnesium concentrations have been multiplied by 100 along the horizontal axis.

5.3 Dynamic properties of intermediate states

Until now only two-state DNA and RNA hairpins have been deeply characterized using DFS techniques. However, biomolecules are not always structurally that simple and sometimes they present intermediate configurations and misfolded states that might be crucial for the their correct or incorrect performance. These states are usually metastable, in the sense that they have short lifetimes under equilibrium conditions. When generated dynamically, *i. e.* under non-equilibrium conditions, these states are called *kinetic states*.

In calorimetry or UV absorbance experiments metastable states are hard to observe because output signals are the results of an average over all the configurations in equilibrium, and therefore short-lifetime states usually get masked. Consequently, the characterization of such states is a challenging problem in equilibrium conditions.

In contrast, non-equilibrium experiments can favor the observation of kinetic states. With the proper protocol, it is possible to dynamically generate them and characterize their FEL. In this section, DFS experiments will be performed on DNA hairpins that have been designed to contain one or more kinetic states on-pathway. Therefore, when mechanically unfolding or folding such hairpins these states will be observed between the unfolded and the native conformations, and hence the name of "intermediates". In what follows it will be shown how to extend the DFS analysis in order to unravel kinetic and thermodynamic properties of such intermediate states.

5.3.1 A single intermediate state

A DNA hairpin, named hereafter as I1, with an internal loop was designed and synthesized (details of the synthesis are provided in Appendix A). Its sequence and secondary structure are provided in Fig. 5.18a. The theoretical FEL evaluated at f_c shows an intermediate state, I, at $n = 11 - 14$, which corresponds to the configuration where the

base pairs of the stem that precede the internal loop are open (Fig. 5.18b). A scheme of the predicted unfolding and folding pathway of the hairpin is shown in Fig. 5.18c.

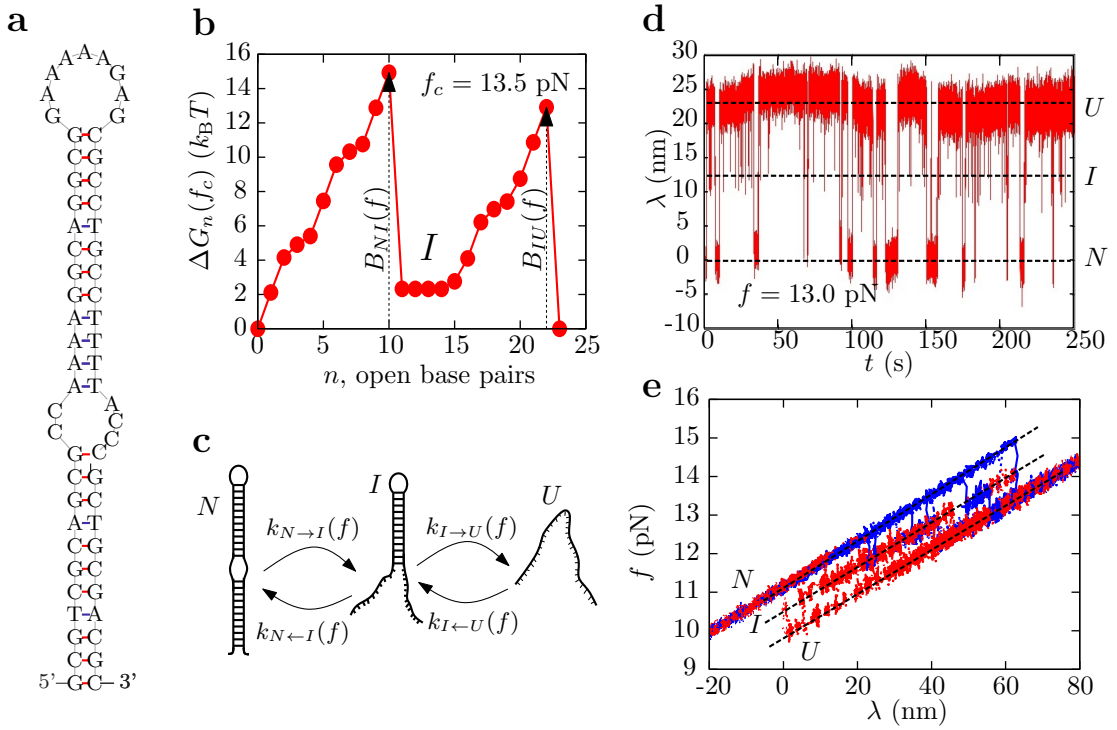

Figure 5.18: Molecule I1 with one intermediate state. **a.** Sequence and secondary structure. **b.** Predicted FEL computed at f_c = 13.5 pN using the NN model with the UO set of base-pair free energies [Sant 98]. **c.** Scheme of the unfolding and folding pathway of I1. **d.** Constant force experiments. The molecule executes transitions between N (at $\lambda \sim 0$ nm), I (at $\lambda \sim$13 nm) and U (at $\lambda \sim$23 nm). **e.** Collection of unfolding (blue) and folding (red) curves measured in pulling experiments carried out at 60 nm/s. Three branches of force are observed, each one assigned to one molecular state: N, I or U.

The intermediate state I of hairpin I1 is observed in DFS experiments. In constant-force experiments performed at low (high) forces only state N (U) is visible. However, at forces close to f_c (equal to 13.5 pN, according to predictions obtained using the NN model with the UO base-pair free energies) the hairpin stochastically jumps between states N, I and U, as can be seen in Fig. 5.18d. A similar behavior is observed in pulling experiments: three branches of force are measured along the FDC. At low forces the hairpin is folded and as force increase the hairpin first reaches the intermediate configuration before complete unfolding at large forces (Fig. 5.18e).

In pulling experiments each state has its own survival probability $P_N(f)$, $P_I(f)$ and $P_U(f)$, respectively. In order to obtain their force-dependent profile along the unfolding process the following steps must be performed:

1. A threshold force $f_{\rm th}$ is defined at which the molecule can be found at any of the three states N, I or U for different values of λ. In Fig. 5.19 two values of $f_{\rm th}$ are

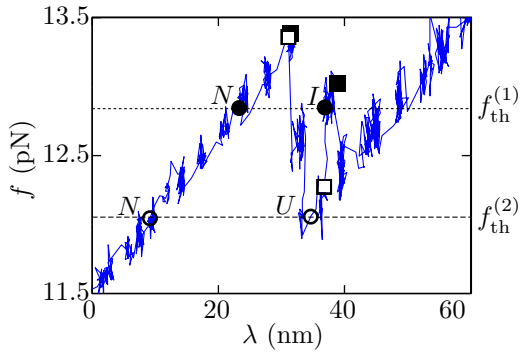

Figure 5.19: Measurement of survival probabilities for I1 along the unfolding process. An unfolding trajectory (blue) and two examples of threshold forces, $f_{\rm th}^{(1)}$ and $f_{\rm th}^{(2)}$ (dotted and dashed lines, respectively), are shown. For the given trajectory, the molecule is found at states N or I (solid circles) at $f_{\rm th}^{(1)}$, and at states N or U at $f_{\rm th}^{(2)}$ (empty circles). Unfolding forces for $f_{\rm th}^{(1)}$ are $f_U^{(N)} = 13.4$ pN and $f_U^{(I)} = 13.0$ pN (solid squares), whereas for $f_{\rm th}^{(2)}$ they are $f_U^{(N)} = 13.4$ pN and $f_U^{(U)} = 12.3$ pN (empty squares).

represented. For the trajectory shown, the molecule is found at states N or I at $f_{\rm th}^{(1)}$ (solid circles), and at states N or U at $f_{\rm th}^{(2)}$ (empty circles).

2. Unfolding forces $f_U^{(N)}$, $f_U^{(I)}$ and $f_U^{(U)}$ are defined as the first rupture forces leaving states N, I and U, respectively, that are larger than $f_{\rm th}$. In the case of $f_U^{(N)}$ ($f_U^{(U)}$) events, these correspond to the transition of state N (U) toward I. In the case of $f_U^{(I)}$ two different molecular transitions can take place: $I \to U$ and $N \leftarrow I$. In the example provided in Fig. 5.19, unfolding forces for $f_{\rm th}^{(1)}$ are $f_U^{(N)} = 13.4$ pN and $f_U^{(I)} = 13.0$ pN (solid squares), whereas for $f_{\rm th}^{(2)}$ they are $f_U^{(N)} = 13.4$ pN and $f_U^{(U)} = 12.3$ pN (empty squares).

3. For a given value of $f_{\rm th}$ and the corresponding set of values for $f_U^{(N)}$, $f_U^{(I)}$ and $f_U^{(U)}$ measured in the whole set of unfolding trajectories, the survival probabilities along unfolding experiments can be measured as:

$$P_N(f) = 1 - \frac{N\left(f_U^{(N)} < f\right)}{N_U^{(N)}} \tag{5.8a}$$

$$P_I(f) = 1 - \frac{N\left(f_U^{(I)} < f\right)}{N_U^{(I)}} \tag{5.8b}$$

$$P_U(f) = 1 - \frac{N\left(f_U^{(U)} < f\right)}{N_U^{(U)}}, \tag{5.8c}$$

where $N\left(f_U^{(i)} < f\right)$ (with $i = N, I$ or U) is the number of unfolding force events where $f_U^{(i)}$ is smaller than f and $N_U^{(i)}$ is the total number of times state i is observed at $f_{\rm th}$. Even though $P_N(f)$ and $P_I(f)$ will always be equal to zero at large forces, this is not necessarily true for $P_U(f)$. By repeating this procedure for different values of $f_{\rm th}$ it is possible to obtain the force-dependence of the survival probabilities for a wide range of forces.

The same procedure is used to analyze folding trajectories. Now, three folding forces $f_F^{(U)}$, $f_F^{(I)}$ and $f_F^{(N)}$ are defined as the first folding forces that leave state U, I or N, respectively, along a folding trajectory and that are smaller than a given threshold force $f_{\rm th}$. The survival probabilities along folding experiments are calculated according to:

$$P_i(f) = 1 - \frac{N\left(f_F^{(i)} > f\right)}{N_F^{(i)}},\ i = U,\ I,\ N, \tag{5.9}$$

being $N\left(f_F^{(i)} > f\right)$ (with $i = U, I$ or N) the number of folding force events where $f_F^{(i)}$ is larger than f and $N_F^{(i)}$ is the total number of times state i is seen at $f_{\rm th}$.

As for two-state DNA hairpins (Eq. 4.22), survival probabilities for I1 are described by first order Markov processes. The master equations that model the kinetics of I1, according to the scheme shown in Fig. 5.18c, are:

$$\frac{dP_N(f)}{df} = -\frac{k_{N\to I}(f)}{r} P_N(f) \tag{5.10a}$$

$$\frac{dP_I(f)}{df} = -\frac{k_{N\leftarrow I}(f) + k_{I\to U}(f)}{r} P_I(f) \tag{5.10b}$$

$$\frac{dP_U(f)}{df} = -\frac{k_{I\leftarrow U}(f)}{r} P_U(f)\ \ , \tag{5.10c}$$

where $k_{N\to I}(f)$ and $k_{I\to U}(f)$ ($k_{N\leftarrow I}(f)$ and $k_{I\leftarrow U}(f)$) are the kinetic rates along the unfolding (folding) pathway. Therefore, once the survival probabilities are experimentally calculated it is straightforward to determine $k_{N\to I}(f)$, the sum $k_{N\leftarrow I}(f) + k_{I\to U}(f)$, and $k_{I\leftarrow U}(f)$ from Eqs. (5.10a), (5.10b) and (5.10c), respectively. The decoupling of the kinetic rates $k_{N\leftarrow I}(f)$ and $k_{I\to U}(f)$ is carried out by the following equation:

$$\frac{k_{I\to U}(f)}{k_{N\leftarrow I}(f)} = \frac{\phi_{I\to U}(f)}{\phi_{N\leftarrow I}(f)}, \tag{5.11}$$

where $\phi_{I\to U}(f)$ and $\phi_{N\leftarrow I}(f)$ are the fraction of rupture forces in the range $[f, f + df]$ associated with transitions of the type $I \to U$ and $N \leftarrow I$, respectively. These fractions can be experimentally determined from the FDCs. Therefore, the four kinetic rates associated to the unfolding and folding pathway of I1 can be extracted from pulling experiments.

Figure 5.20a shows the estimated kinetic rates obtained from pulling experiments using Eqs. (5.8)–(5.11) (empty symbols) and from hopping experiments at constant force, using the inverse of the lifetime of each state [Forn 11] (solid symbols). The agreement is good and results for each kinetic rate collapse into a single master curve. Again, the range of forces at which kinetic rates are measurable is wider in pulling experiments than in equilibrium experiments. This is due to hysteresis effects: in non-equilibrium pulling experiments molecular transitions can occur at forces at which one would hardly populate the three different states under equilibrium conditions at reasonable timescales.

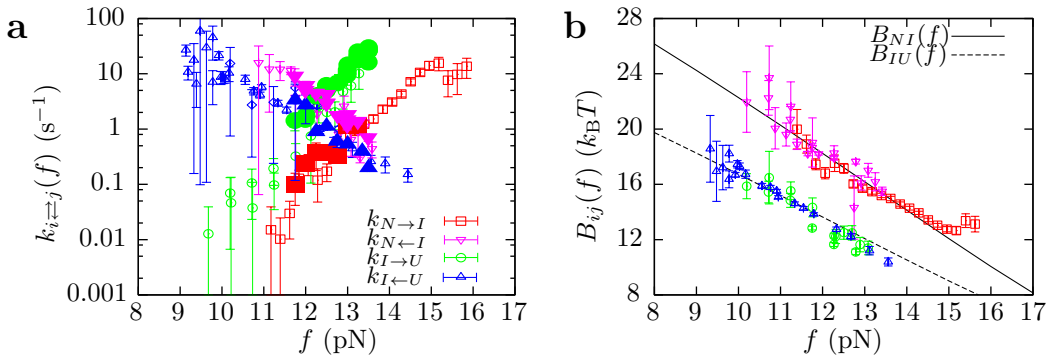

Figure 5.20: Molecule I1 with one intermediate state. a. Kinetic rates $k_{N\rightarrow I}(f)$ (red squares), $k_{I\rightarrow U}(f)$ (green circles), $k_{N\leftarrow I}(f)$ (blue up-pointing triangles) and $k_{I\leftarrow U}(f)$ (purple down-pointing triangles), extracted from pulling experiments (empty symbols) and equilibrium hopping experiments at constant force (solid symbols). **b.** Comparison between the experimentally measured kinetic barriers $B_{NI}(f)$ and $B_{IU}(f)$ and the theoretical prediction obtained using Eq. (4.32) and the NN model with the UO set of base-pair free energies [Sant 98].

The existence of state I enforces the presence of two TS, as already revealed in the FEL for molecule I1 (dotted arrows in Fig. 5.18b). $B_{NI}(f)$ mediates the transition between N and I, and $B_{IU}(f)$ between I and U. Both $B_{NI}(f)$ and $B_{IU}(f)$ can be theoretically estimated using the Kramers solution given by Eq. (4.32), where the first sum runs over sequential configurations between states N and I, and I and U, respectively. Moreover, it is possible to experimentally measure the force-dependence of the two kinetic barriers by using the following expressions for the kinetic rates:

$$k_{N\rightarrow I}(f) = k_0^{NI} \exp\left(-\frac{B_{NI}(f)}{k_\mathrm{B}T}\right) \tag{5.12a}$$

$$k_{N\leftarrow I}(f) = k_0^{NI} \exp\left(-\frac{B_{NI}(f) - \Delta G_{NI}(f)}{k_\mathrm{B}T}\right), \tag{5.12b}$$

and

$$k_{I\rightarrow U}(f) = k_0^{IU} \exp\left(-\frac{B_{IU}(f)}{k_\mathrm{B}T}\right) \tag{5.13a}$$

$$k_{I\leftarrow U}(f) = k_0^{IU} \exp\left(-\frac{B_{IU}(f) - \Delta G_{IU}(f)}{k_\mathrm{B}T}\right). \tag{5.13b}$$

Here, k_0^{NI} and k_0^{IU} are the attempt frequencies at zero force between states N and I, and I and U; and $\Delta G_{NI}(f)$ and $\Delta G_{IU}(f)$ are the free-energy differences between states N and I, and I and U, at force f. From here on, one can proceed as it was done for two-state nucleic acid hairpins by matching the experimental and the theoretical prediction of each kinetic barrier $B_{NI}(f)$ and $B_{IU}(f)$ using the elastic properties derived for ssDNA in Chapter 3 ($P = 1.35 \pm 0.05$ nm). Results are shown in Fig. 5.20b. The values for the attempt frequencies that best fit the experimental data are $\log k_0^{NI} = 14.4 \pm 0.5$ and $\log k_0^{IU} = 13.2 \pm 0.5$, or equivalently, $k_0^{NI} = (1.1 \pm 0.5) \times 10^7$ s^{-1} and

$k_0^{IU} = (5.4 \pm 0.4) \times 10^6$ s^{-1}. These results are in the range of values found for previous estimations of the attempt frequency for two-states nucleic acid hairpins. To finish, the values for the free-energy differences ΔG^0_{NI} and ΔG^0_{IU} that best fit the results are summarized in Table 5.10, where it can be seen that a good agreement is found between theoretical predictions and experimental results.

	Experiments	NN model
ΔG^0_{NI} ($k_{\mathrm{B}}T$)	30±2	30±1
ΔG^0_{IU} ($k_{\mathrm{B}}T$)	27±3	28±1

Table 5.10: Free-energy differences at zero force for I1. Experimental values were obtained by matching the experimental and the theoretical profile of the kinetic barriers $B_{NI}(f)$ and $B_{IU}(f)$ shown in Fig. 5.20b. Theoretical values are obtained by averaging predicted values from the NN model using the UO and the UNZ set of base-pair free energies [Sant 98, Hugu 10b]. Error bars are standard errors.

5.3.2 Several molecular intermediate states

The DFS analysis presented in Section 5.3.1 can be extended to analyze molecules with several molecular intermediate states. For example, for a given molecule with K intermediate states, the unfolding or folding pathway can be written as:

$$N \rightleftarrows I_1 \rightleftarrows I_2 \rightleftarrows \cdots \rightleftarrows I_{K-1} \rightleftarrows I_K \rightleftarrows U$$

The survival probabilities of each state $P_N(f)$, $P_k(f)$ $(k = 1, \ldots K)$, $P_U(f)$ satisfy the following set of differential equations:

$$\frac{dP_N(f)}{df} = -k_{N\to I_1}(f) P_N(f) \tag{5.14a}$$

$$\frac{dP_{I_k}(f)}{df} = -\left(k_{I_k\to I_{k+1}}(f) + k_{I_{k-1}\leftarrow I_k}(f)\right) P_{I_k}(f) \quad \text{for } k = 1, \ldots, K \tag{5.14b}$$

$$\frac{dP_U(f)}{df} = -k_{I_K\leftarrow U}(f) P_U(f), \tag{5.14c}$$

where $I_{K+1} = U$ and $I_0 = N$. Now, assuming the existence of a kinetic barrier $B_{ij}(f)$ between any sequential states i and j $(i, j = N, I_k, U;\ k = 1, \ldots, K)$, it is possible to generalize the expressions for the kinetic rates $k_{i\rightleftarrows j}(f)$ as:

$$k_{i\to j}(f) = k_0^{ij} \exp\left(-\frac{B_{ij}(f)}{k_{\mathrm{B}}T}\right) \tag{5.15a}$$

$$k_{i\leftarrow j}(f) = k_0^{ij} \exp\left(-\frac{B_{ij}(f) - \Delta G_{ij}(f)}{k_{\mathrm{B}}T}\right), \tag{5.15b}$$

where k_0^{ij} is the attempt frequency at zero force between states i and j, and $\Delta G_{ij}(f)$ is the free-energy difference between states j and i at force f.

To prove the validity of this extension, results obtained with a DNA three-way junction are presented below. This molecule consists of a stem of 21 base pairs that, instead

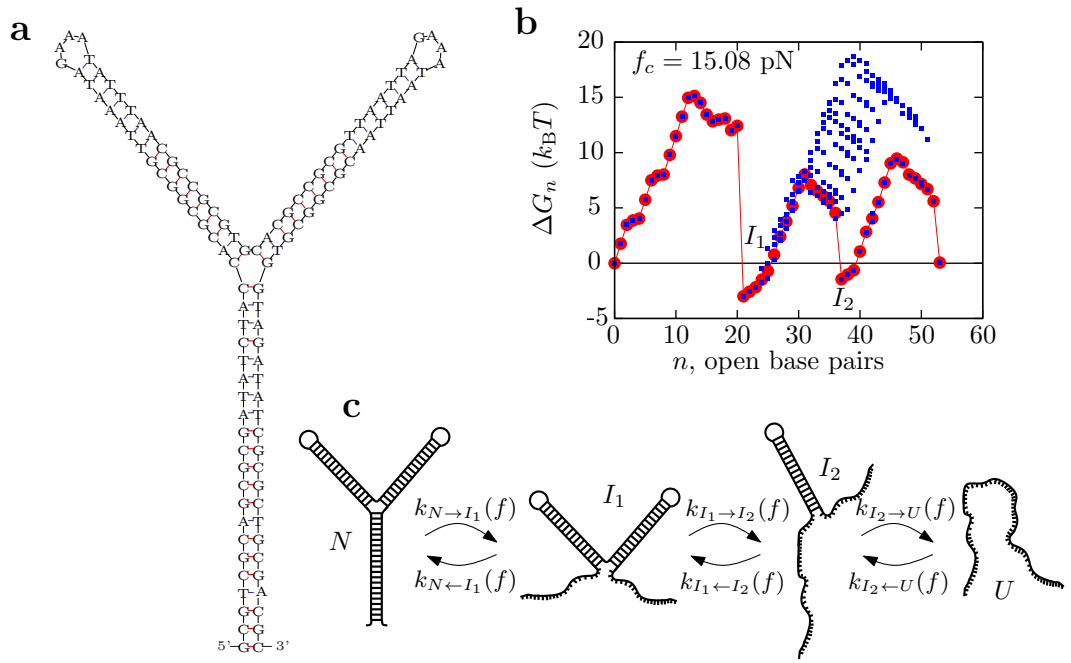

Figure 5.21: Molecule I2 with two intermediate states. a. Sequence and secondary structure. **b.** FEL calculated at the coexistence force (15.08 pN) predicted by the NN model using the UO parameters. Blue squares are the free energies of all the possible sequential configurations that have a given number n of unzipped base pairs distributed between the two hairpins. Red circles are the Boltzmann average (*i. e.* the potential of mean force) taken by summing over all configurations constrained by a given number of unzipped base pairs n. **c.** Scheme of the unfolding and folding pathway. I_1 corresponds to the state where all the stem is open and the two hairpins of the bifurcation are folded ($n = 21$). I_2 corresponds to two identical structures, where all the stem and one hairpin in the bifurcation are open ($n = 37$).

of a single loop at its end, has a multi-branched loop (Fig. 2.3) made of two DNA hairpins of 16 base pairs each (Fig. 5.21a). This molecule has two intermediate states, and it is given the name of I2.

In order to obtain the FEL of such molecule, the total number of open base pairs n is used as the reaction coordinate. Once the initial stem is completely unfolded (*i. e.*, for $n > 21$) several configurations are compatible with a given n because the total number of open base pairs distribute between the two hairpins of the multi-branched loop. That is, if n_1 and n_2 are the number of open base pairs in each of the two hairpins, compatible configurations satisfy $n_1 + n_2 = n$. Blue squares in Fig. 5.21b are the free-energy values for all the possible configurations for the different values of n. The potential of mean force, or FEL, for I2 is calculated as the Boltzmann average taken over all the configurations constrained by the given number of unzipped base pairs n (Fig. 5.21b, red circles). Two intermediates, hereafter referred as I_1 and I_2, appear along the unfolding pathway as revealed by the FEL calculated at f_c (Fig. 5.21b-c). I_1 is found at $n = 21$ and corresponds to a structure where all the stem is open and the two hairpins of the bifurcation are closed. I_2 is found at $n = 37$, where several configurations are

possible but the two states that most contribute to the Boltzmann average are identical structures, where all the stem and one hairpin in the bifurcation are open. Since these two configurations have identical molecular extension, I_2 is treated as a single state.

The four molecular states N, I_1, I_2 and U are observed in both pulling and equilibrium experiments at constant force (Fig. 5.22). However, hopping between the four conformations is never observed at the accessible experimental timescales, which precludes the measurement of the kinetic rates $k_{i \rightleftarrows j}(f)$. Nonetheless, kinetic rates measurements are straightforward possible by doing pulling experiments and using Eq. (5.14) (Fig. 5.23a).

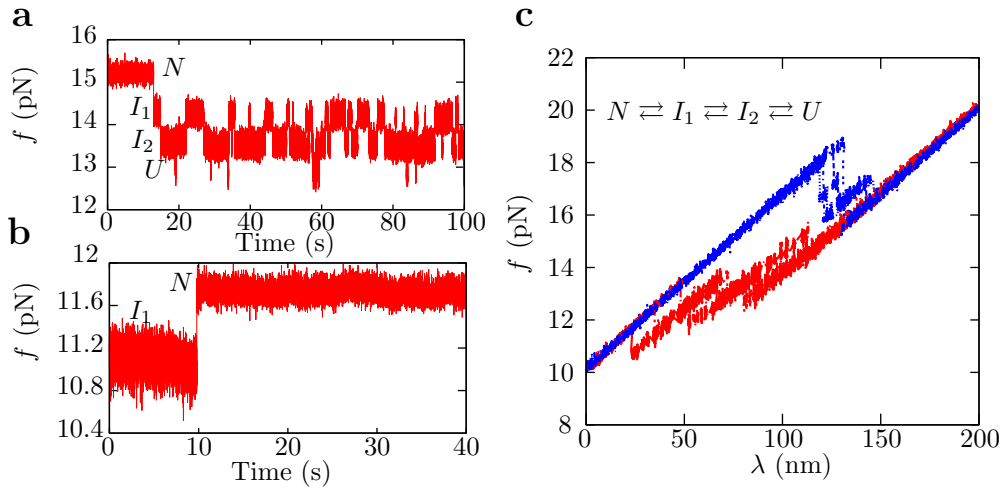

Figure 5.22: Molecule I2 with two intermediate states. a. Example of a passive hopping trace measured at high forces: the molecule is initially set to N. Once the molecule partially unfolds (after ∼15 in the figure), the system subsequently hops between states I_1, I_2 and U, but never folds back to N. **b.** Example of a hopping experiment at low forces: the molecule is initially set to I_1 (it was not possible to stabilize the molecule in neither U nor I_2 states at such low forces) and once the molecule folds to N it remains there. Equilibrium hopping experiments were repeated at different forces and the native state was never observed to coexist with any other state. **c.** Collection of unfolding (blue) and folding (red) curves measured in pulling experiments carried out at 60 nm/s.

Theoretical predictions for the three different kinetic barriers $B_{NI_1}(f)$, $B_{I_1I_2}(f)$ and $B_{I_2U}(f)$ are obtained from Eq. (4.32), where the first sum runs over sequential configurations between states N and I_1, I_1 and I_2, and I_2 and U, respectively. The resulting profiles are shown in Fig. 5.23b as black-solid, dashed, and dotted lines. The barriers $B_{I_1I_2}(f)$ and $B_{I_2U}(f)$ are almost identical. From the best matches between theoretical and experimental profiles of the different kinetic barriers, estimates for the attempt frequencies k_0^{ij} and free-energy differences at zero force ΔG_{ij}^0 are obtained. Results are shown in Fig. 5.23b and summarized in Table 5.11. Again, good agreement is found between theoretical and experimental values of the free-energy differences.

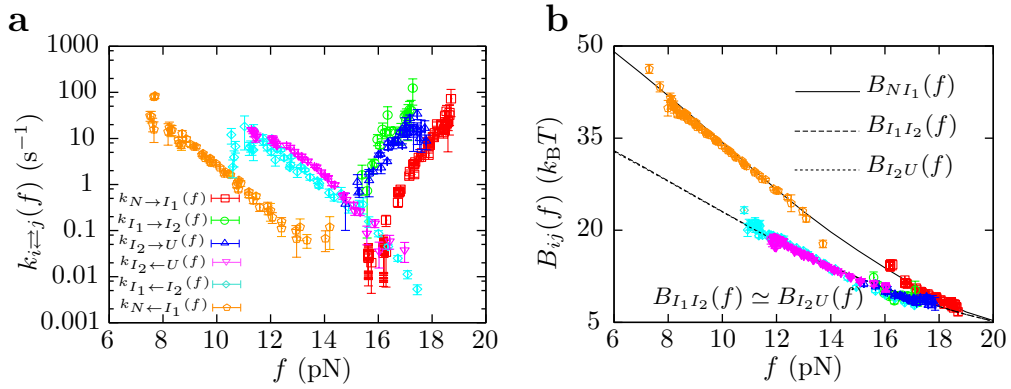

Figure 5.23: Molecule I2 with two intermediate states. a. Unfolding and folding kinetic rates between different states, $k_{i\rightleftarrows j}(f)$ ($i = N,\ I_1,\ I_2$ and $j = I_1,\ I_2,\ U$), extracted from pulling experiments. **b.** Experimentally measured kinetic barriers $B_{ij}(f)$ and the theoretical prediction obtained using Eq. (4.32) and the NN model with the UO set of base-pair free energies [Sant 98].

	$\log k_0^{ij}$	k_0^{ij} (s^{-1})	ΔG_{ij}^0 (k_BT)	
			Experiments	NN model
$i = N,\ j = I_1$	11.0±0.5	$(6 \pm 1) \times 10^5$	57±4	52±2
$i = I_1,\ j = I_2$	12.0±0.5	$(2 \pm 1) \times 10^6$	38±3	40±2
$i = I_2,\ j = U$	12.0±0.5	$(9 \pm 2) \times 10^5$	39±2	40±2

Table 5.11: Kinetic analysis for hairpin I2. Estimation of the attempt frequencies and free-energy differences at zero force, k_0^{ij} and $\Delta G_{ij}^0 = \Delta G_j^0 - \Delta G_i^0$ respectively, obtained by matching the experimentally measured $B_{ij}(f)$ with theoretical predictions computed using Eq. (4.32) and the NN model. Theoretical values for ΔG_{ij} are the average over estimations obtained using the UO and the UNZ set of base-pair free energies (Table 4.1) [Sant 98, Hugu 10b].

5.4 Specific binding of a peptide intercalating DNA

In Section 5.2 the non-specific effect of Na^+ and Mg^{2+} on an RNA hairpin was studied using DFS techniques [Biza 12]. In this section a more challenging situation is addressed: using DFS methods, the specific binding that occurs between a DNA hairpin and a peptide that intercalates dsDNA is characterized. To this end, hairpin CD4 (Fig. 4.2a) is pulled with the mini-tweezers instrument in a buffer solution containing Thiocoraline (hereafter, concentrations will be referred to as [Thio]), which is the dsDNA-intercalating peptide under study [Erba 99, Negr 07]. DFS experiments are carried out at 1 M NaCl concentration and at varying concentrations of Thiocoraline (10, 50, 100 and 200 nM [Thio]). The pulling speed is always kept at 180 nm/s.

In Fig. 5.24a, experimental FDC measured with standard pulling experiments performed in the absence of intercalator (inset) and at 200 nM [Thio] (main panel) are shown. It is observed that in presence of Thiocoraline the two-state behavior of CD4 is preserved since the hairpin is either folded or unfolded. Moreover, Thiocoraline does not have any effect on the measured values of folding forces. In contrast, in presence of intercalator, unfolding forces are occasionally much larger than the ones measured in the

absence of peptide. The unfolding and folding force histograms, $\rho(f_U)$ and $\rho(f_F)$ respectively, obtained at different concentrations are shown in Fig. 5.24b. It can be observed that $\rho(f_F)$ does not depend on the concentration, whereas above 10 nM a bimodal distribution is obtained for $\rho(f_U)$. In this latter case, the peak at low forces coincides with the peak of $\rho(f_U)$ measured in the absence of intercalator, which suggests the absence of binding in such rupture events; the peak found at high forces is attributed to events where Thiocoraline specifically binds to the folded DNA hairpin. Therefore, the double helix is stabilized upon binding of Thiocoraline, since unfolding forces are higher than the ones reported in the absence of peptide [Koir 11].

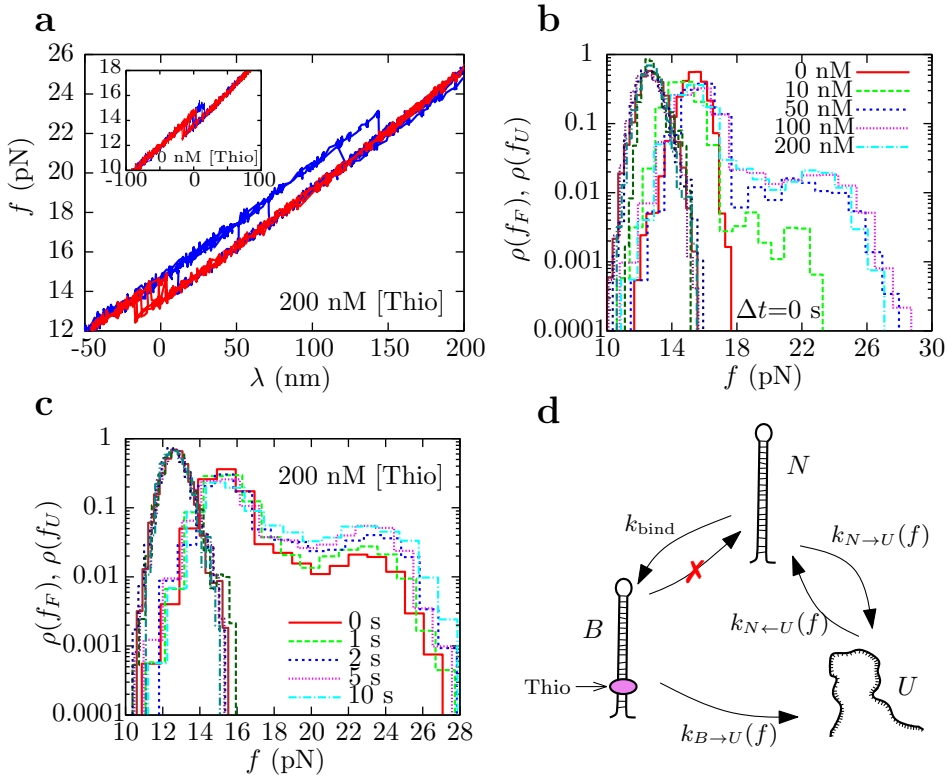

Figure 5.24: Thiocoraline binds dsDNA. a. Pulling experiments carried out without (inset) and in a buffer solution of 200 nM [Thio] (main panel). Folding (unfolding) traces are shown in red (blue). **b.** Rupture force distributions obtained at different concentrations of intercalator, from 0 to 200 nM [Thio]. Light(dark) colors refer to unfolding(folding) forces distributions $\rho(f_U)(\rho(f_F))$. **c.** Unfolding (light colors) and folding (dark colors) force distributions measured at 200 nM [Thio] for different pause intervals at the minimum force of the pulling protocol, from 0 to 10 s. **d.** Scheme of the unfolding and folding reaction pathways. The hairpin can be found at states N (folded without any bound peptide), B (folded with a bound peptide, represented as a purple circle), or U (unfolded).

Additional pulling experiments are performed where the hairpin is kept in the folded state at the minimum force of the protocol for a given period of time Δt, between 0 and 10 seconds. Hence, a time interval is introduced between the folding process and the subsequent unfolding one. In Fig. 5.24c the resulting unfolding and folding

force histograms measured at 200 nM [Thio] are shown for different intervals of time. As before, the time pause does not have any effect on $\rho(f_F)$. In contrast, since the binding probability increases with the time the hairpin spends in state N, the force peak measured at large forces in $\rho(f_U)$ becomes larger as the pause increases.

The unfolding/folding pathway along pulling experiments can be modeled as sketched in Fig. 5.24d: there are two different folded states for the DNA hairpin, which are N (native state without bound peptide) and B (a peptide, plotted as a purple circle in the figure, is bound to the hairpin double helix). The hairpin can only access state B through N with a binding rate k_{bind}, which is proportional to the concentration of Thiocoraline, $k_{\text{bind}} = k_{\rightarrow}[\text{Thio}]$, being $k_{\rightarrow}$ the kinetic rate of the binding of a Thiocoraline molecule. The binding process is considered to be independent of the force applied. This is justified because force does not appreciably distort the shape of the hairpin double helix. Next, Thiocoraline can spontaneously unbind from the hairpin with an unbinding rate $k_{\leftarrow}$, which is independent of [Thio]. However, at the experimentally accessed time-scales, no unbinding events were observed in the absence of applied force and, thus, this transition can be neglected (*i. e.,* $k_{\text{bind}} \gg k_{\leftarrow}$). Under the action of a mechanical force, both states N and B can access state U with the force-dependent kinetic rates $k_{N\rightarrow U}(f)$ and $k_{B\rightarrow U}(f)$, respectively. Finally, from the unfolded-stretched conformation U the hairpin can only get to state N. The master equations describing the survival probabilities of states N, B and U ($P_N(f)$, $P_B(f)$ and $P_U(f)$, respectively) according to such reaction pathway are:

$$\frac{dP_N(f)}{df} = -\frac{k_{N\rightarrow U}(f) + k_{\text{bind}}}{r} P_N(f) \tag{5.16a}$$

$$\frac{dP_B(f)}{df} = -\frac{k_{B\rightarrow U}(f)}{r} P_B(f) \tag{5.16b}$$

$$\frac{dP_U(f)}{df} = -\frac{k_{N\leftarrow U}(f)}{r} P_U(f). \tag{5.16c}$$

In pulling experiments, state U is unambiguously identified. All folding forces correspond to the folding transition $U \rightarrow N$ and it is possible to estimate $P_U(f)$ and $k_{N\leftarrow U}(f)$ using Eqs. (4.20b) and (5.16c), respectively. Since the presence of Thiocoraline in the buffer solution does not interfere with the folding of the DNA hairpin, $k_{N\leftarrow U}(f)$ is expected to be independent of the concentration of peptide (Fig. 5.25).

On the other hand, states N and B look identical along the FDC. In order to characterize the unfolding of the hairpin a joint folded state, NB, made of the union of states N and B is defined. Its survival probability along the unfolding pathway, $P_{NB}(f)$, satisfies the following first-order Markov process:

$$\frac{dP_{NB}(f)}{df} = -\frac{k_{NB\rightarrow U}(f)}{r} P_{NB}(f), \tag{5.17}$$

where $k_{NB\rightarrow U}(f)$ is the unfolding kinetic rate of the joint state NB. The survival probability $P_{NB}(f)$ can be calculated from the whole set of measured unfolding forces using

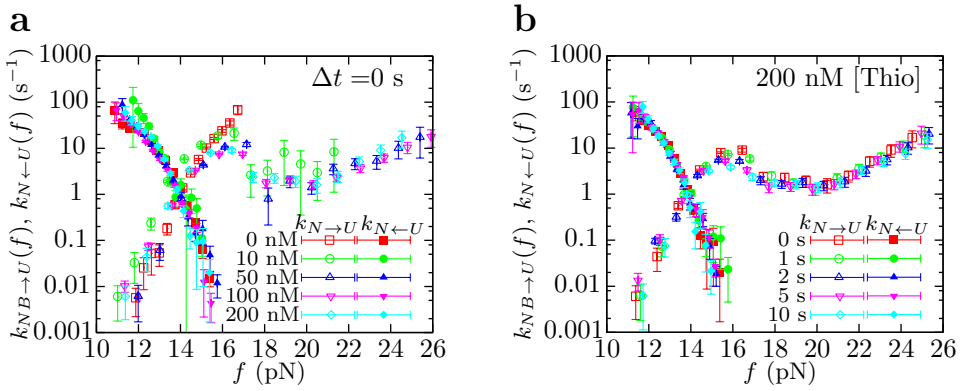

Figure 5.25: Unfolding and folding kinetic rates, $k_{NB\to U}(f)$ and $k_{N\leftarrow U}(f)$. a. Results obtained at different concentrations of Thiocoraline and with no time pause between folding and subsequent unfolding events. **b.** Results obtained at 200 nM [Thio] and at different intervals of time Δt between folding and subsequent unfolding events. In both panels, open (solid) symbols refer unfolding (folding) kinetic rates.

Eq. (4.20a), and from simple mathematical manipulation of Eq. (5.17) it is straightforward to estimate $k_{NB\to U}(f) = -r\frac{dP_{NB}(f)}{df}/P_{NB}(f)$.

The resulting force-dependent profiles for $k_{NB\to U}(f)$ and $k_{N\leftarrow U}(f)$ are shown in Fig. 5.25a. These have been obtained from standard pulling experiments performed at 180 nm/s (no time pause Δt is introduced between the folding and subsequent unfolding processes in the example shown), and at different concentrations of peptide. Results of the same kinetic rates obtained for different durations of the time pause are shown in Fig. 5.25b. Two main features must be emphasized:

i. As expected, neither the presence of intercalator nor the duration of the time pause does affect the force-dependence of $k_{N\leftarrow U}(f)$. Since Thiocoraline only binds to dsDNA, it does not interfere with the folding of the hairpin.

ii. The unfolding kinetic rate $k_{NB\to U}(f)$ does not satisfy the Leffler-Hammond rule, since it decreases with force at the range between 16 and 20 pN. This is because the process mediated by $k_{NB\to U}(f)$ is made of two different reaction pathways ($N \to U$ and $B \to U$), while the Leffler-Hammond postulate is only valid along a single reaction pathway. Hence, its violation under DFS experiments suggests the existence of multiple unfolding pathways.

Now, $k_{NB\to U}(f)$ can be decomposed into the individual kinetic rates of unfolding of N and B (*i. e.*, $k_{N\to U}(f)$ and $k_{B\to U}(f)$ respectively) by:

$$k_{NB\to U}(f) = \phi(f)k_{N\to U}(f) + (1-\phi(f))k_{B\to U}(f), \tag{5.18}$$

being $\phi(f)$ the probability of the folded hairpin to be in N. To find a theoretical expression for $\phi(t)$ it is assumed that binding kinetics are independent of force. Therefore, the time-dependent probability $\phi(t)$ satisfies the following differential equation:

$$\frac{d\phi(t)}{dt} = -k_{\text{bind}}\phi(t), \tag{5.19}$$

where k_{bind} has been previously defined as the kinetic rate of binding of the intercalator (Fig. 5.24d) and where spontaneous unbinding is neglected (since $k_{\text{bind}} \gg k_{\leftarrow}$). From Eq. (5.19) it is straightforward that:

$$\phi(t) = \phi_0 \exp\left(-k_{\text{bind}} t\right), \tag{5.20}$$

where ϕ_0 is an unknown constant defined by the initial conditions. In order to get $\phi(f)$ from $\phi(t)$, the time-dependence of the force in the experimental protocol is required. In the case of an unfolding process this is $f(t) = f_0 + rt$, where f_0 is the minimum force. Consequently:

$$\phi(f) = \phi_0^* \exp\left(-\frac{k_{\text{bind}}}{r} f\right), \tag{5.21}$$

and $\phi_0^* = \phi_0 \exp\left(\frac{k_{\text{bind}}}{r} f_0\right)$ is a constant.

At large forces $k_{NB\to U}(f)$ is dominated by $k_{B\to U}(f)$ since $\phi(f) \overset{f\to\infty}{\longrightarrow} 0$ (Eqs. 5.18 and 5.21). Therefore, when large rupture forces are explored, the unfolding pathway for such events is always given by the reaction $B \to U$. Along the same lines of what has been done so far, the existence of a kinetic barrier $B_{BU}(f)$ that mediates such transition is assumed and, according to Eq. (5.15a), it is mathematically related to $k_{B\to U}(f)$ as:

$$k_{B\to U}(f) = k_0^{BU} \exp\left(-\frac{B_{BU}(f)}{k_{\text{B}}T}\right), \tag{5.22}$$

where k_0^{BU} is the attempt rate at zero force for the pathway $B \to U$.

It is reasonable to assume that the binding position of the peptide along the hairpin stem corresponds to the position of the TS found along the unfolding pathway $B \to U$. In order to unfold the hairpin it is necessary to overcome the kinetic barrier produced from binding the ligand at the binding position. Two facts support that Thiocoraline mostly binds at the beginning of the stem. First, in pulling experiments there are no transitions at low forces to any intermediate state of the hairpin that precedes full unfolding (as it was the case for molecules I1 and I2 in Section 5.3). Second, the relative distance between state B and the TS can be calculated using Eq. (4.19),

$$x_{B-\text{TS}} = -\frac{\partial B_{BU}(f)}{\partial f}. \tag{5.23}$$

Now, $x_{B-\text{TS}}$ can be obtained from the experimental data shown in Fig. 5.25 by using the approximation $\log k_{NB\to U}(f) \sim \log k_{B\to U}(f)$ in Eq. (5.18). Therefore:

$$x_{B-\text{TS}} = k_{\text{B}}T \frac{d \log k_{B\to U}(f)}{df} \tag{5.24a}$$

$$\simeq k_{\text{B}}T \frac{d \log k_{NB\to U}(f)}{df}\Big|_{f_{\text{large}}} = 1.8 \pm 0.4 \ \text{ nm}, \tag{5.24b}$$

which implies that the TS (or, equivalently, the peptide binding site) is two base pairs ahead of the stem of the hairpin (at the hairpin configuration $n = 2$, corresponding to

the bases 5'-CG-3'). Once this is known, a first order perturbative approach can be used to write a general expression for the free energy of formation at zero force of the configuration n of the hairpin with a bounded ligand at position $n = 2$:

$$\Delta G_n^0 = \sum_{i=n+1}^{\mathrm{N}-1} (g_{i,i+1} - g_{\mathrm{bind}}\delta_{i2}) + g_{\mathrm{loop}}(1 - \delta_{n\mathrm{N}}), \tag{5.25}$$

where δ_{i2} and δ_{nN} are both Kronecker deltas, equal to 1 if $i = 2$ and $n = \mathrm{N}$ respectively, and zero otherwise. Equation (5.25) is formally identical to Eq. (4.2), but a binding free energy term g_{bind} has been added to the base-pair free energy term $g_{2,3}$, which corresponds to an additional energetic term in the binding position of Thiocoraline along the hairpin stem. Therefore, it is considered that the binding event does not have any effect on the base-pair free energies of the DNA hairpin other than at $n = 2$. In Appendix H it is demonstrated that g_{bind} is equal to $\mu - \varepsilon$, being μ the chemical potential of Thiocoraline, and ε the gain in energy due to binding. Hence, g_{bind} depends on [Thio] according to the law of mass action,

$$\log\left(\frac{K_{\mathrm{eq}}}{[\mathrm{Thio}]}\right) = -\frac{g_{\mathrm{bind}}}{k_{\mathrm{B}}T}, \tag{5.26}$$

where $K_{\mathrm{eq}} = k_{\leftarrow}/k_{\rightarrow}$ is the equilibrium constant of the binding reaction.

In Fig. 5.26a, the FEL of the molecular system made of CD4 plus a binding peptide at $n = 2$ is represented at the coexistence force f_c for different values of the binding free energy g_{bind}, ranging from 0 to 10 kcal/mol. The effect of force into the molecular FEL is included as described in Section 4.1.1. It can be seen that a new TS at $n = 2$ competes with the original TS of CD4 at $n = 7$. In addition, in Fig. 5.26b it is shown how f_c increases with g_{bind}, which reasserts the stabilizing effect of the bound peptide on the hairpin double helix. Finally, in Fig. 5.26c theoretical profiles of $B_{BU}(f)$ computed for different values of g_{bind} according to the Kramers solution (Eq. 4.32) are shown.

On the other hand, an effective kinetic barrier $B_{(NB)U}(f)$ between the joint folded state NB and state U is defined so that it satisfies:

$$k_{(NB)\to U}(f) = k_0^{(NB)U} \exp\left(-\frac{B_{(NB)U}(f)}{k_{\mathrm{B}}T}\right), \tag{5.27}$$

where $k_0^{(NB)U}$ is a constant. Because at large forces $k_{(NB)\to U}(f)$ tends to $k_{B\to U}(f)$, as shown by Eq. (5.18), it can be accordingly assumed that, at large forces, $B_{(NB)U}(f) = B_{BU}(f)$ and as a consequence $k_0^{(NB)U} = k_0^{BU}$. Hence, the experimental profile of $B_{(NB)U}(f)$ determined using unfolding forces and Eq. (5.27) can be matched to the theoretical estimation of $B_{BU}(f)$ at large forces (above 20 pN in the case of study). The matching procedure is performed by looking at the value of $\log k_0^{(NB)U}$ (or, equivalently, of $\log k_0^{BU}$) for which χ^2 is minimized (Appendix E). In this case, P is set to 1.35 nm and the method is performed for 10, 50, 100 and 200 nM [Thio] at different values of g_{bind}. An example obtained at 200 nM [Thio] is shown in Fig. 5.27a: the red-solid line is

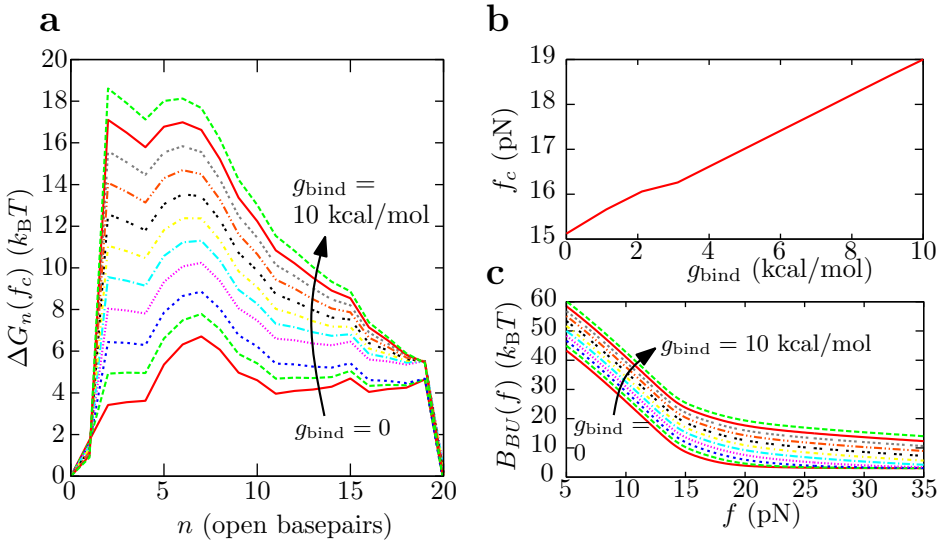

Figure 5.26: Effect of the bound intercalator on the molecular FEL of CD4. a. FEL of CD4 calculated by assuming one bound Thiocoraline at $n = 2$ with a binding energies $g_{\rm bind} = 0$, 1, 2, 3, 4, 5, 6 ,7, 8, 9 and 10 kcal/mol (indicated with different line styles in the direction of the arrow). **b.** Coexistence force f_c of CD4 with a bound peptide at $n = 2$ as a function of the binding free energy, $g_{\rm bind}$. **c.** Theoretical profiles of the force-dependent kinetic barrier $B_{BU}(f)$ as a function of the binding free energy $g_{\rm bind}$, for values ranging from 0 to 10 kcal/mol. Color code as in panel a.

the theoretical estimation of the kinetic barrier $B_{BU}(f)$ obtained for $g_{\rm bind} = 5$ kcal/mol, the magenta-solid squares are the experimental estimation of $B_{(NB)U}(f)$ obtained from the measured unfolding forces. At large forces, $B_{(NB)U}(f)$ is matched with $B_{BU}(f)$ in order to determine $\log k_0^{BU}$ as justified by Eq. (5.18). Theoretical and experimental estimations for $B_{NU}(f)$ are also shown. In Fig. 5.27b the estimated values for $\log k_0^{BU}$ for each value of $g_{\rm bind}$ are represented.

In order to determine the kinetic properties of the binding of Thiocoraline into ds-DNA, Eqs. (5.18) and (5.27) are merged together and the following mathematical expression for $B_{(NB)U}(f)$ is obtained:

$$B_{(NB)U}(f) = -k_{\rm B}T \log\left[\phi(f)\frac{k_0^{NU}}{k_0^{BU}}\exp\left(-\frac{B_{NU}(f)}{k_{\rm B}T}\right) + (1-\phi(f))\exp\left(-\frac{B_{BU}(f)}{k_{\rm B}T}\right)\right], \tag{5.28}$$

where $\phi(f)$ is given by Eq. (5.21) and $k_0^{NU} = (8 \pm 3) \times 10^4$ s^{-1} for CD4 was already determined in Section 4.3.2. This expression, that depends on the ligand concentration, can be simultaneously fit to the experimental $B_{(NB)U}(f)$ measured at 10, 50, 100 and 200 nM [Thio] in order to get estimates for $k_{\rightarrow}$ and the three values of ϕ_0^* (on for each value of [Thio]). The procedure starts by taking a fixed value for $g_{\rm bind}$ at a given concentration of [Thio], which determines $g_{\rm bind}$ and $B_{BU}(f)$ for the other concentrations through the law of mass action (Eq. 5.26). Next, the fit of the experimental data to Eq. (5.28) using the corresponding profile of $B_{BU}(f)$ for each concentration is performed. Interestingly, results for $k_{\rightarrow}$ obtained from the fit do not depend on the value of

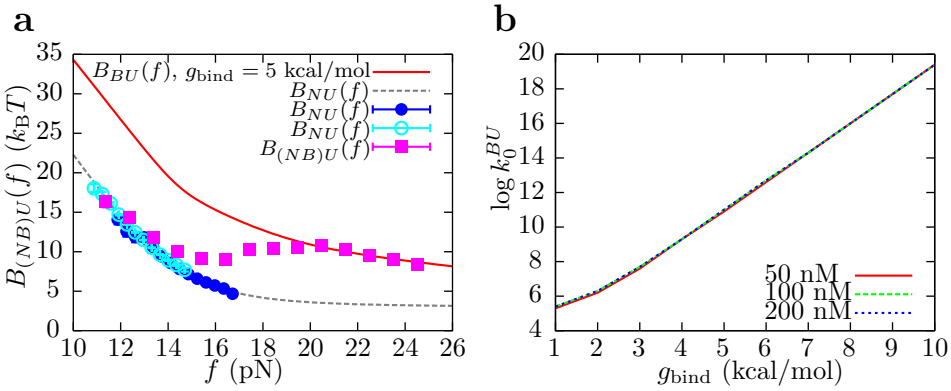

Figure 5.27: The matching procedure for $B_{BU}(f)$. a. Match between the theoretical and experimental profiles for $B_{NU}(f)$ (dashed-gray line, and green-solid and cyan-open circles, respectively), theoretical profile of $B_{BU}(f)$ obtained at $g_{\rm bind} = 5$ kcal/mol (red-solid line), and experimental measurement of $B_{(NB)U}(f)$ (magenta-solid squares). **b.** Estimated value of $\log k_0^{BU}$ obtained from χ^2-minimization by matching the theoretical profile of $B_{BU}(f)$ obtained for different values of $g_{\rm bind}$ with the experimentally measured $B_{(NB)U}(f)$ at large forces.

$g_{\rm bind}$, neither on the time interval Δt introduced between a folding and the subsequent unfolding in the pulling experiments. The average value of $k_{\rightarrow} = 0.070 \pm 0.005$ s^{-1}nM^{-1} is recovered. In Fig. 5.28 results of the fit for the different concentrations of Thiocoraline obtained using $g_{\rm bind} = 5$ kcal/mol are shown for data obtained at $\Delta t = 5$ s. Results are identical for fits obtained using data sets with different preset values of Δt and using different values for $g_{\rm bind}$.

The fact that $k_{\rightarrow}$ does not depend on $g_{\rm bind}$ for the range of concentrations explored obstructs the determination of $k_{\leftarrow}$ and $K_{\rm eq}$ for the binding reaction. In case the unbinding kinetic rate is not neglected, it can be demonstrated that $\phi(t)$ satisfies:

$$\phi(t) = \frac{1}{1 + \frac{[\text{Thio}]}{K_{\rm eq}}} + \phi_0 \exp\left[-k_{\leftarrow}\left(\frac{[\text{Thio}]}{K_{\rm eq}} + 1\right) t\right], \tag{5.29}$$

where $K_{\rm eq} = k_{\leftarrow}/k_{\rightarrow} = [\text{Thio}] \exp(g_{\rm bind}/k_{\rm B}T)$, according to the law of mass action (Eq. 5.26 and Appendix H). If Eq. (5.29) is introduced in Eq. (5.28) and the fit of the experimental estimation of $B_{(NB)U}(f)$ is performed, again $k_{\rightarrow} \sim 0.07$ s^{-1}nM^{-1} for any value of $g_{\rm bind}$ (and consequently, for any value of $K_{\rm eq}$). This suggests that the recovered value for $k_{\rightarrow}$ using this DFS method is a diffusion-limited upper bound.

5.5 Conclusions

DFS experiments are useful to unravel the kinetic properties of single molecules. In literature, the majority of studies extract thermodynamic information of multi-state molecular systems only from equilibrium experiments. In addition, simplified models such as the BE approach are used to interpret experimental data [Rief 99, Schl 04, Schl 06, Wood 06a, Wood 06b, Wood 08, Forn 11].

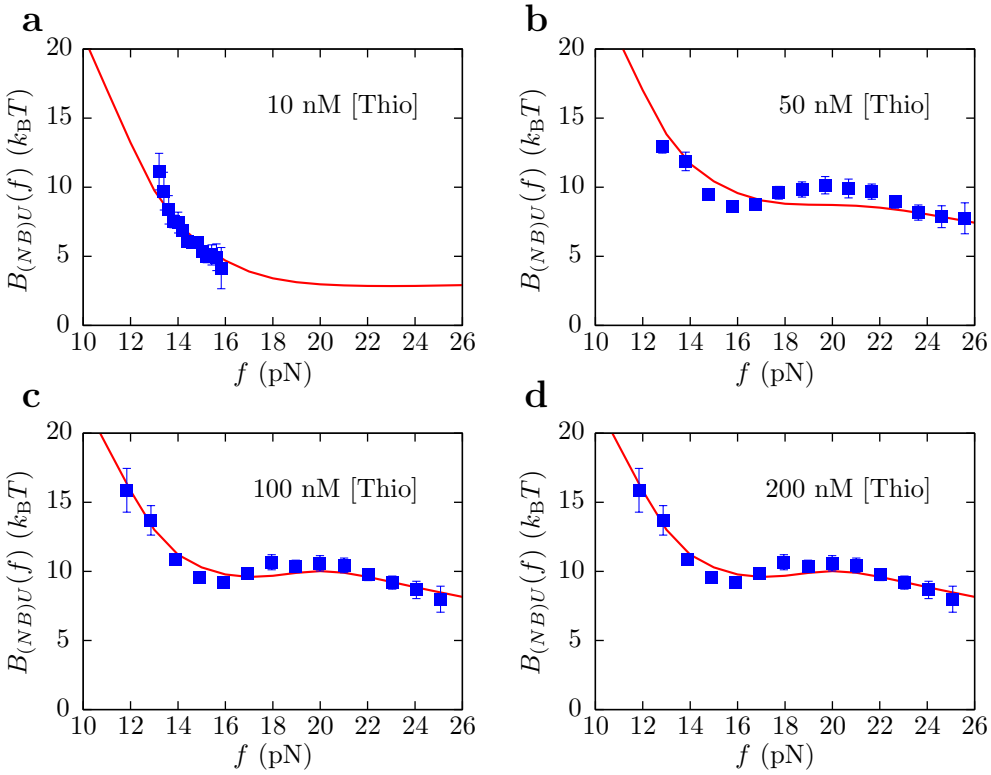

Figure 5.28: Kinetic barrier $B_{(NB)U}(f)$. Fit of the theoretical expression for $B_{(NB)U}(f)$ (Eq. 5.28, red-solid line) to the experimental data obtained by setting $\Delta t = 5$ s (blue-solid squares) for 10 (**a**), 50 (**b**), 100 (**c**) and 200 nM [Thio] (**d**). Numerical results for the fit give $k_{\rightarrow} = 0.070 \pm 0.005$ s^{-1}nM^{-1}, $\phi_0^* = 1.2,\ 27.5,\ 14483$ and 5.5×10^8 for 10, 50, 100 and 200 nM [Thio], respectively.

In this chapter, DFS has been extended beyond current frontiers, and a variety of complex molecular situations have been addressed using small nucleic acid molecules. To this end, the Kramers-based model introduced in Chapter 4 to study the kinetic barrier is used to analyze experimental results from non-equilibrium mechanical denaturation of nucleic acid hairpins. The standard version of this method succeeds in extracting the unfolding and folding kinetic rates, the height and position of the kinetic barrier, and the free energy of formation of molecules [Mano 06]. Results are in agreement with those obtained with a standard equilibrium approach, which provides a simple validation of the non-equilibrium method. In the presence of high kinetic barriers along the molecular FEL, it turns out that non-equilibrium experiments allow to sample rare states or transitions at forces where they would never occur in equilibrium conditions. Therefore, phenomena observed in wider range of forces can be investigated with non-equilibrium approaches using the aforementioned Kramers-based analysis.

In the present state of the art of DFS, the BE model occupies a privileged position in the analysis of two-state molecules. It is a very simple model that allows to easily fit force-dependent unfolding and folding kinetic rates to successfully get the general order of magnitude of the position and the height of the kinetic barrier in two-state molecular

systems. However, it is shown in this chapter that when the kinetic barriers involved in the mechanical unfolding and folding of nucleic acid hairpins are too high, hysteresis effects are large and the use of the BE model may give underestimated results for the TS position and height. The direct observation of detailed balance is the most important experimental condition that must be fulfilled in order for the results of the BE fit to be reliable. That is, unfolding and folding kinetic rates must be measured at identical values of force, and close to the condition of coexistence of the unfolded and the folded states. Otherwise, there is no guarantee that the kinetic barrier governing each transition is the same. An explanation of such behavior is given by the Leffler-Hammond postulate [Leff 53, Hamm 55]: as the force applied to the ends of the molecule increases, the kinetic barrier mediating unfolding approaches the native/folded state. Hence, at large forces the kinetic barrier is close to the native state, whereas as force increases it moves towards the unfolded state. As a result, transitions –either folding and unfolding– occurring at extreme values of force can be mediated by different kinetic barriers. Within this scenario, it is important to carefully select a small range of forces where both unfolding and folding kinetic rates are sampled in order to properly interpret experimental data from DFS experiments using the BE model.

The Kramers-based method used in this chapter to analyze results from DFS experiments is more robust, and indeed it allows to extract the force-dependent position of the kinetic barrier. Hence, the validation of the Leffler-Hammond rule can be experimentally tested. To this end, three different DNA hairpins, with identical structural properties and different sequences, have been used and the effect of applied force on the position of the TS (or, equivalently, on the molecular fragility) has been investigated. It is observed that the closer the TS to the unfolded state, the more compliant the molecule behaves under applied force.

Second, the non-specific binding of Na^+ and Mg^{2+} ions into a short RNA hairpin is characterized. The effect on the position and height of the kinetic barrier and the free energy of formation of the molecule can be unraveled, and the ability of Mg^{2+} ions to stabilize RNA at much lower concentrations than monovalent ions is experimentally tested. Furthermore, it is shown how the Kramers-based method can be extended in order to access the elastic properties of ssDNA at different ionic strengths. Despite the biological relevance of ions on RNA structures, it is the first time that their effect on the persistence length is experimentally investigated in mixed monovalent/divalent salt conditions. Both elastic and thermodynamic results are appealing because they allow to test several theories where the effect of ions and counterions on nucleic acids in solution are theoretically modeled, such as the Poisson-Boltzmann theory, the CC model or the TBI model –the former one being in better agreement with our experimental data.

Third, it is shown how to characterize kinetic states, which are metastable states observed under non-equilibrium conditions. Based on simple Markov models, it is possible to apply the Kramers theory to different reaction pathways and characterize the kinetic and thermodynamic properties of intermediate states, that are observed on pathway along the folding and the unfolding reaction. The method is successfully applied to two DNA hairpins with one or two intermediate states, respectively (hairpins I1 and

I2). Interestingly, for hairpin I1 kinetic rates can be measured under both equilibrium and non-equilibrium conditions, and are shown to be compatible. Hairpin I2 presents a situation where equilibrium experiments are not suitable to characterize the kinetic molecular properties and where non-equilibrium experiments are more appropriate.

Finally, the binding of a DNA intercalating peptide into dsDNA is also characterized. Again, first order Markov models and the Kramers theory are combined in order to obtain the binding position of the intercalating ligand, and to obtain an upper bound for the value of the binding kinetic rate. These results are valuable, since at the experimental conditions explored it is not possible to observe unbinding events under equilibrium experiments. Hence, the present analysis shows the power of the non-equilibrium approach to characterize molecular interactions.

In sum, in this chapter a Kramers-based approach is used to characterize a wide range of molecular situations involving small nucleic acid molecules, from the unfolding and folding of two-states DNA hairpins under different experimental conditions to molecules with kinetic properties of intermediate and ligand-bound states. These are powerful results since they prove the capacity of DFS experiments to understand several molecular processes.

Future studies could address situations where molecular transitions are not detected through a sudden jump in force when the molecule partially folds or unfolds. For instance, misfolded states usually show a smooth unfolding and folding FDC along pulling experiments because of their rough FEL characterized by low kinetic barriers. Recent studies discuss how to obtain the unfolding and folding kinetic rates in such situations from non-equilibrium experiments [Sola 14]. Hence, the Kramers-based approach used here could be used to characterize the molecular FEL and the free energy of formation of the misfolded structure. Other possible studies could address the measurement of entropic and enthalpic barriers by repeating pulling experiments of different nucleic acid molecules at different temperatures. Finally, different populations could be characterized in a molecular sample made of nucleic acid mutants. A first application could be to check the variability of unfolding and folding kinetic rates over different members of the whole ensemble. Moreover, the ability of an intercalating peptide to bind the different molecules could be used to carry out molecular evolution studies.

Chapter 6

Affinity and elasticity of single bonds

Intermolecular recognition is an ubiquitous phenomenon in biological systems, like cellular adhesion, cell signaling, or the regulation of replication and transcription of nucleic acids. Noteworthy, molecular recognition is crucial in the humoral immune system of vertebrate organisms, where antibodies can specifically identify a single foreign body (antigen) among thousands of molecules (Section 2.1.3).

A proper understanding of the mechanisms that govern the immune response is critical to the optimal performance of antibody-based therapies and techniques. Therefore, the investigation of molecular interactions and the characterization of antigen-binding affinity constitutes a major area of interest. Traditional studies [Fiel 96, Homo 03, Rose 09, Liu 10] use nuclear magnetic resonance spectroscopy, crystallographic analysis, surface plasmon resonance or fluorescence microscopy to characterize biophysical properties of bonds such as affinity, dissociation rate constants or epitope binding sites.

Several studies have indirectly shown that specificity and rigidity are related properties of intermolecular bonds [Berg 99, Mani 00, Ma 02, Sund 02, Natk 13]. Recently, Romesberg and collaborators [Jime 03, Jime 04, Thor 07] used photon echo spectroscopy to relate the time scales of intermolecular bond motion to the degree of flexibility of the interaction, which provided a novel method to characterize bond elasticity as a function of bond affinity. Experimental evidence suggests that bond elasticity is a key feature controlling the response of the immune system to infection: at the initial stages, the detection of intrusive agents is mainly non-specific; antibodies are very elastic and continuously explore different conformations, which enables them to bind as many antigens as possible. Resulting intermolecular bonds are expected to be very flexible and varied, and despite being suboptimal interactions they may reorganize their structure to increase binding strength [Boeh 09]. After the detection, a complex immunitary response entailing the massive production of different antibodies takes place [Grey 64, Jane 02]. This process, known as the maturation line of the immune system, triggers the appearance of new antibodies which specifically bind to destroy a given foreign body through rigid bonds. Interestingly, such new antibodies do not lose their ability to explore multiple

conformations and cross-react with different antigens, which promotes the recognition of newer infections through the establishment of elastic non-specific bonds. Therefore, direct measurements of the elastic properties of ligand-receptor bonds can be crucial to understand the relation between bond affinity and bond rigidity.

In the last decade, advances in the mechanical manipulation of receptor-ligand bonds at the level of individual molecular pairs have been possible thanks to the improvement of experimental techniques such as atomic force microscopy, biomembrane force probe transducer and laser optical tweezers [Evan 01, Huge 01, Rito 06b]. It was shown in Chapter 3 that OT are ideally suited to characterize the elastic response of polymers to applied mechanical forces [Smit 96, Tskh 97a, Wang 97]. Moreover, they are an emergent tool to investigate the effect of force on single antibody-antigen interactions because of their accessible force-range (0-80 pN) and resolution (∼0.1 pN) [Kuo 93, Helm 97, Sims 99, Stri 99, Baum 00, Stou 01, Perr 02, Salo 08, Evan 10, Wagn 11, Stan 13]. In this chapter it is shown how OT can be used to characterize the correlation between bond flexibility and mechanical strength in intermolecular recognition.

6.1 Dynamic force spectroscopy experiments of single bonds

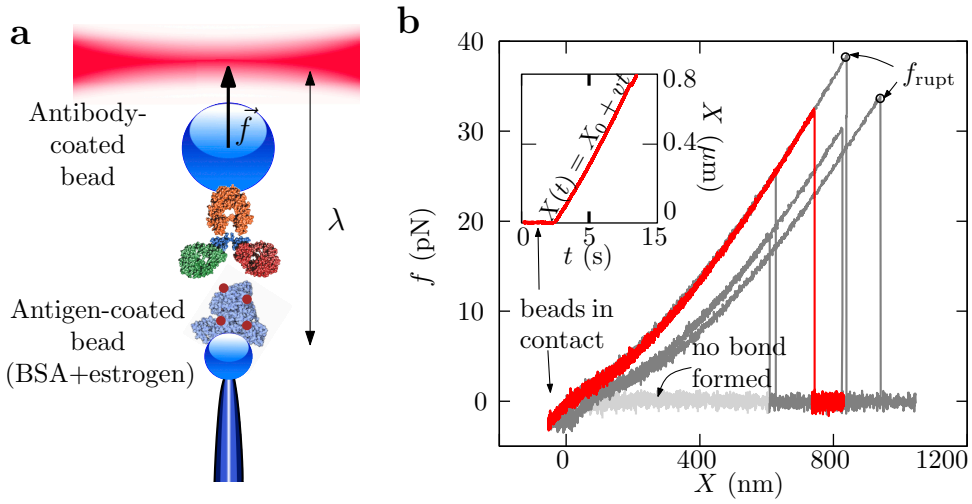

Figure 6.1: DFS experiments of single bonds. **a.** Experimental scheme (figure not to scale). One bead coated with antibodies is captured in the optical trap. Another bead coated with antigens is immobilized in the tip of a micropipette by air suction. The beads are approached by varying the trap-pipette distance λ. **b.** Pulling experiments. Inset: beads are kept in mutual contact at -3 pN for 2 s. Next λ is increased at a constant speed v, around 140 nm/s if not stated otherwise. Main panel: different examples of force-distance curves. After contact, the force remains zero in the absence of a binding event (light-gray curve). When a bond is successfully established the force increases until the value $f_{\rm rupt}$ is reached and then it drops to zero (red and dark-gray curves). A minimum of 5 and a maximum of 70 pairs of beads were tested for each interaction under study, and at least 80 pulling curves per pair of beads were recorded.

In order to study the antibody-antigen interaction with the minitweezers, a 3 μm-

diameter polystyrene bead is coated with antibodies an captured in the optical trap. A second 2 μm-diameter polystyrene bead is coated with the antigen and is kept fixed in the tip of a micropipette by air-suction (Fig. 6.1a). The coating protocol is summarized in Appendix A.3.

At the beginning of a pulling experiment (Fig. 6.1b) the two beads are kept at a contact force of -3 pN for 2 seconds, enabling the molecules on the surface to interact. Next, the distance between the center of the optical trap and the tip of the micropipette, λ, is increased at a constant pulling speed v (Fig. 6.1b, inset). After contact either no interaction is established and the force remains zero (light-gray curve in Fig. 6.1b), or a bond is formed (red and dark-gray curves) and the force increases until the bond breaks at the so-called rupture force, f_{rupt}, and the force drops to zero.

6.1.1 Bond strength and stiffness

The bond strength is related to the resistance of the bond to be broken by the application of a mechanical force. Intuitively, this must be proportional to average values of rupture forces. Therefore, bond strength is quantified from the measured rupture forces.

The slope of experimental FDC, denoted by k_{eff}, is used to quantify the flexibility of binding. It is observed in Fig. 6.1b that this slope increases with force. As a criterion, k_{eff} is measured at the preset force value of 5 pN for all the measured antibody-antigen interactions.

To find an expression for the effective stiffness of the antigen-antibody bond, the same steps followed in Chapter 3 in case of DNA hairpins (Eq. 3.2b) are followed. Here, k_{eff} only contains contributions from the rigidity of the antibody-antigen bond, k_{bond}, and the rigidity of the optical trap, $k_{\text{OT}} \simeq 0.078 \pm 0.005$ pN/nm [Forn 11]. The mathematical relation between the three different rigidities is given by:

$$\frac{1}{k_{\text{eff}}} = \frac{1}{k_{\text{bond}}} + \frac{1}{k_{\text{OT}}}, \tag{6.1}$$

which leads to:

$$k_{\text{bond}} = \frac{k_{\text{OT}} \cdot k_{\text{eff}}}{k_{\text{OT}} - k_{\text{eff}}}. \tag{6.2}$$

k_{bond} is always a positive quantity, and is the result of the serially connected stiffnesses of the antigen, of the antibody and of the bond. Low values for k_{bond} are indicative of flexible bonds, whereas high values correspond to stiff/rigid bonds.

6.1.2 Removal of multiple interactions

Some traces (less than 10%) show two or more rupture forces, indicating multiple binding, *i. e.* the formation of parallel bonds. To focus the attention on single bonds such events are not taken into account. However, sometimes multiple bonds can break at the same time and cannot be discriminated within the limited time resolution measurement.

A statistical method, previously introduced by Evans and collaborators [Evan 10], is used to remove these simultaneous multiple unbinding events. The method is based on

Bond	N	Successful binding events	Multiple bonds (Eq. 6.4)
Pab-MB-BSA	2674	1080	255
PreI-MB-BSA	1597	212	15
Pab-BSA	386	12	0
PreI-BSA	312	17	0

Table 6.1: Polyclonal recognition and multiple binding events. Number of experiments N, number of successful binding events, and estimated number of simultaneous multiple binding events according to the Poisson distribution for the different polyclonal interactions investigated (Section 6.2).

Bond	N	Successful binding events	Multiple bonds (Eq. 6.4)
Mab-B-BSA	2767	944	183
Mab-T-BSA	3347	756	93
Mab-BSA	2215	414	41

Table 6.2: Monoclonal recognition and multiple binding events. Number of experiments N, number of successful binding events, and estimated number of simultaneous multiple binding events according to the Poisson distribution for the different monoclonal interactions investigated (Section 6.3).

a Poissonian analysis of rupture forces: the binding probability p is defined as the ratio between the number of successful binding events and the total number N of experiments performed. Next, the probability to find a given number x of parallel tethers is assumed to follow a Poisson distribution,

$$P(x) = \frac{e^{-\mu}\mu^x}{x!}. \tag{6.3}$$

Since $p = P(x > 0)$, it can be shown that $\mu = -\log(1-p)$ and that the probability to find multiple bonds is equal to:

$$Q = P(x > 1) = p - (p-1)\log(1-p). \tag{6.4}$$

Therefore, $N \times Q$ gives an estimation of the number of simultaneous multiple binding events. Tables 6.1 and 6.2 show the number N of experiments, the number of successful binding events and the number of multiple binding events $N \times Q$ for the different interactions investigated in this chapter.

It is reasonable to assume that parallel tethers are strong and dissociate at large forces. On the other hand, the total stiffness measured in a pulling experiment is the sum of the different stiffnesses of each individual bond. Therefore, both large rupture forces and high stiffness values suggest the presence of multiple tethers. A distance d is defined as:

$$d = \sqrt{\left(\frac{f_{\text{rupt}}}{f_{\text{rupt}}^{\max}}\right)^2 + \left(\frac{k_{\text{bond}}}{k_{\text{bond}}^{\max}}\right)^2}, \tag{6.5}$$

where $f_{\text{rupt}}^{\max}$ and $k_{\text{bond}}^{\max}$ are the maximum values of rupture forces and bond rigidities experimentally measured. In order to remove multiple binding, the $N \times Q$ events with largest values of d are removed from the whole set of rupture events.

6.2 The maturation line of polyclonal recognition

A rabbit was inoculated with a complex made by methyl-boldenone (MB) linked to the carrier protein horseshoe crab hemocyanin (HCH). Such molecular complex is used to guarantee the immune response of the animal since MB by itself is a very small compound unable to turn on the immune system of the animal. Polyclonal antibodies present in the animal blood before immunization (hereafter referred as pre-immunological antibodies, PreI) and after immunization (Pab) were isolated [Kram 07].

Different sets of polystyrene beads were coated with PreI, Pab and the antigen, respectively. In order to focus the attention on antibodies which specifically recognize MB, an antigenic complex made by linking MB to the carrier protein bovine serum albumin (BSA) was used.

Then, force-spectroscopy experiments were carried out using different sets of beads: PreI-coated beads and Pab-coated beads were tested, separately, against MB-BSA-coated beads. In order to evaluate the degree of affinity of each set of antibodies to MB, control experiments between Pab-/PreI-coated beads and BSA-coated beads (that is, without MB) were also carried out.

Large variability in the slopes of different FDC is observed (Fig. 6.1b), which suggests a diversity of antibody-antigen bonds. Two sources may contribute to such variability:

1. The heterogeneous population of antibodies present in the PreI and Pab samples.

2. The diversity of conformational sub-states for each antibody that are functional and bind to the antigen [Berg 99, Kosh 58, Ma 02].

Figure 6.2a shows the binding efficiencies of pulling experiments, defined as the ratio between the number of successful binding events to the total number of times the pair of beads were put in mutual contact, for the different samples. In order to discriminate unbinding events from thermally-induced force fluctuations, rupture force events below 3 pN were discarded. The binding efficiency between Pab and the complex MB-BSA is almost 35%, whereas between Pab and BSA is lower than 15%. This proves that the immune system develops antibodies that accurately detect MB, even if the immunization has been performed by linking MB to a big carrier protein such as HCH. On the other hand, the binding efficiencies measured between PreI and MB-BSA or BSA (10-15%) are comparable. This indicates that both PreI and Pab detect BSA with an almost identical low affinity, which suggests a similar non-specific recognition.

The histograms of f_{rupt} for successful binding events in the case of PreI-MB-BSA and Pab-MB-BSA are very similar (Fig. 6.2b top). In contrast, the histogram of bond rigidities measured at 5 pN, k_{bond}, for the Pab-MB-BSA case extends to higher values as compared to the one obtained for the PreI-MB-BSA case (Fig. 6.2b bottom).

In order to gain more information about differences between the pre-immunological and specific polyclonal recognition to MB the correlation between bond elasticity and rupture forces is studied. 2D contour plots are used with histograms of f_{rupt} plotted against histograms of k_{bond} [vZiv 10], which reveal direct correlation between specific

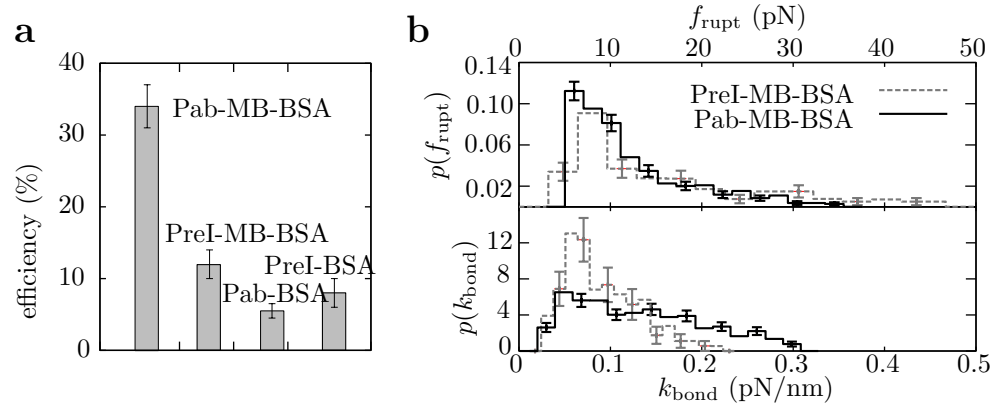

Figure 6.2: Force-spectroscopy experiments with polyclonal antibodies. a. Binding efficiencies in pulling experiments with polyclonal antibodies PreI and Pab tested against the complex MB-BSA (error bars are standard errors over different tested pairs of beads). **b.** Histograms of $f_{\rm rupt}$ (top) and $k_{\rm bond}$ evaluated at 5 pN (bottom) measured with polyclonal antibodies PreI (dashed line) and Pab (solid line) tested against the complex MB-BSA. Error bars are the standard deviation, and were obtained using the Bootstrap method.

recognition and bond elasticity. Figures 6.3a and 6.3b show such plots for the recognition events measured in pulling experiments testing PreI and Pab against MB-BSA respectively. In the case of PreI, most binding events occur at $f_{\rm rupt}$ below 10 pN and $k_{\rm bond}$ below 0.1 pN/nm (Fig. 6.3a, purple-orange-yellow region). Moreover, the purple scattered regions in Fig. 6.3a show that some PreI-MB-BSA bonds, despite being non-specific, are able to structurally reorganize in order to increase binding strength, which implies an increase in $f_{\rm rupt}$. In the case of Pab, successful binding events cover a wider interval of rupture forces and new regions in the 2D contour plot are populated at high values of $k_{\rm bond}$ (from 0.1 to 0.3 pN/nm, Fig. 6.3b). This indicates that interactions measured between PreI and MB-BSA are weaker and more flexible than the ones measured using Pab and the same antigen. Therefore, the immune system has generated new antibodies that recognize MB through more rigid bonds.

6.3 The cross-reactivity of monoclonal antibodies

A similar procedure can be applied to study the cross-reactivity of antibodies. In this case, a monoclonal antibody (Mab) was produced against the anabolic steroid boldenone (B) [Kram 07]. It is expected that this antibody also has affinity for testosterone (T), which is another steroid hormone that shares the rings B, C and D in common with B (Fig. 6.4a). Pulling experiments were carried out to prove the ability of Mab to cross-react against the two similar haptens and to quantify its different binding strengths to different antigens. As in the case of MB, both B and T were linked to BSA.

Figure 6.4b shows the experimental binding efficiencies obtained between Mab and B-BSA, T-BSA or BSA. As expected, the highest value is found for Mab tested against the complex B-BSA, followed by the one measured between Mab tested against T-BSA.

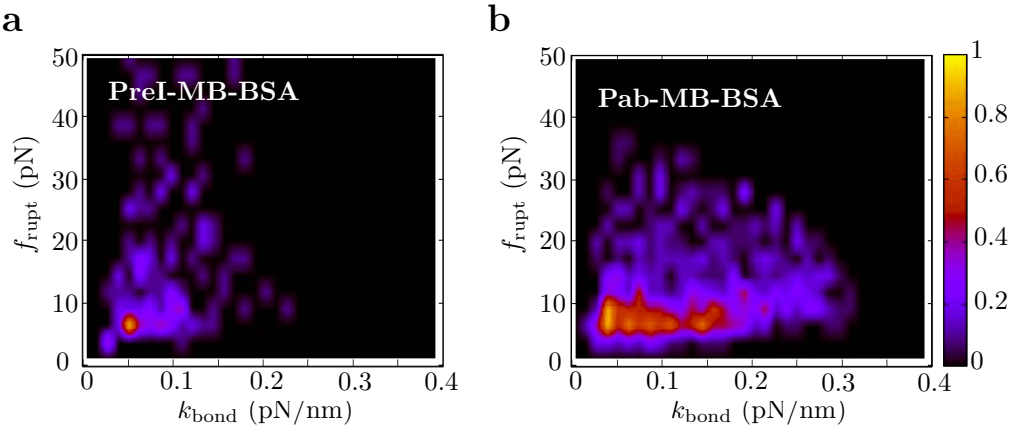

Figure 6.3: Maturation line for MB. 2D contour plot with histograms of f_{rupt} plotted against histograms of k_{bond} measured in pulling experiments using the complex MB-BSA and PreI **(a)** or Pab **(b)**.

Finally, the efficiency reported for the recognition between Mab and BSA is the lowest.

The histograms of f_{rupt} (Fig. 6.4c, top) for the three bonds Mab-B-BSA (solid line), Mab-T-BSA (dashed line) and Mab-BSA (dotted line) show different features from each other. In the Mab-B-BSA case, a maximum in rupture forces around 30 pN is revealed, whereas in the Mab-T-BSA case a shoulder can be seen around 20 pN with positive deviation with respect to the Mab-BSA case. Both behaviors suggest recognition from Mab to B and T with different degrees of affinity. The histograms of k_{bond} (Fig. 6.4c, bottom) are also evaluated and a maximum around 0.1 pN/nm is obtained for the three cases. However, $p(k_{\mathrm{bond}})$ for Mab-B-BSA shows a lower deviation with respect to the other cases at larger values of k_{bond}.

Again, the 2D contour plot of histograms of f_{rupt} and k_{bond} reveal features of the different antibody-antigen bonds that get masked with the conventional histograms $p(k_{\mathrm{bond}})$ and $p(f_{\mathrm{rupt}})$. In Fig. 6.5a the case for the recognition between Mab and BSA is shown. A broad range of bond elasticities at low values of rupture forces is covered (purple-orange-yellow region). This suggests that the bond Mab-BSA is able to explore several structural conformations, with different degrees of flexibility, in order to increase its binding affinity. Such result combined with the observed low efficiency in pulling experiments (<20%, Fig. 6.4b) indicates that the interaction between Mab and BSA is mainly non-specific, in agreement with results obtained with polyclonal antibodies (Fig. 6.2). An extremely different pattern is obtained for the recognition between Mab and B-BSA (Fig. 6.5b): at forces below 10 pN the 2D contour plot is very similar to the one measured for the Mab-BSA bond (Fig. 6.5a), which indicates that in this region interactions between Mab and BSA mainly occur. Remarkably, the Mab-B-BSA bond also covers values of f_{rupt} ranging from 20 to 40 pN (orange-yellow area). Such region is not populated by the Mab-BSA bond. This is a signature of specific recognition between Mab and B.

In the 2D contour plot of histograms of f_{rupt} and k_{bond} for Mab-T-BSA (Fig. 6.5c) it is observed that most rupture forces are found below 30 pN and a wide interval of bond

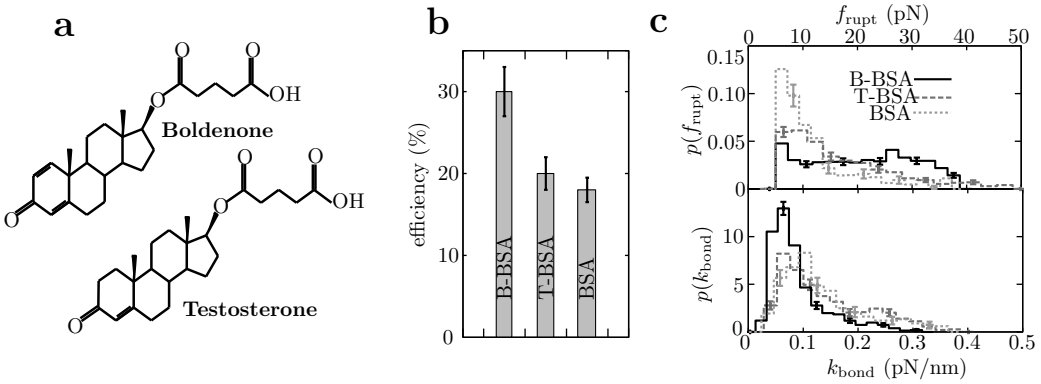

Figure 6.4: Cross-reactivity of a monoclonal antibody. a. Chemical structures for the haptens of boldenone (B) and testosterone (T) used in force-spectroscopy experiments. **b.** Binding efficiencies in pulling experiments using Mab against different antigens (error bars are standard errors over different tested pairs of beads). **c.** Histograms of $f_{\rm rupt}$ (top) and $k_{\rm bond}$ evaluated at 5 pN (bottom) measured with Mab tested against B-BSA (solid line), T-BSA (dashed line) and BSA (dotted line). Error bars are the standard deviation, and were obtained using the Bootstrap method.

rigidities is covered ($k_{\rm bond}$ ~0.05-0.3 pN/nm). Such results reveal that Mab recognizes T and it is able to establish stronger bonds with T than with BSA. However, the large range of rigidities covered suggests that the bond explores different conformations in order to improve recognition. Hence, Mab shows cross-reactivity, being able to recognize antigens that are similar to B (such as T) but with different degrees of affinity, as revealed by the different correlation patterns of $f_{\rm rupt}$ and $k_{\rm bond}$.

The scattered purple regions observed in Figs. 6.5a-c suggest diversity of antibody-antigen bonds. It has been hypothesized that such diversity is the seed of improved recognition specificity by antibodies [Boeh 09, Ma 02, Mani 00, Rose 09]. Accordingly, non-specific antibody-antigen bonds explore multiple conformations in order to improve specificity and therefore they must have low activation barriers and multiple binding states. In contrast, activation barriers for specific bonds must be large to guarantee an optimized response to the infection [Stru 00, Thie 08].

6.4 Specific antibody-antigen bonds

In both the study of the maturation line of polyclonal antibodies and the cross-reactivity of monoclonal antibodies, non-specific interaction between antibodies and the carrier protein BSA were directly measured (Fig. 6.5b). Such non-specific recognition events are unavoidable, yet they do not play any relevant role in the conclusions so far obtained in this chapter. However, in order to investigate the FEL of specific antibody-antigen bonds it is crucial to remove such non-specific events to obtain conclusive results.

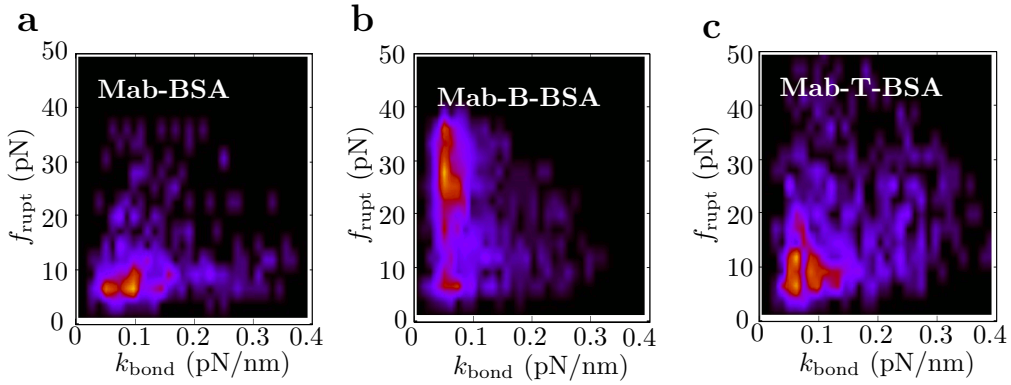

Figure 6.5: Cross-reactivity of a monoclonal antibody. 2D contour plots with histograms of f_{rupt} plotted against histograms of k_{bond} measured in pulling experiments using Mab and BSA **(a)**, B-BSA **(b)** and T-BSA **(c)**. Color legend as in Fig. 6.3.

6.4.1 Removal of non-specific interactions

The set of measured unbinding events from pulling experiments carried out using Mab and the complex B-BSA or T-BSA contains not only specific interactions between Mab and B or T, but also non-specific interactions between Mab and BSA. It is important to remove the former interactions to properly characterize the thermodynamics and kinetics of the specific bonds Mab-B or Mab-T. Several one-dimensional approaches, based only on rupture forces, have been previously suggested for this purpose [Evan 09, Wagn 11]. Here, a 2-dimensional Bayesian inference method is developed to extract the rupture force distributions of specific binding events, by including both k_{bond} and f_{rupt} in the analysis.

In what follows, the capital letter X will refer to any of the two anabolic steroids B or T. With that in mind, $p(k_{\mathrm{bond}}, f_{\mathrm{rupt}})$ for any of the two cases can be written in terms of specific and non-specific binding events as:

$$p(k_{\mathrm{bond}}, f_{\mathrm{rupt}}) = p(k_{\mathrm{bond}}, f_{\mathrm{rupt}}|X)p(X) + p(k_{\mathrm{bond}}, f_{\mathrm{rupt}}|\mathrm{BSA})p(\mathrm{BSA}). \tag{6.6}$$

where $p(X)$ and $p(\mathrm{BSA})$ are the probabilities of measuring binding to X (*i. e.* to B or T) or BSA respectively, and they satisfy $p(X) + p(\mathrm{BSA}) = 1$; $p(k_{\mathrm{bond}}, f_{\mathrm{rupt}}|X)$ and $p(k_{\mathrm{bond}}, f_{\mathrm{rupt}}|\mathrm{BSA})$ are the conditional probabilities of bond rigidities and rupture forces given the binding events Mab-X or Mab-BSA respectively.

The quantity $p(k_{\mathrm{bond}}, f_{\mathrm{rupt}}|\mathrm{BSA})$ is directly measured by carrying out pulling experiments between Mab- and BSA-coated beads (Fig. 6.5b). In order to evaluate $p(k_{\mathrm{bond}}, f_{\mathrm{rupt}}|B)$ a distance $I(k_0, f_0)$ is defined between $p(k_{\mathrm{bond}}, f_{\mathrm{rupt}})$ and $p(k_{\mathrm{bond}}, f_{\mathrm{rupt}}|\mathrm{BSA})$ as:

$$I(k_0, f_0) = \int_0^{f_0} df_{\mathrm{rupt}} \int_0^{k_0} dk_{\mathrm{bond}} \quad \left[p(k_{\mathrm{bond}}, f_{\mathrm{rupt}}) - p(k_{\mathrm{bond}}, f_{\mathrm{rupt}}|\mathrm{BSA})\right]^2 . \tag{6.7}$$

The evaluation of $I(k_0, f_0)$ for the recognition between Mab and B or T is shown in Fig. 6.6. For the two cases, the profile of $I(k_0, f_0)$ shows a region (black area in the

2D contour plot) where it can be assumed that the probability $p(k_{\rm bond}, f_{\rm rupt}|{\rm B})$ or $p(k_{\rm bond}, f_{\rm rupt}|{\rm T})$ of having specific binding Mab-B or Mab-T is negligible. Consequently, the probability of having non-specific binding events, $p({\rm BSA}) = 1 - P(X)$, can be derived from Eq. (6.6) as follows:

$$\int\int_{\text{black area}} df_{\rm rupt} dk_{\rm bond}\; p(k_{\rm bond}, f_{\rm rupt}) = \tag{6.8a}$$

$$= \underbrace{\int\int_{\text{black area}} df_{\rm rupt} dk_{\rm bond}\; p(k_{\rm bond}, f_{\rm rupt}|{\rm B})}_{=0}\, p({\rm B}) +$$

$$+ \int\int_{\text{black area}} df_{\rm rupt} dk_{\rm bond}\; p(k_{\rm bond}, f_{\rm rupt}|{\rm BSA}) p({\rm BSA}) = \tag{6.8b}$$

$$= \int\int_{\text{black area}} df_{\rm rupt} dk_{\rm bond}\; p(k_{\rm bond}, f_{\rm rupt}|{\rm BSA}) p({\rm BSA}) \tag{6.8c}$$

$$\tag{6.8d}$$

$$p({\rm BSA}) = \frac{\int\int_{\text{black area}} df_{\rm rupt} dk_{\rm bond}\; p(k_{\rm bond}, f_{\rm rupt})}{\int\int_{\text{black area}} df_{\rm rupt} dk_{\rm bond}\; p(k_{\rm bond}, f_{\rm rupt}|{\rm BSA})}. \tag{6.9}$$

In both cases under consideration, the value $p({\rm BSA}) \simeq 0.9$ is obtained.

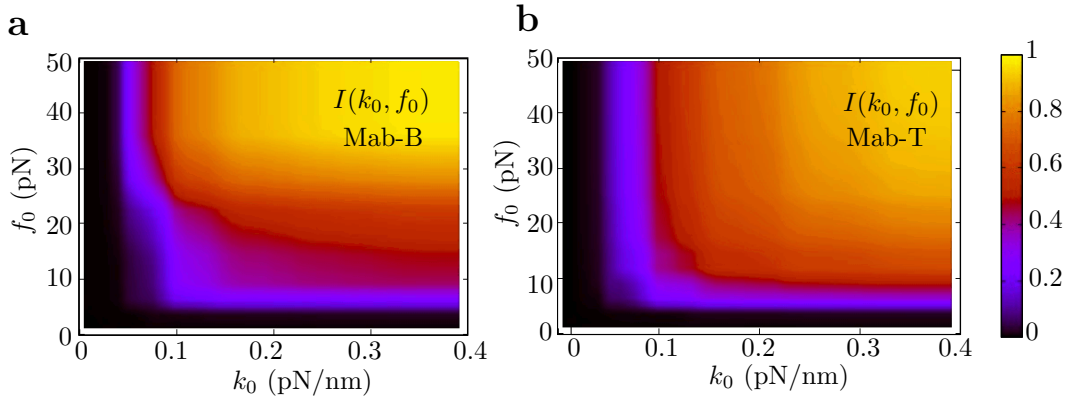

Figure 6.6: Distance $I(k_0, f_0)$ for B and T. 2D contour plot of $I(k_0, f_0)$ (Eq. 6.7) evaluated for the recognition events between Mab and the complex B-BSA **(a)** and T-BSA **(b)**.

Finally, using Eq. (6.6) the probability $p(k_{\rm bond}, f_{\rm rupt}|X)$ associated only to specific binding events can be determined (Fig. 6.7a-b for B and T). The histogram of specific rupture forces is obtained by integrating over bond rigidities:

$$p(f_{\rm rupt}|X) = \int dk_{\rm bond}\; p(k_{\rm bond}, f_{\rm rupt}|X). \tag{6.10}$$

In Fig. 6.7c-d the histogram of specific rupture forces, $p(f_{\rm rupt}|{\rm B})$ and $p(f_{\rm rupt}|{\rm T})$, obtained using the 2-dimensional Bayesian approach is shown (dotted curve). The histogram is also evaluated using other approaches where rigidity is not included, like

plain subtraction of rupture force histograms, $p(f_{\mathrm{rupt}}) - p(f_{\mathrm{rupt}}|\mathrm{BSA})$ (dashed curve) [Wagn 11], or setting a threshold force to separate specific from non-specific events (dotted curve) [Evan 09].

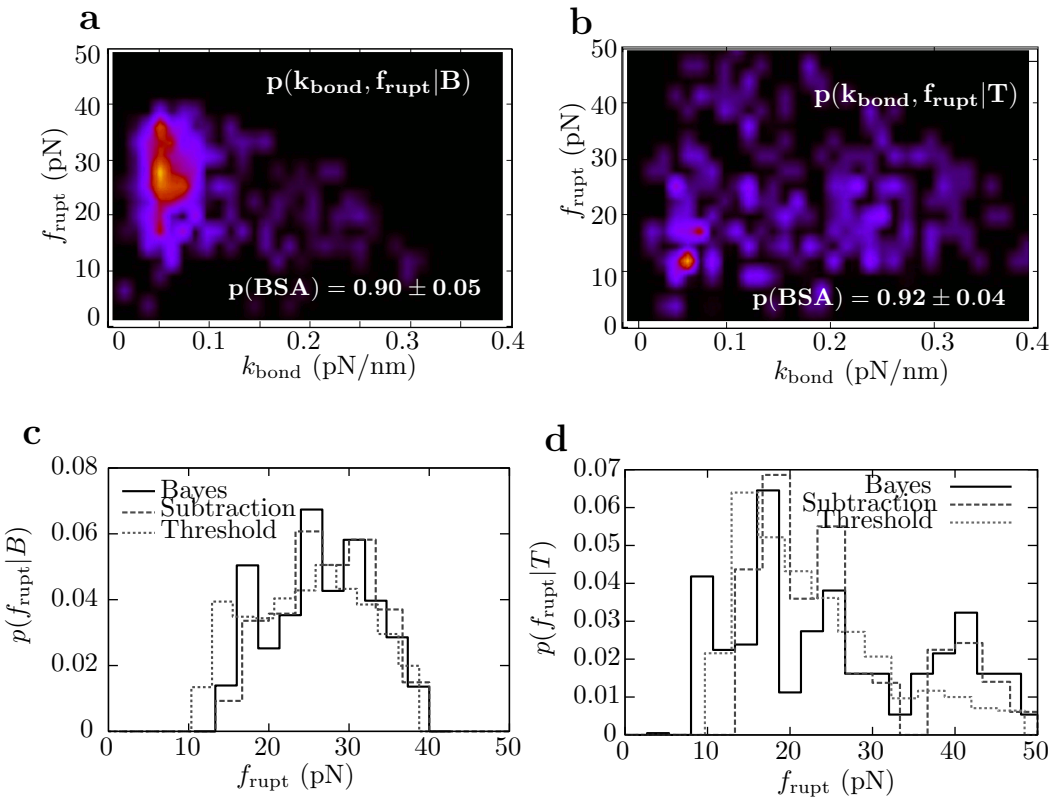

Figure 6.7: Removal of non-specific interactions. 2D contour plot of specific recognition between Mab and B **(a)** or Mab and T **(b)**. Specific rupture force histogram for the recognition between Mab and B **(c)** or Mab and T **(d)**, measured using Eq. (6.10) (solid line); by subtracting rupture force histograms $p(f_{\mathrm{rupt}})$ and $p(f_{\mathrm{rupt}}|\mathrm{BSA})$ (dashed line) [Wagn 11]; or by setting a threshold force to separate specific from non-specific events (dotted line) [Evan 09].

6.4.2 The free-energy landscape of single bonds

Mechanical bond dissociation is a statistical process that can be described as a first order Markov process for the survival probability $P(f)$ (being the probability of the bond to remain intact at force f) [Sims 99, Stri 99, Evan 09], as it is done for nucleic acid hairpins in Chapters 4 and 5. For an applied force that increases linearly with time, $f(t) = rt$, the master equation is given by:

$$\frac{dP(f)}{df} = -\frac{k(f)}{r}P(f), \tag{6.11}$$

where r is the loading rate in pN/s and $k(f)$ is the force-dependent rate of bond dissociation. It is predicted by the TST that $k(f)$ depends on the shape of the FEL of the

interaction (Section 4.1.2) [Stru 00].

A standard approach in the literature is to analytically solve Eq. (6.11) using a theoretical model for $k(f)$ and fit the resulting analytical force distribution function to experimental data. In the study of single bonds the most well-accepted models of the force-induced breakage of molecular bonds are the BE and the DHS models, already introduced in Section 4.3.2. In what follows, results obtained in each case are briefly summarized.

Bell-Evans model

In the BE model [Bell 78] a force-independent kinetic barrier is assumed, and $k(f)$ is considered to be equal to:

$$k(f) = k_m \exp\left(\frac{fx^\dagger}{k_\mathrm{B}T}\right), \qquad k_m = k_0 \exp\left(-\frac{\Delta G^\dagger}{k_\mathrm{B}T}\right), \tag{6.12}$$

where k_m is the kinetic rate of bond dissociation at zero force; $\Delta G^\dagger$ the height of the kinetic barrier (corresponding to B_{BE} in Section 4.3.2); $x^\dagger$ is the distance between the bonded and the transition state (corresponding to x_{BE} in Section 4.3.2); k_B is the Boltzmann constant; and T is the temperature (taken equal to 298 K) [Stri 99, Evan 09, Merk 99, Evan 10]. The attempt rate k_0 has been estimated in previous chapters for DNA and RNA hairpins, and values in the range $10^5 - 10^7$ s^{-1} have been obtained [Mano 06, Enge 11, Biza 12].

It is shown in Appendix C.4 that the differential equation in Eq. (6.11) can be analytically solved using the kinetic rate given by Eq. (6.12). The resulting analytical formula for the probability density function of the rupture force, $p(f_\mathrm{rupt})$, is:

$$p(f_\mathrm{rupt}) = \frac{k_m}{r} \exp\left(\frac{f_\mathrm{rupt}x^\dagger}{k_\mathrm{B}T}\right) \exp\left\{-\frac{k_m k_\mathrm{B}T}{rx^\dagger}\left[\exp\left(\frac{f_\mathrm{rupt}x^\dagger}{k_\mathrm{B}T}\right) - 1\right]\right\}. \tag{6.13}$$

Therefore, the experimental specific rupture force distributions $p(f_\mathrm{rupt}|X)$ can be fitted to Eq. (6.13) in order to obtain numerical estimations for k_m and $x^\dagger$ for both Mab-B and Mab-T bonds. Results are shown in Fig. 6.8 and summarized in Table 6.3.

	k_m (s^{-1})	$x^\dagger$ (Å)
Mab-B	0.008±0.003	6.0±1.0
Mab-T	0.05±0.01	2.5±1.0

Table 6.3: Fit of specific rupture force histograms to the BE model. To estimate $\Delta G^\dagger$ a value for k_0 must be assumed. For instance, by taking $k_0 \sim 10^6 - 10^7$ s^{-1} [Enge 11, Biza 12] the range of values $\Delta G^\dagger \sim 19 - 21\ k_\mathrm{B}T$ for Mab-B and $\Delta G^\dagger \sim 17 - 19\ k_\mathrm{B}T$ are obtained. Error bars are standard errors from fits.

Dudko-Hummer-Szabo model

The DHS model is a Kramers-based approach that uses two different analytical *ansatzs* for the potential of mean force $V(x)$ at zero force (Section 4.3.2 and Appendix C.5). The Kramers equation (4.16) is then solved in each case assuming a force-dependent potential $V(x, f) = V(x) - fx$. The results of $k(f)$ for the two approaches can be unified in a single analytical expression, which is given by:

$$k(f) = k_m \left(1 - \gamma \frac{f x^\dagger}{\Delta G^\dagger}\right)^{1/\gamma - 1} \exp\left\{\frac{\Delta G^\dagger}{k_\mathrm{B} T}\left[1 - \left(1 - \gamma \frac{f x^\dagger}{\Delta G^\dagger}\right)^{1/\gamma}\right]\right\}, \tag{6.14}$$

being $\Delta G^\ddagger$ and $x^\ddagger$ the height and position of the TS along the FEL respectively, k_m the kinetic rate of bond dissociation at zero force, and γ a parameter related to the shape of the FEL. For $\gamma = 1/2$, the FEL is assumed to have a parabolic shape, whereas for $\gamma = 2/3$ it is expected have a cubic profile [Dudk 06, Dudk 08]. For $\gamma = 1$ the expression from the BE model (Eq. 6.12) is recovered.

Again, the differential equation obtained by introducing Eq. (6.14) into (6.11) can be analytically solved and the following expression for the probability density function of rupture forces can be derived:

$$p(f_\mathrm{rupt}) = \frac{k_m}{r}\left(1 - \gamma \frac{f_\mathrm{rupt} x^\dagger}{\Delta G^\dagger}\right) \exp\left\{\frac{\Delta G^\dagger}{k_\mathrm{B} T}\left[1 - \left(1 - \gamma \frac{f_\mathrm{rupt} x^\dagger}{\Delta G^\dagger}\right)^{1/\gamma}\right] - \right.$$
$$\left. - \frac{k_m k_\mathrm{B} T}{r x^\dagger}\left(e^{\frac{\Delta G^\dagger}{k_\mathrm{B} T}\left[1 - \left(1 - \gamma \frac{f_\mathrm{rupt} x^\dagger}{\Delta G^\dagger}\right)^{1/\gamma}\right]} - 1\right)\right\}. \tag{6.15}$$

Therefore, the experimental specific rupture force distributions $p(f_\mathrm{rupt}|X)$ can be fitted to Eq. (6.15) in order to obtain numerical estimations for k_m, $x^\dagger$, and $\Delta G^\dagger$ for both Mab-B and Mab-T bonds. Results are shown in Fig. 6.8 and summarized in Table 6.4.

Mab-B				Mab-T			
γ	k_m (s^{-1})	$x^\dagger$ (Å)	$\Delta G^\dagger$ ($k_\mathrm{B}T$)	γ	k_m (s^{-1})	$x^\dagger$ (Å)	$\Delta G^\dagger$ ($k_\mathrm{B}T$)
1/2	0.003±0.001	9.5±1.0	12±4	1/2	0.03±0.01	5±1	4±1
2/3	0.005±0.002	8.0±1.0	14±4	2/3	0.04±0.01	3±1	4±1

Table 6.4: Fit of specific rupture force histograms to the DHS model. Fits were obtained by fixing the value of γ at 1/2 or at 2/3 to find values related to a parabolic or a cubic potential, respectively (Eq. 4.31a,b). Error bars are standard errors from fits.

Values obtained using either the BE or the DHS models for $x^\dagger$ for both Mab-B and Mab-T bonds are in agreement between themselves (Tables 6.3 and 6.4) and are of the same order of magnitude of estimations reported in previous single-molecule studies carried out on different antibody-antigen systems [Evan 10, Wagn 11]. For any fit, the

estimated value of k_m for the Mab-T bond is almost ten times larger than the one estimated for the Mab-B bond, indicating a larger bond lifetime at zero force (*i. e.* k_m^{-1}) for the latter. Moreover, a larger kinetic barrier, $\Delta G^\dagger \sim 13\ k_\mathrm{B}T$, is found for the bond Mab-B as compared to the one reported for Mab-T, $\Delta G^\dagger \sim 4\ k_\mathrm{B}T$, which indicates a larger kinetic stability for the former. Finally, it is observed that the rupture force histogram for the specific bond Mab-T is not accurately fitted by any model (Fig. 6.8b). All these facts suggest that the specific recognition between Mab and B has a large affinity and can be modeled with a FEL made of a single bonded state and a single escape barrier (Fig. 6.8c). In contrast, the affinity for the Mab-T bond is lower and a single escape barrier does not fit the rupture force histogram. In this case the recognition is less specific, and its FEL should be more appropriately described by several states and lower kinetic barriers [Stru 00, Raib 06, Thie 08] (Fig. 6.8d).

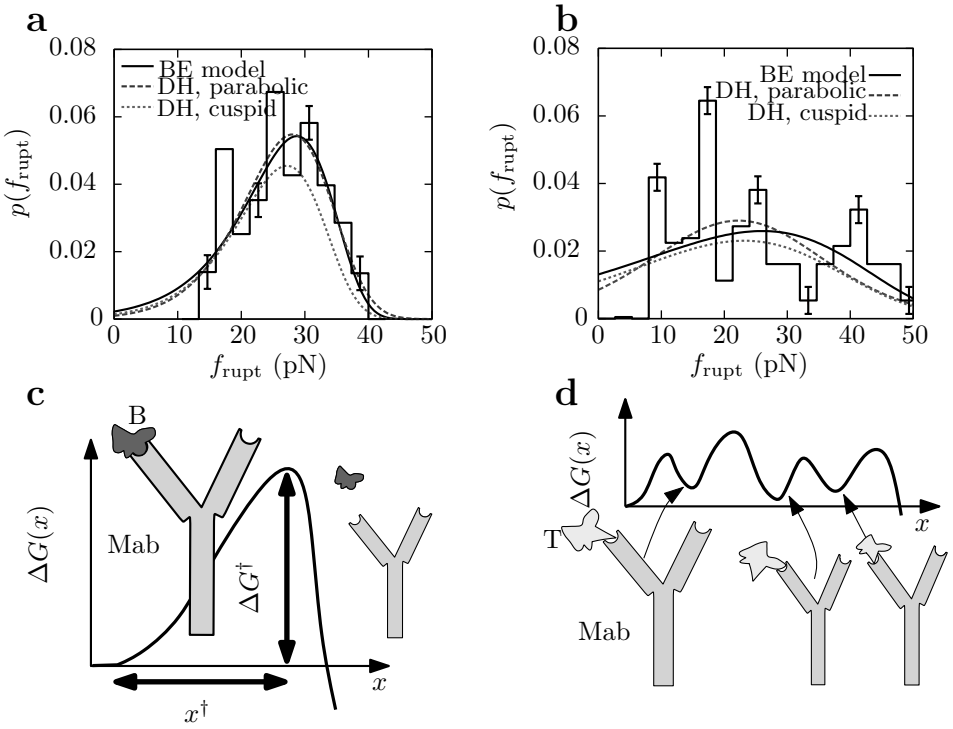

Figure 6.8: FEL of the specific bonds Mab-B and Mab-T. Experimental rupture force histogram and fits to the BE model (continuous line), the DHS model with a parabolic FEL (dashed) or a cubic FEL (dotted) for the bond Mab-B (**a**) and the bond Mab-T (**b**). **c.** Sketch of the FEL of the Mab-B bond, with a single escape barrier and a single bond state. **d.** Sketch of the FEL of the Mab-T bond, with multiple states and low kinetic barriers.

6.4.3 Study of high-affinity bonds

Taking into account that the FEL of the Mab-B bond can be described with a single escape barrier, a more detailed analysis of rupture forces was performed. Pulling

experiments between Mab-coated beads and B-BSA-coated beads were carried out at three different pulling speeds (v ∼20, 70 and 140 nm/s) and non-specific interactions between Mab and BSA were removed as described in Section 6.4.1. Figure 6.9 shows the measured histograms of f_{rupt} (panel a), and most probable rupture forces $\langle f_{\text{rupt}} \rangle$ and standard deviations σ_f as a function of the pulling speed v (panel b, top and bottom respectively).

Using the probability density function of rupture forces $p(f_{\text{rupt}})$ predicted by the BE model (Eq. 6.13 and Appendix C.4) it is obtained:

$$\langle f_{\text{rupt}} \rangle = \frac{k_{\text{B}}T}{x^{\dagger}} \log\left(\frac{x^{\dagger}v}{k_m k_{\text{B}}T}\right) \tag{6.16}$$

$$\sigma_f = 0.96\frac{k_{\text{B}}T}{x^{\dagger}}, \tag{6.17}$$

whereas using the probability density function of rupture forces $p(f_{\text{rupt}})$ predicted by the DHS model (Eq. 6.15) [Dudk 06, Dudk 08] it is obtained:

$$\langle f_{\text{rupt}} \rangle = \frac{\Delta G^{\dagger}}{\gamma x^{\dagger}} \left\{ 1 - \left[\frac{k_{\text{B}}T}{\Delta G^{\dagger}} \log\left(\frac{k_m k_{\text{B}}T e^{\frac{\Delta G^{\dagger}}{k_{\text{B}}T}+0.577}}{x^{\dagger}v} \right) \right]^{\gamma} \right\} \tag{6.18}$$

$$\sigma_f = \frac{k_{\text{B}}T\pi}{\sqrt{6}x^{\dagger}} \left[\frac{k_{\text{B}}T}{\Delta G^{\dagger}} \log\left(\frac{k_m k_{\text{B}}T e^{\frac{\Delta G}{k_{\text{B}}T}+1.064}}{x^{\dagger}v} \right) \right]^{\gamma-1} \tag{6.19}$$

Simultaneous fits to $\langle f_{\text{rupt}} \rangle$ and σ_f can be carried out using analytical expressions derived for the BE (Eqs. 6.16 and 6.17) and the DHS model (Eqs. 6.18 and 6.19). Results, shown in Fig. 6.9b and summarized in Table 6.5, are in good agreement within error bars with previous estimations (Tables 6.3 and 6.4).

FEL model	$x^{\dagger}$ (Å)	k_m (s^{-1})	$\Delta G^{\dagger}$ ($k_{\text{B}}T$)
BE	5.0±0.3	0.008±0.002	–
Parabolic	8.0±1.0	0.0002±0.0001	13±4
Cubic	7.0±1.0	0.0002±0.0001	12±3

Table 6.5: FEL parameters for Mab-B. Estimations of $x^{\dagger}$, k_m and $\Delta G^{\dagger}$ obtained from pulling experiments using the BE and the DHS fits of $\langle f_{\text{rupt}} \rangle$ and σ_f as a function of the pulling speed v (Fig. 6.9b).

6.5 Conclusions

In this chapter the relation between the decrease in flexibility with the increase in specificity in intermolecular bonds is investigated using a polyclonal system obtained before and after immunization [Alem 13]. DFS techniques have been applied to study the specificity of binding between antigens and antibodies by measuring the spectrum of bond

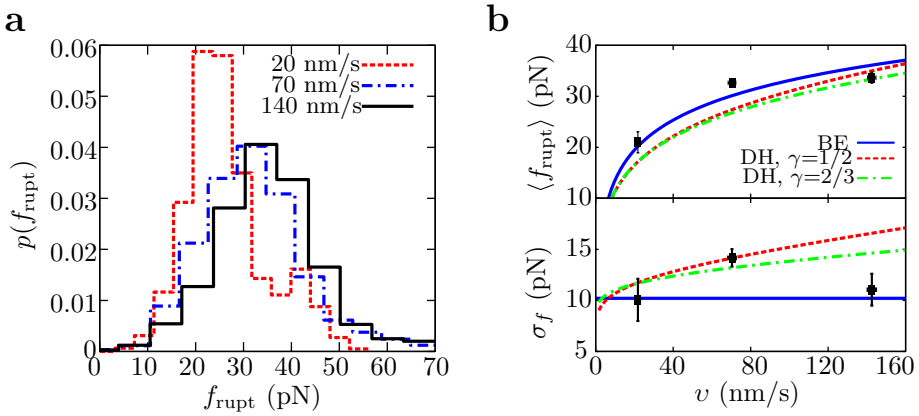

Figure 6.9: Pulling experiments for the specific bond Mab-B. a. Pulling rate dependence of the specific rupture force histogram for the interaction Mab-B measured in pulling experiments. **b.** Loading rate dependence of the average rupture forces (top) and standard deviations (bottom) for Mab-B. Simultaneous fits using the BE (solid blue line) and the DHS (dashed red for γ=1/2 and dashed-dotted green for γ=2/3) models.

elasticities and mechanical strengths in single-bond pulling experiments using OT. This novel approach has been used to unravel different binding mechanisms used by polyclonal and monoclonal antibodies tested against different antigens.

It has been observed that in order to improve the specificity in the antibody-antigen recognition, antibodies generated after immunization establish stronger bonds with lower flexibilities than antibodies prior to immunization (Fig. 6.2). Consequently, bonds strengthen and become more rigid as ligand-receptor affinity increases. Results on the cross-reactivity of monoclonal antibodies also suggest that non-specific antibody-antigen interactions explore different conformations to increase bond-specificity and improve molecular recognition. This is revealed by the broad spectrum of rigidities that can be measured in pulling experiments for different non-specific interactions. Such diversity might be due both to an heterogeneous population of polyclonal antibodies present in the sample, and to a diversity of conformational sub-states for each monoclonal antibody that are functional and bind to the antigen. This behavior is in agreement with previous studies [Berg 99, Jern 55, Kosh 58, Ma 02, Sund 02, Paul 46], where the non-specific antibody-antigen bond is modeled by combining the conformational-change and the induced-fit mechanisms of binding. In the first case, antigens and antibodies explore different conformations before binding and transient molecular states interact when structures match each other. In the second case, initial unoptimized bonds explore the conformational space and reorganize their structure in order to increase the binding strength. Such multiplicity of conformational states or kinetic barriers is not surprising. Seminal investigations have already shown functionality of multiple conformational states characterized by different activation barriers [Aust 74, Solo 11]. These results suggest a crucial role of bond heterogeneity in the humoral immune system. Overall, they reinforce the intriguing possibility of understanding the immune system in terms of evolutionary allosteric macromolecular ensembles governed by the flexural

property of antibody-antigen bonds.

Future studies should address the detailed continuous adaptation in time of the humoral immune system to a given infection, by characterizing the full time-evolution of the degree of flexibility of antibody-antigen bonds. Ultimately, the mechanistic approach might be suitable to unravel and characterize different binding mechanisms used by the antibody repertory to identify infectious agents.

Chapter 7

Mechanical unfolding and folding of protein Barnase

Proteins are essential molecules for living organisms. They perform a wide variety of biological functions, such as regulation, transport, motion, structure, or storage, among others. A protein consists of a chain of aminoacids that folds into itself in a very specific structure. Shape and intra-chain connectivity determine how the protein interacts with its environment, and therefore both are closely related to the function. In contrast to nucleic acid hairpins, whose folded state can be easily determined according mainly to the formation of Watson-Crick basepairs, it is very difficult to predict which is the native structure of a folded protein given its sequence of aminoacids. Hydrogen bonds, Van der Waals forces, electrostatic interactions, hydrophobic forces, and steric effects determine in the protein folding process. The huge amount of possible degrees of freedom requires the development of efficient computational techniques and theoretical models based on statistical physics to accurately predict the native state and the FEL of a protein.

The ability of DFS experiments combined with the theoretical framework provided by TST to predict the molecular FEL has been deeply tested using nucleic acid hairpins. In that case, the one-to-one relation between the molecular end-to-end distance and the molecular structure through the number of open basepairs provides a great insight about states that drive the molecule along its FEL during the mechanical unfolding and folding processes. However, in the case of proteins the correspondence between molecular extension and structure is not straightforward: proteins are typically structurally much more complex than nucleic acid hairpins and sequential configurations in the folding/unfolding pathway do not necessarily imply the absorption/release of sequential aminoacids along the peptide chain, but might imply the folding/unfolding of separated domains of the protein structure. For this reason, a multidimensional FEL is indispensable to unravel the molecular folding/unfolding process of many proteins. Even though DFS experiments by themselves only grant access to the molecular extension, when combined with simulation, protein engineering, or bulk techniques, they can be extremely helpful to describe the mechanical unfolding and folding pathways of proteins and to obtain key features of the molecular structure and thermodynamical stability of

proteins [Diet 04, Schl 04, Cecc 05, Mick 07, Shan 10, Gebh 10, Stig 11, Heid 14].

Barnase is a a bacterial globular protein that degrades RNA in the absence of its inhibitor Barstar. It has been deeply studied using a protein-engineering approach [Mato 92, Serr 92] and doubts about the possible existence of an intermediate state along the folding pathway have been raised [Take 00, Fers 00]. Made of 110 aminoacids, Barnase does not contain any disulfide bond in its native state, which makes it an ideal protein to be pulled using OT. In this chapter, DFS experiments are performed to extract the elastic properties of the peptide chain and to unravel the unfolding and folding mechanisms of Barnase under mechanical denaturation.

7.1 Pulling experiments

In order to perform DFS experiments on Barnase, both N- and C- terminus of the protein are linked to 500-basepairs long dsDNA handles [Kell 97, Tskh 97b, Cecc 05, Borg 08, Shan 10, Heid 12]. Details of the molecular synthesis are provided in Appendix A.4. One end of the molecular construct is attached to a SA bead and the other to an AD bead using biotin-streptavidin and digoxigenin-antidigoxigenin bonds, respectively. The SA bead is immobilized in the tip of the micropipette and the AD bead is captured in the optical trap. A sketch of the experimental setup is shown in Fig. 7.1a.

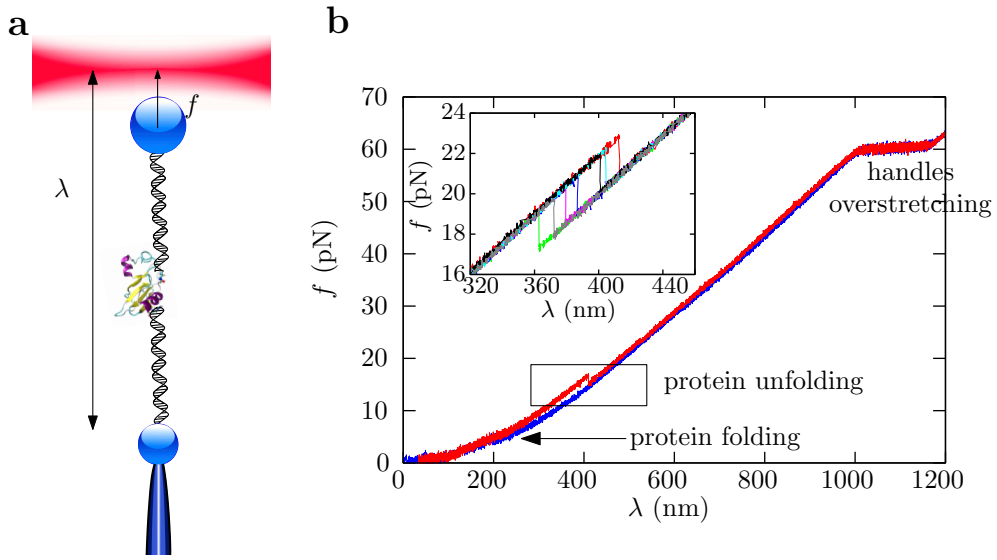

Figure 7.1: Force-spectroscopy experiments with Barnase. **a.** Not-to-scale scheme of the molecular experimental setup: Barnase is linked to two dsDNA handles and the whole construct is inserted between two polystyrene beads. One bead is immobilized in the tip of the micropipette by air suction whereas the other is captured in the optical trap. **b.** Example of unfolding and folding FDC measured in pulling experiments. Inset: collection of unfolding FDC.

Figure 7.1b shows an unfolding (red) and a folding (blue) FDC obtained in a pulling experiment while stretching and releasing the molecular construct. The drop in force

measured in the unfolding curve is related to the mechanical unfolding of the protein, which apparently behaves as a two-state system. Unfolding forces change upon independent repetitions of identical pulling protocols, as is the case for DNA hairpins (inset in Fig. 7.1b). At ~ 60 pN, the overstretching transition of the dsDNA handles is observed. The folding of Barnase usually occurs at forces below 5 pN. This transition is difficult to observe and, in some occasions, the protein does not fold as revealed by the absence of a force rip in the next unfolding curve. To increase the probability of folding a time pause can be introduced between the end of a folding process and the beginning of the subsequent unfolding.

As it can be seen in Fig. 7.1b, there are hysteresis effects between unfolding and folding trajectories measured in pulling experiments with Barnase, which implies that the pulling speed is faster than the rate at which molecular states equilibrate. Such hysteresis effects subtly increase with the pulling speed. Similar FDC patterns have also been observed with the protein Ribonuclease H [Cecc 05].

7.2 Elastic response of the peptidic chain

For the case of Barnase, both the folded and the unfolded force-branches can be measured in a wide interval of forces, ranging approximately from 4 to 20 pN (Fig. 7.1b). Following the same steps as in Chapter 3 for DNA hairpins [Alem 14], it can be demonstrated that the effective stiffness of each branch, k_{eff}^{F} and k_{eff}^{U} respectively, satisfy:

$$\frac{1}{k_{\text{eff}}^{F}} = \frac{1}{k_{\text{b}}} + \frac{1}{k_{\text{h}}} + \frac{1}{k_{\text{d}}} \tag{7.1a}$$

$$\frac{1}{k_{\text{eff}}^{U}} = \frac{1}{k_{\text{b}}} + \frac{1}{k_{\text{h}}} + \frac{1}{k_{\text{c}}}, \tag{7.1b}$$

where k_{b} is the stiffness of the optical trap, k_{h} is the stiffness of the dsDNA handles, k_{d} is the stiffness of the protein when it is folded in its native globular conformation, and k_{c} is the stiffness of Barnase when it is unfolded, *i. e.* it is the stiffness of the peptide chain.

The elastic response of the peptide chain is assumed to satisfy the ideal WLC model, which as already shown in Chapter 3 (Eq. 3.11) can be analytically described by the following interpolation formula:

$$k_{\text{c}}(x) = \frac{k_{\text{B}}T}{PL}\left[\frac{1}{2\left(1 - x/L\right)^{3}} + 1 + \sum_{n=2}^{7} n a_n \left(\frac{x}{L}\right)^{n-1}\right], \tag{7.2}$$

being here x the equilibrium end-to-end distance of the peptide chain, P the persistence length and $L = n_{\text{aa}} d_{\text{aa}}$ the contour length of the peptide. n_{aa} is the total number of aminoacids, equal to 110 for Barnase, while d_{aa} is the equilibrium distance between consecutive aminoacids. The numerical coefficients a_n ($n = 2, \ldots, 7$) depend on the analytical expression used to model the WLC behavior. According to reference [Bust 94], these can be taken equal to zero, whereas in reference [Bouc 99] these are taken equal

to $a_2 = -0.5164228$, $a_3 = -2.737418$, $a_4 = 16.07497$, $a_5 = -38.87607$, $a_6 = 39.49944$, and $a_7 = -14.17718$.

On the other hand, the elastic response of the folded protein under applied mechanical force is described as made for the double helix diameter of nucleic acid hairpins: the globular structure is modeled as a single bond of length d that is oriented under mechanical force as a single magnetic dipole does under applied magnetic field. The result is identical to the analytical expression for the FJC elastic model with identical Kuhn and contour lengths, d. For Barnase, d is taken equal to 3 nm, which is the end-to-end distance of the protein at zero force and has been measured by X-ray crystallography [Wilt 09]. Consequently, according to Eq. (3.10), k_{d} satisfies:

$$k_{\mathrm{d}}^{-1}(f) = \frac{d^2}{k_{\mathrm{B}}T}\left[-\frac{1}{\sinh^2\left(\frac{fd}{k_{\mathrm{B}}T}\right)} + \left(\frac{k_{\mathrm{B}}T}{fd}\right)^2\right] \tag{7.3}$$

Both $k_{\mathrm{eff}}^F(f)$ and $k_{\mathrm{eff}}^U(f)$ can be estimated from the slopes of the FDC along the folded and unfolded force-branches measured in pulling experiments (Fig. 7.2a). The subtraction of their inverse values at a given force f gives:

$$\frac{1}{k_{\mathrm{eff}}^U(f)} - \frac{1}{k_{\mathrm{eff}}^F(f)} = \frac{1}{k_{\mathrm{c}}(f)} - \frac{1}{k_{\mathrm{d}}(f)} \tag{7.4}$$

Therefore, the experimental measurement of $(k_{\mathrm{eff}}^U)^{-1} - (k_{\mathrm{eff}}^F)^{-1}$ can be fitted to the theoretical expression for $k_{\mathrm{c}}^{-1} - k_{\mathrm{d}}^{-1}$ (Eqs. 7.2 and 7.3), taking P and d_{aa} as free parameters, in order to obtain the elastic response of the peptide chain. Results from such fit are shown in Fig. 7.2b and summarized in Table 7.1. Recovered values for P and d_{aa} using the WLC interpolation formula proposed by Bouchiat and collaborators in [Bouc 99] are in very good agreement with values proposed in previous single-molecule studies [Tskh 97b, Shan 10].

WLC model	P (nm)	d_{aa} (nm/aa)
[Bust 94], $a_i = 0$	0.85±0.05	0.36±0.01
[Bouc 99], $a_i \neq 0$	0.60±0.05	0.34±0.01

Table 7.1: Elastic parameters for the peptide chain. Results of the fit of $\left((k_{\mathrm{eff}}^U)^{-1} - (k_{\mathrm{eff}}^F)^{-1}\right)^{-1}$ to the theoretical expression given by Eqs. (7.2), (7.3) and (7.4) using the interpolation formula for the WLC elastic model proposed in references [Bust 94] (top) and [Bouc 99] (bottom). Error bars are standard errors from the fit.

The jump in force measured in the unfolding events that take place along FDC in pulling experiments for two-state DNA hairpins is related to the elastic properties of ssDNA according to Eq. (3.5) (Chapter 3). An identical formulation can be applied in the case of proteins that behave as two-state systems along unfolding, as is the case for Barnase. Hence, it is satisfied that:

$$\frac{\Delta f}{k_{\mathrm{eff}}^F} = x_{\mathrm{c}}(f) - x_{\mathrm{d}}(f), \tag{7.5}$$

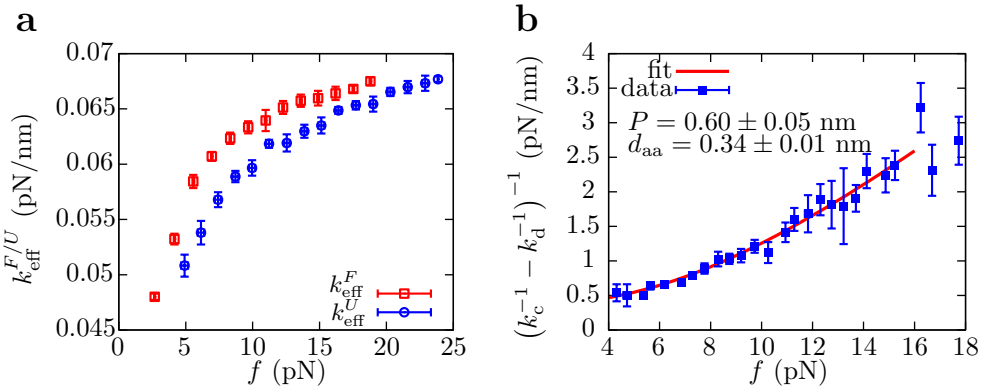

Figure 7.2: Elastic behavior of the peptide chain. a. Effective stiffnesses along the folded and unfolded branches, $k_{\rm eff}^F$ and $k_{\rm eff}^U$ respectively, obtained for one molecule. **b.** Experimental measurement of $\left((k_{\rm eff}^U)^{-1}-(k_{\rm eff}^F)^{-1}\right)^{-1}$ and fit to Eq. (7.4) using Eqs. (7.2) and (7.3). Numerical results for P and $d_{\rm aa}$ are given in Table 7.1. Data is obtained by averaging over 11 molecules (error bars are standard statistical errors).

where Δf is the size of the sudden drop in force measured when Barnase unfolds along an unfolding experiment, f is the force acting on the system in the unfolded branch during the transition, and $x_{\rm c}(f)$ and $x_{\rm d}(f)$ are the equilibrium end-to-end distances of the peptide chain and the folded protein at force f that are described according to the WLC and the FJC elastic models, respectively. According to Eq. (7.5), the pair of values P, $d_{\rm aa}$ (Table 7.1) obtained by fitting the force-dependent stiffness to Eq. (7.4) allow to estimate the quantity $x_{\rm c}(f)-x_{\rm d}(f)$. Values recovered are in good agreement with experimental results obtained for $\Delta f/k_{\rm eff}^F$. Thus, Barnase behaves as a two-state system upon mechanical unfolding.

7.3 Free-energy landscape of Barnase

There is not a unique approach to theoretically determine the molecular FEL of small proteins. The use of the Random Energy model, mean-field approaches or lattice models, among other statistical techniques, has been extensive in the determination of the native structure of a sequence of aminoacids and in the characterization of the molecular folding pathway [Bryn 87, Broo 98, Colu 03, Juni 08, Lei 10]. Despite of the fact that the specific choice of the reaction coordinate might be disputed, typical theoretical approaches use the radius of gyration, the fraction of native contacts between protein residues, or the mean distance of intermediate structures to the native state. However, it has been argued that the kinetic properties of proteins might depend on the reaction coordinate used [Hinc 10]. In fact, it is difficult *a priori* to say which are the best reaction coordinates to characterize the configurations along the unfolding and folding reaction pathways for proteins. Additionally, experimental approaches usually only grant access to a low-dimensional projection of the FEL, nevertheless providing valuable information about the molecular kinetic properties [Borg 08, Ferr 11, vZol 13].

Here, the unfolding and folding behavior of Barnase is investigated using DFS experiments carried out with optical tweezers. DFS experiments give access to the protein end-to-end distance, which in what follows will be the chosen reaction coordinate [Best 08]. Barnase is a model protein that has been extensively studied using protein engineering, fluorescence and X-ray spectroscopy, among other bulk techniques [Mato 90, Mato 92, Serr 92]. Hence, information extracted from DFS is extremely useful to complement already available data. In the following sections, the unfolding and folding pathways of Barnase under mechanical denaturation are unraveled, the force-dependent unfolding and folding kinetic rates determined, and basic properties of the FEL extracted.

7.3.1 The unfolding pathway and unfolding kinetic properties

A sudden drop in force is observed along unfolding curves measured in pulling experiments performed with Barnase (Fig. 7.1). As already discussed in Section 7.2, this force jump is attributed to the complete unfolding of Barnase due to the action of applied mechanical force (Eq. 7.5). Hence, this protein behaves according to a two-state system upon unfolding.

Figure 7.3a shows histograms of unfolding forces, $\rho(f_U)$, obtained from pulling experiments performed at different pulling speeds. The unfolding force distributions are slightly shifted to higher force values as the pulling speed increases. This result can be better visualized in Fig. 7.3b, where the average unfolding force $\langle f_U \rangle$ is plotted against the loading rate r. Both the BE and the DHS model provide analytical expressions which can be fitted to this data in order to obtain estimations of the kinetic rate of unfolding at zero force, k_m, the distance between state N and the TS, $x_{N-\mathrm{TS}}$, and the height of the kinetic barrier, $\Delta G_{N-\mathrm{TS}}$ (Eqs. C.30, C.38). Results are shown in Fig. 7.3b for the different models (red-solid, green-dashed and blue-dotted lines for the BE, the DHS with $\gamma = 1/2$ and the DHS with $\gamma = 2/3$ models, respectively), and numerical values are summarized in Table 7.2. Discrepancies are observed between the values obtained from the fit to the BE and the DHS model which will be further discussed below.

Model	k_m (s^{-1})	$x_{N-\mathrm{TS}}$ (nm)	$\Delta G_{N-\mathrm{TS}}$ ($k_\mathrm{B}T$)
BE	$(4 \pm 2)\times 10^{-8}$	3.5±0.4	–
DHS, $\gamma = 1/2$	$(7 \pm 1)\times 10^{-10}$	6±1	31±2
DHS, $\gamma = 2/3$	$(10 \pm 3)\times 10^{-11}$	6±1	28±5

Table 7.2: Parameters for the FEL of Barnase estimated using the BE and DHS models by fitting $\langle f \rangle$ versus the pulling speed. Numerical values for k_m, $x_{N-\mathrm{TS}}$ and $\Delta G_{N-\mathrm{TS}}$ obtained by fitting the average unfolding force $\langle f_U \rangle$ versus the pulling speed to analytical expressions provided by the BE and the DHS model, with $\gamma = 1/2$ or $2/3$ depending on the shape assumed for the FEL (Eqs. C.30, C.38, respectively). Results are shown in Fig. 7.3b.

Unfolding of Barnase can be modeled as a first order Markov process, as it was done for two-state nucleic acid hairpins. Therefore, the survival probability of the native state N along an unfolding experiment satisfies Eq. (4.22a) and the force-dependent unfolding

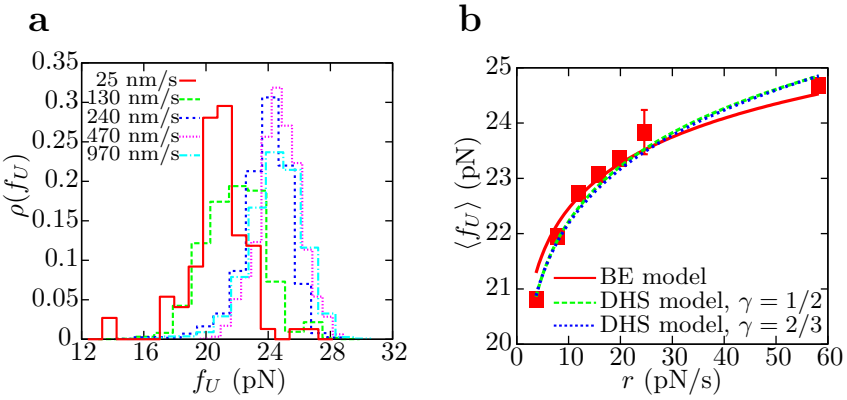

Figure 7.3: Mechanical unfolding of Barnase. a. Unfolding force histograms obtained at different pulling speeds. **b.** Dependence of average unfolding forces $\langle f_U \rangle$ with loading rate r ($r = vk_{\text{eff}}$, where $k_{\text{eff}} \sim 0.069$ pN/nm). Fits to the analytical expressions provided by the BE (red-solid line) and DHS (green-dashed and blue-dotted lines for $\gamma = 1/2$ and $\gamma = 2/3$, respectively) models are shown.

kinetic rate $k_{N\to U}(f)$ can be extracted from a collection of unfolding forces measured in pulling experiments (Eq. 4.23a). Results obtained at different loading rates are shown in Fig. 7.4a, where it can be observed that all experimental points fall into the same master curve within error bars.

Both the BE and the DHS models provide analytical expressions for $k_{N\to U}(f)$ that can be fitted to experimental data in order to get estimations for k_m, $x_{N-\text{TS}}$ and $\Delta G_{N-\text{TS}}$ (Eqs. C.26 and C.36, respectively). In the case of the BE model, it was already discussed in Chapter 4 that the predicted linear dependence between $\log k_{N\to U}(f)$ and f it is only valid in a narrow range of forces. However, in this case the experimentally measured $k_{N\to U}(f)$ covers the force interval between 12 and 28 pN, which is extremely large. Hence, the restricted range of forces $f \in [12, 18]$ pN will be used in order to fit $k_{N\to U}(f)$ to the analytical expression provided by the BE model (Eq. C.26). In contrast, the DHS model does indeed predict a convex curvature for $\log k_{N\to U}(f)$ versus f and therefore the entire range of forces will be used to perform the fit to Eq. (C.36). Results for the two approaches can be seen in Fig. 7.4a and numerical values are summarized in Table 7.3. Again, differences between the two models are obvious. Moreover, estimations for k_m, $x_{N-\text{TS}}$ and $\Delta G_{N-\text{TS}}$ obtained using the BE model by fitting the rate-dependent average unfolding force (Table 7.2) and the force-dependent unfolding kinetic rate (Table 7.3) are different, probably because the range of forces used to perform the fits are different in the two cases. In contrast, the DHS model (specially the case of $\gamma = 2/3$) provides similar values within error bars for the two fits. Hence, the interpretation of these results will be performed according to the DHS model.

According to the DHS model, $x_{N-\text{TS}} = 7.5$ nm (obtained by averaging results obtained for $\gamma = 1/2$ and $\gamma = 2/3$), which at 20 pN corresponds to the release of ~ 32 aminoacids. Interestingly, this number of released aminoacids is equal to the number of aminoacids contained along unstructured regions close to the N–terminus of the protein (blue regions in Fig. 7.4b). This is in reasonable agreement with previous ob-

Model	k_m (s^{-1})	x_{N-TS} (nm)	ΔG_{N-TS} (k_BT)
BE	$(2.8 \pm 0.5)\times 10^{-10}$	4.6±0.2	–
DHS, $\gamma = 1/2$	$(1.1 \pm 0.5)\times 10^{-13}$	9±1	35±2
DHS, $\gamma = 2/3$	$(6.0 \pm 0.4)\times 10^{-11}$	6.0±0.4	28±1

Table 7.3: Parameters for the FEL of Barnase estimated using the BE and DHS models by fitting $k_{N\to U}(f)$ versus f. Numerical values for k_m, x_{N-TS} and ΔG_{N-TS} obtained by fitting the force-dependent unfolding kinetic rate to the analytical expressions provided by the BE and the DHS model with $\gamma = 1/2$ or $\gamma = 2/3$ (Eqs. C.26 and C.36, respectively). Results are shown in Fig. 7.4a.

servations from protein engineering studies reported by L. Serrano and collaborators in reference [Serr 92], where they find that the first events in Barnase unfolding are, sequentially, the unfolding of some loops close to the N–terminus (*i. e.* unstructured regions), the unwinding of the α-helices, and the weakening of the whole protein structure despite that β-sheets remain intact. Hence, according to the DHS and in agreement with the protein engineering-based approach, in DFS experiments the TS upon unfolding is characterized by the unfolding of unstructured regions that are located close to the N–terminus. Additionally, the DHS model provides a value for the height of the kinetic barrier, $\Delta G_{N-TS} \sim 30\ k_BT$, which is also in good agreement with previous measurements performed with protein engineering and nuclear magnetic resonance [Mato 90].

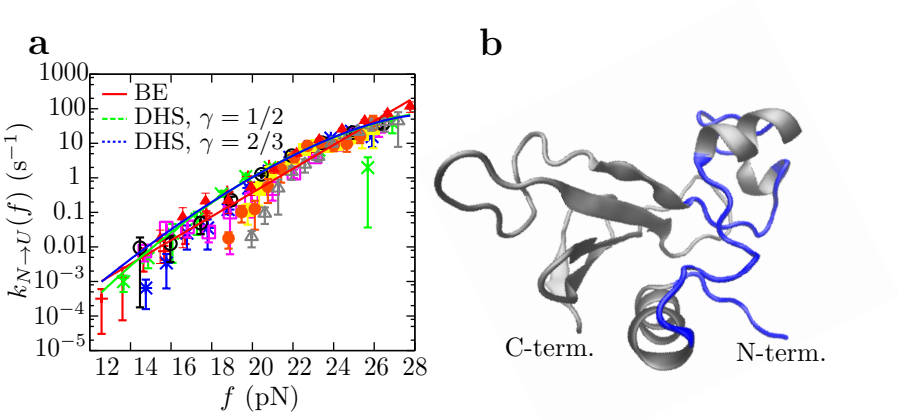

Figure 7.4: Mechanical unfolding and TS of Barnase. a. Unfolding kinetic rate as a function of force. Each symbol is associated to a different pulling speed in the range 25–970 nm/s. Fits to the analytical expressions provided by the BE (red-solid line) and DHS (green-dashed and blue-dotted lines for $\gamma = 1/2$ and $\gamma = 2/3$, respectively) models are shown. **b.** Structure of Barnase in the native state. Unstructured regions (made of a total of 32 aminoacids) indicated in blue are the first to unfold and hence correspond to the Barnase TS along the unfolding process. This makes protein engineering-based measurements and DFS experiments analyzed according to the DHS model compatible with each other.

A different lecture can be performed on the profile of the experimentally measured $k_{N\to U}(f)$. Despite the fact that the convex curvature observed in the logarithmic plot of the unfolding kinetic rate against force is well-fitted by the DHS model and a reasonable interpretation for x_{N-TS} can be given(Fig. 7.4), such curvature can also be explained

in terms of the Leffler-Hammond effect [Leff 53, Hamm 55]. As already discussed in Sections 4.3.1 and 5.1.1, the Leffler-Hammond postulate predicts that $x_{N-\mathrm{TS}}$ decreases as force increases. This behavior implies a decrease in the slope of the logarithm of $k_{N\to U}(f)$ at large forces, since $x_{N-\mathrm{TS}} = -\partial B_{NU}(f)/\partial f$ (Eq. 4.19) and $B_{NU}(f) \propto -\log k_{NU}(f)$ (Eq. 4.25a).

Thus, the BE model is not accurate enough to determine the main features of the FEL. Both the fit of the average unfolding force as a function of the pulling speed and the fit of the force-dependent unfolding kinetic rate in a restricted range of forces give underestimated values for $x_{N-\mathrm{TS}}$. In contrast, the DHS is able to reproduce results obtained using different experimental techniques, such as protein engineering and nuclear magnetic resonance [Mato 90, Serr 92]. Therefore, the Leffler-Hammond effect might play a minor role in the mechanical unfolding of Barnase in the range of forces explored, and the main features of the FEL may be determined using the DHS model exclusively.

7.3.2 The folding pathway

In pulling experiments the folding transition of Barnase is usually observed as a smooth change in the slope of the FDC (Fig. 7.1b) and the folding force is difficult to measure. As a consequence, the folding kinetic rate $k_{N\leftarrow U}(f)$ cannot be estimated using Eq. (4.23b). Here, pulling and passive-mode experiments are combined in order to extract $k_{N\leftarrow U}(f)$ (Fig. 7.5a). The protocol works as follows:

- First, the molecular system is set at a maximum value of the force, $f_{\max}$, where the protein is always unfolded, and then the system is released at a constant pulling speed v until a preset value of the force f_p, between 4 and 6 pN, is reached.
- Once f_p is reached, the system is let to evolve for a few seconds ($\Delta t \sim 15$ s) at a constant trap-pipette position λ (experimental passive mode, Section 2.3.2). During this time-interval, the protein can either fold, which is observed as a small jump in the value of force (red curve in Fig. 7.5a), or remain in the unfolded state (blue curve in Fig. 7.5a).
- Next, the molecule is stretched at the constant pulling speed until $f_{\max}$ is reached and the protocol starts again.

Figure 7.5b shows different examples of folding traces measured at the preset force f_p =3.6 pN with a sampling rate of 4 kHz during the time interval where the protocol is operating in the passive mode. A jump in force takes place when the protein folds. No hopping is observed for the case of Barnase: once the protein folds it remains folded for the rest of the time. The size of the force jump ($\sim$0.5 pN in the examples depicted in Fig. 7.5b) is compatible to the absorption of $\sim$ 110 aminoacids, which corresponds to the total number of aminoacids of Barnase. This has been estimated using Eq. (7.5), where the elastic properties of the peptide chain and the numerical value of the effective stiffness k^F_{eff} are taken from Table 7.1 and Fig. 7.2a, respectively. This result suggests that Barnase behaves as a two-state folder under the action of mechanical force.

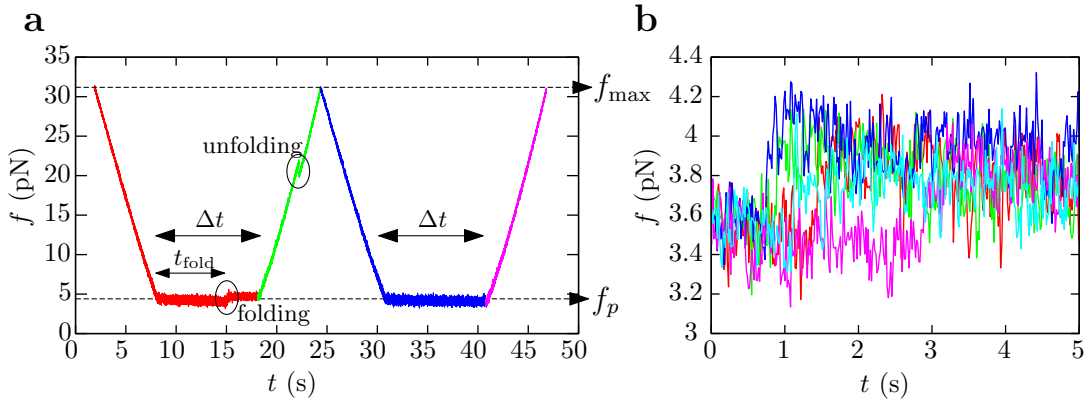

Figure 7.5: Folding of Barnase. a. Combination of pulling and passive-mode experiments to study the folding behavior of Barnase at low forces: standard pulling experiments are performed between the maximum force $f_{\max}$ and a preset force f_p, where the molecular system is held at a constant trap-pipette distance λ (passive-mode) for a time interval Δt. During this time interval, the protein can either fold (red curve) or remain in the unfolded state (blue curve). **b.** Examples of folding trajectories observed at low forces ($f_p = 3.6$ pN) during the passive-mode interval of the experimental protocol (each trajectory is indicated with a different color). Data has been filtered at 4 ms. Each force jump corresponds to a different folding event.

In order to study whether fast intermediate states are present in the folding pathway of Barnase, passive-mode traces were also recorded at a high sampling rate (50 kHz). An example of an experimental trace is shown in Fig. 7.6a (red). The histogram of the points recorded for a single folding trajectory is well-fitted by a double-Gaussian function (Fig. 7.6a, right), which also suggests that Barnase folds in a two-state manner. In fact, despite being two Gaussians very close to each other, each one can be assigned without error to either state U (initial state at low forces) or N (final state at high forces). To reduce Brownian fluctuations and further check that Barnase behaves as a two-state system upon mechanical folding, different folding trajectories were aligned at the center of the transition using 1-second data trace anterior and posterior to the transition (Fig. 7.6a, blue). Then, they were averaged on temporal windows of 0.5 ms, as it is done in reference [Neup 12]. Results of such alignment obtained from folding traces recorded at different values of f_p are shown in Fig. 7.6b, where a zoom around the transition is performed. At first sight, no evidence of intermediate states is found.

In order to check if deviations from the two-state behavior do take place, the force-time curves resulting from the alignment are fitted to the following sigmoid function:

$$f(t) = \frac{f_{\max} + f_{\min}}{2} + \frac{f_{\max} - f_{\min}}{2} \tanh\left(\frac{t - t_0}{\tau}\right), \tag{7.6}$$

where $f_{\min}$ and $f_{\max}$ are the minimum and the maximum forces of the trace, corresponding to the unfolded and the folded state of Barnase, respectively; t_0 is the time corresponding to the center of the transition and here it is arbitrarily fixed at 1 s; and τ gives a measurement of the total time for the folding transition to occur. Fits to Eq. (7.6) are shown in Fig. 7.6b for aligned curves obtained with folding trajectories measured at

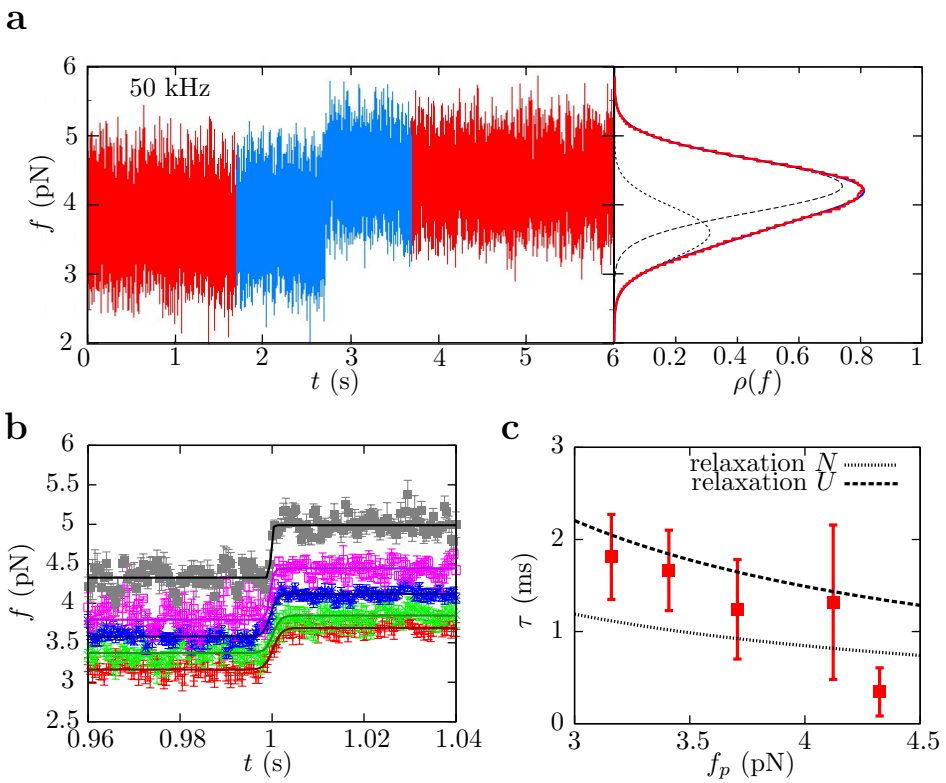

Figure 7.6: Folding of Barnase observed under passive-mode experiments at high temporal resolution. a. Folding of Barnase recorded at 50 kHz of sampling rate measured using the passive-mode experimental protocol. Red data is a complete trajectory. Data used for the alignment with other folding trajectories is represented in blue. **b.** Example of force-time curves obtained by aligning at the center of the transition different folding trajectories obtained at the same preset force. **c.** Measured τ as a function of the preset force, and relaxation time of the experimental setup (made by the bead in the optical trap, the handles and the protein) as a function of force when Barnase is at state N (dashed line) or U (dotted line).

different preset forces. It can be observed that the fit to the two-state function is very accurate. Moreover, it is again observed that $\Delta f = f_{\max} - f_{\min}$ does correspond to the release of ∼110 aminoacids for all the cases, according to a two-state folding process (Eq. 7.5).

Figure 7.6c shows the dependence of τ on the preset force, f_p. A tendency of τ to decrease with f_p is observed. To properly interpret the temporal parameter τ, the power spectrum of the passive traces recorded at a sampling rate of 50 kHz is computed. In Fig. 7.7a power spectra $S(\nu)$ obtained for 1-second long passive traces measured at 4 pN when Barnase is in state U (red) or in state N (blue) are shown. Both power spectra can be well-fitted by a double Lorentzian function,

$$S(\nu) = \frac{A_1}{\nu_1^2 + \nu^2} + \frac{A_2}{\nu_2^2 + \nu^2}, \tag{7.7}$$

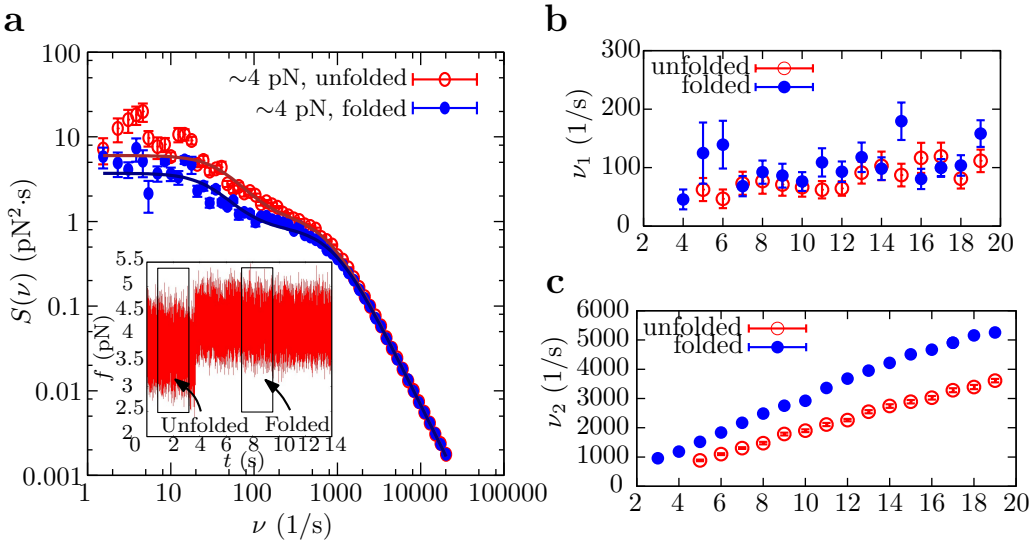

Figure 7.7: Power spectrum of Barnase. a. Inset: example of a passive trace measured at ~ 4 pN. Boxed regions are the traces where Barnase is unfolded (left) or unfolded (right) to which the power spectrum is calculated. Main panel: example of power spectrum of passive traces measured at 4 pN obtained when Barnase is folded or folded. Experimental results are fitted to a double Lorentzian function. Force-dependence of the slow-frequency (**b**) and fast-frequency (**c**) modes.

where ν_1 and ν_2 are slow- and fast-frequency modes, and A_1 and A_2 are the corresponding amplitudes. It can be seen that both frequency modes increase with force, being ν_1 independent of the configurational state of Barnase (N or U), while ν_2 indeed depends on the state of Barnase. Similar results have been previously reported in the study of dsDNA tethers using double trap OT setups [Ribe 13], where it is justified that the slow contribution (ν_1) is due to misalignment and fluctuations along the optical axis, while the fast contribution (ν_2) is due to fluctuations along the pulling direction. Hence, the relaxation time of the molecular setup at a given force when Barnase is either folded or unfolded can be estimated using the inverse of the corresponding fast-frequency mode. In Fig. 7.6c the inverse of the extrapolations for both the folded (dotted line) and the unfolded (dashed line) fast-frequency branches shown in Fig. 7.7c are plotted. It can be seen that the parameter τ falls, within error bars, between the relaxation time of the molecular system when Barnase is either folded or unfolded. These results allows us to easily explain why τ decreases with force: in Fig. 7.2a it can be seen that both k_{eff}^F and k_{eff}^U increase with force. Since τ is the relaxation time of the experimental setup, it satisfies $\tau = \gamma/k_{\text{eff}}$ (being γ the friction coefficient, and k_{eff} either k_{eff}^F or k_{eff}^U). Hence, τ is expected to decrease with force. Summing up, any potential observation of fast intermediate states along the protein folding is limited by the time resolution of the mini-tweezers instrument. The improvement of such resolution would not only grant access to the observation of fast intermediate states, but would make it possible to measure the transition path time of the protein [Chun 09, Yu 12, Neup 12, Chun 12].

Two decades ago the folding pathway of Barnase was studied in great detail using protein engineering, nuclear magnetic resonance, calorimetry, and urea denaturation experiments [Mato 90, Mato 92, Fers 93]. In these studies an observed non-linearity of the logarithm of unfolding and folding kinetic rates under the action of urea was attributed to the possible presence of a sub-millisecond intermediate. However, posterior hydrogen-exchange and denaturant-dependent unfolding experiments questioned the interpretation of former results [Take 00]. In fact, results obtained in this thesis using passive mode experiments and summarized in Figs. 7.6 and 7.7 point towards a two-state folding process at least in a time resolution of the order of few milliseconds. In order to confirm or discard the hypothesis of the existence of a sub-millisecond intermediate state in the mechanical folding of Barnase with DFS experiments, the experimental time resolution of the instrument must be improved or the relaxation time of the experimental setup must be reduced.

7.3.3 Folding kinetics properties

The protein folding process can be described as a first order Markov process. Therefore, the survival probability of the unfolded state, $P_U(t)$, satisfies the master equation:

$$\frac{dP_U(t)}{dt} = -k_{N\leftarrow U}(f)P_U(t), \tag{7.8}$$

where $k_{N\leftarrow U}(f)$ is the force-dependent folding kinetic rate. Equation (7.8) can be easily solved by considering that the force f, and consequently $k_{N\leftarrow U}(f)$, does not change with time:

$$P_U(t) = \exp\left[-k_{N\leftarrow U}(f)t\right], \tag{7.9}$$

where it is assumed that $P_U(t=0) = 1$. Equation (7.9) is a reasonable approximation to model the behavior of Barnase during the passive-mode time interval of the experimental protocol shown in Fig. 7.5. The survival probability $P_U(t)$ can be extracted from the passive-mode experiments carried out at a given preset force f_p according to:

$$P_U(t) = 1 - \frac{n(t_{\text{fold}} < t)}{N_{\text{total}}} \tag{7.10}$$

where N_{total} is the total number of measured trajectories at a given preset force f_p and $n(t_{\text{fold}} < t)$ is the number of events at which Barnase folds at times t_{fold} smaller than t. Fits of the experimentally measured survival probabilities $P_U(t)$ to Eq. (7.9) obtained at different preset forces are shown in Fig. 7.8. It can be seen that data is well-fitted by a single exponential.

Figure 7.9 shows the force-dependent folding kinetic rate estimated from the fits to Eq. (7.9) (Fig. 7.8). As already done to characterize the molecular FEL of Barnase upon unfolding (Section 7.3.1), the BE and the DHS models will be fitted to the experimental profile of $k_{N\leftarrow U}(f)$ to get estimations for the relative distance between state U and the TS, $x_{\text{TS}-U}$, the height of the kinetic barrier, $\Delta G_{U-\text{TS}}$, and the kinetic rate of folding at zero force, k_m^f (the super index f is here added in order to distinguish it from the

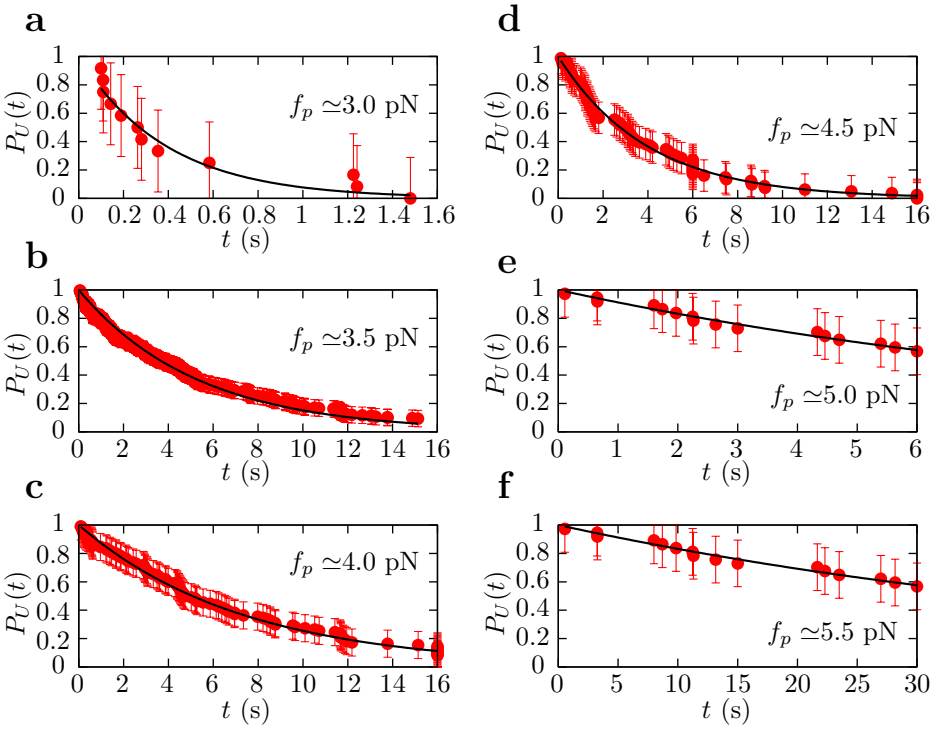

Figure 7.8: Survival probabilities of U at different present forces. Fit of the experimental survival probabilities extracted from passive traces using Eq. (7.10) to Eq. (7.9) at the preset force f_p of 3.0 pN (**a**), 3.5 pN (**b**), 4.0 pN (**c**), 4.5 pN (**d**), 5.0 pN (**e**), and 5.5 pN (**f**).

kinetic rate of unfolding at zero force, k_m). Results from such fits are shown in Fig. 7.9 and summarized in Table 7.4. Again, even though the different fits to the experimental data look identical (Fig. 7.9), there are discrepancies between the different numerical estimations.

Model	k_m^f (s^{-1})	x_{TS-U} (nm)	ΔG_{U-TS} (k_BT)
BE	40±20	5±1	–
DHS, $\gamma = 1/2$	2±1	2.8±0.6	5±1
DHS, $\gamma = 2/3$	7±3	3.6±0.8	5±1

Table 7.4: Parameters for the FEL estimated using the BE and DHS models by fitting $k_{N\leftarrow U}(f)$ versus f. Numerical values for k_m^f, x_{TS-U} and ΔG_{U-TS} obtained by fitting the force-dependent unfolding kinetic rate to the analytical expressions provided by the BE and the DHS model with $\gamma = 1/2$ or $\gamma = 2/3$ (Eqs. C.26 and C.36, respectively). Results are shown in Fig. 7.9.

For the BE model, $x_{TS-U} = 5$ nm, which at 4 pN implies that the difference in extension between the TS and state U is of ~ 48 aminoacids. In the case of the DHS, $x_{TS-U} = 3.2$ nm (obtained by averaging results using $\gamma = 1/2$ and $\gamma = 2/3$, Table 7.4), which corresponds to ~ 31 aminoacids at 4 pN. In both cases the result is compatible with the TS being related to the formation of part of the hydrophobic core of the

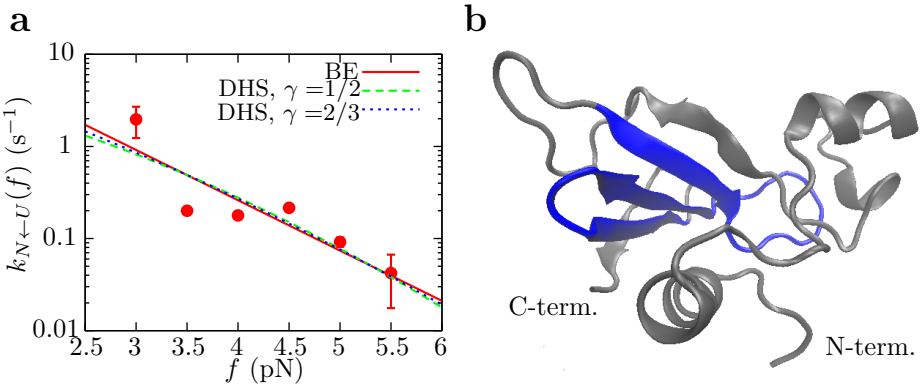

Figure 7.9: Mechanical folding and TS for Barnase. **a.** Folding kinetic rates versus force determined by the fit of Eq. (7.9) to the experimentally measured survival probabilities (Eq. 7.10), and fit to the analytical expressions provided by the BE and DHS models (Eqs. C.26 and C.36, respectively). **b.** Structure of Barnase in the native state, with the center of the hydrophobic core (made of 31 aminoacids) indicated in blue. The TS along folding is identified with the formation of the hydrophobic core.

protein, mainly dominated by β-sheets (blue region in Fig. 7.9b). Similar results have been derived using protein engineering [Mato 90].

7.3.4 Kinetic barrier and free energy of formation

Unfolding and folding kinetic rates have been determined in Sections 7.3.1 and 7.3.3, respectively. In Fig. 7.10a, the Chevron plot for Barnase as a function of the applied force is shown. According to Eqs. (4.33) and (4.34) the experimental profile of force-dependent kinetic barrier between states N and U for two-state systems, $B_{NU}(f)$, can be extracted from the logarithm of the kinetic rates of folding and unfolding (corrected with proper elastic contributions). In Chapters 4 and 5 it was shown that from the match between theoretical and experimental estimations of $B_{NU}(f)$ it is possible to extract the attempt frequency at zero force, k_0, and the molecular free energy of formation, ΔG^0_{NU}.

In the particular case of Barnase, and in general for most proteins, there is not a theoretical prediction for the profile of the kinetic barrier, $B_{NU}(f)$. Nevertheless, the profile of the kinetic barrier can be extrapolated from the logarithm of the experimental kinetic rates $\log k_{N\to U}(f)$ and $\log k_{N\leftarrow U}(f)$, using Eqs. (4.34a) and (4.34b) (red-solid squares and blue-open circles in Fig. 7.10, respectively). Imposing the continuity of both extrapolated curves, a rough estimation of $\Delta G^0_{NU} \sim 20 \pm 5\ k_{\rm B}T$ is obtained. Remarkably, using protein engineering the free energy of formation of Barnase has been previously estimated to be equal to $\Delta G^0_{NU} = 16\ k_{\rm B}T$. Moreover, in figure chapters it will be shown how the measure $\Delta G^0_{NU} = 14 \pm 7\ k_{\rm B}T$ for Barnase using fluctuation relations (Section 8.4.2). Therefore, all values are in reasonable agreement among themselves within error bars.

A lower bound estimate for k_0 can also be obtained by assuming that at large forces the kinetic barrier tends to zero. This gives that $\log k_0 \sim 5$ and, hence, $k_0 \sim$

150 s^{-1}. From the attempt frequency k_0 both the unfolding and the folding kinetic rates at zero force can be estimated by assuming $k_m = k_0 \exp(-\Delta G_{N-\mathrm{TS}}/k_\mathrm{B}T)$ and $k_m^f = k_0 \exp(-\Delta G_{U-\mathrm{TS}}/k_\mathrm{B}T)$. Noteworthy, taking the values derived using the DHS model in Sections 7.3.1 and 7.3.3 one gets $k_m \sim 150 \exp(-30) \sim 10^{-11}$ s^{-1} and $k_m^f \sim 150 \exp(-5) \sim 1$ s^{-1}. Therefore, the estimated value for k_0 obtained in this section and values for k_m, k_m^f, $\Delta G_{N-\mathrm{TS}}$ and $\Delta G_{U-\mathrm{TS}}$ obtained by fitting the force-dependent unfolding and folding kinetic rates to the DHS model are consistent among themselves.

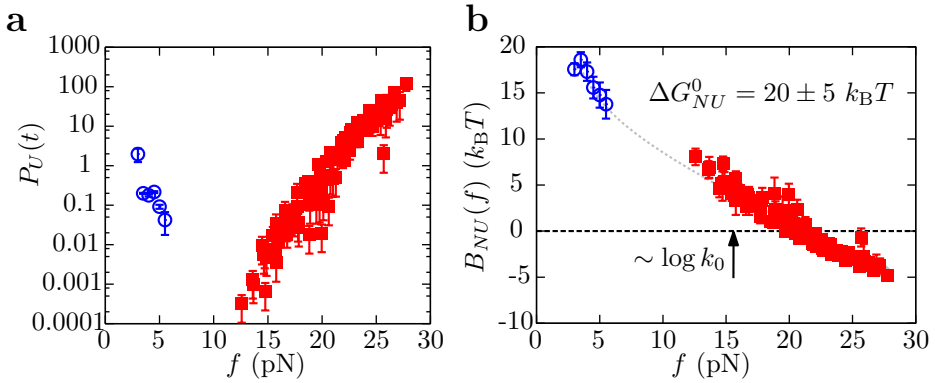

Figure 7.10: Chevron plot and reconstruction of the kinetic barrier. a. Unfolding (red-solid squares) and folding (blue-open circles) kinetic rates as a function of force. **b.** Profile of the kinetic barrier $B_{NU}(f)$ determined using Eq. (4.34a) (red-solid squares) and Eq. (4.34b) (blue-open squares). Imposing the continuity of the extrapolation (gray-dashed curve) and estimation for the free energy of formation of Barnase is obtained, $\Delta G^0_{NU} = 20 \pm 5$ $k_\mathrm{B}T$. Additionally, assuming that at large forces the kinetic barrier goes to zero gives a lower bound for $\log k_0 \sim 5$. Hence, $k_0 \sim 150$ s^{-1}. Error bars are computed using the bootstrap method and by propagation of the errors of the elastic parameters.

A decrease in the slope of $B_{NU}(f)$ is observed in Fig. 7.10b, in agreement with the Leffler-Hammond postulate and results previously obtained using the DHS in Sections 7.3.1 and 7.3.2: at low forces the TS is close to state U (since most of the protein is unfolded with the exception of part of the hydrophobic core, Fig. 7.9b), whereas at large forces the TS is close to state N (most of the protein is folded except stretched unstructured regions, Fig. 7.4b).

7.4 Conclusions

The behavior of Barnase when a mechanical force is applied to its ends has been deeply investigated in this chapter. Barnase is a globular protein that does not contain any disulfide bond. This makes it an ideal system to be studied using DFS with OT. Under pulling experiments, Barnase shows large hysteresis between unfolding and folding FDC: the protein unfolds with a sudden jump above ∼20 pN, and shows a smooth transition around ∼4 pN when it folds.

First, the elastic properties of the peptide chain have been extracted from the direct measurement of the stiffness of the molecular setup, as it was done for short ssDNA

molecules in Chapter 3. Due to the strong hysteresis effects, the stiffness of the setup when Barnase was either folded or unfolded could be accessed in a very large interval of forces, covering from 4 to 25 pN. Fits to different analytical expressions of the ideal WLC model give a persistence length value in the range 0.60-0.85 nm, and an inter-aminoacid distance of 0.36-0.34 nm/aa. These results are in very good agreement with previous estimations.

During the unfolding cycle of pulling experiments, Barnase showed a sudden force jump along unfolding that is compatible with the sudden release of 110 aminoacids. Hence, Barnase behaves as a two-state system upon mechanical unfolding similar to small DNA hairpins. The force-dependent unfolding kinetic rate could be easily obtained from the survival probability of state N along the unfolding process using Markov models (Eq. 4.23). Next, the standard BE and DHS models were used to fit the average unfolding force versus the pulling speed and the unfolding kinetic rate versus force to analytical expressions in order to get estimates for the distance between the TS and N, $x_{N-\mathrm{TS}}$, the height of the kinetic barrier, $\Delta G_{N-\mathrm{TS}}$, and the kinetic rate of unfolding at zero force, k_m. A released number of aminoacids can be associated to the estimated distance $x_{N-\mathrm{TS}}$, and a corresponding structure of the TS along the unfolding pathway has been hypothesized based also on previous measurements carried out combining techniques such as protein engineering with urea denaturation. This structure is very close to state N of the protein, where only the unstructured regions and part of the α-helices are unfolded. Moreover, the estimated value for $\Delta G_{N-\mathrm{TS}}$ is in remarkable agreement with previously determined values using bulk techniques.

The study of the folding transition of Barnase is a bit more difficult, since no clear force jump is observed during the releasing process in pulling experiments. Nevertheless, under equilibrium experiments a force jump compatible with the absorption of 110 aminoacids (which is the number of aminoacids of Barnase) is observed when the protein folds, and no evidence of intermediate states is found at standard temporal resolution (4 kHz). In order to discard the existence of intermediate states above the millisecond temporal scale, equilibrium experiments at a high sampling rate (50 kHz) were also performed and the analysis of the relaxation time of the system upon folding was directly measured. Again, no evidence of intermediate states was found in the timescale of the order of milliseconds. Once the folding transition was assumed to be a two-state process, passive traces recorded at different forces were analyzed accordingly in order to extract the force-dependent folding kinetic rates. As before, the BE and DHS models were used to fit the folding kinetic rate versus force to analytical expressions in order to get estimates for the distance between the TS and U, $x_{\mathrm{TS}-\mathrm{U}}$, the height of the kinetic barrier, $\Delta G_{U-\mathrm{TS}}$, and the kinetic rate of folding, k_m^f. The resulting value for $x_{\mathrm{TS}-\mathrm{U}}$ was associated to a number of absorbed aminoacids and a possible structure for the TS upon folding was hypothesized, where the main hydrophobic core of the protein (mostly made of β-sheets) is formed. In this case, the TS is found close to state U.

Finally, the study of the profile of the force-dependent kinetic barrier $B_{NU}(f)$ was performed using experimentally measured unfolding and folding kinetic rates. A rough estimate for the free energy of formation of Barnase could be extracted by imposing

the continuity of the experimentally determined kinetic barrier using the unfolding kinetic rate (Eq. 4.34a) and the folding kinetic rate (Eq. 4.34b). The result, $\Delta G^0_{NU} = 20 \pm 5\ k_{\mathrm{B}}T$, is in reasonable agreement with the value previously estimated using a protein engineering-based approach ($\Delta G^0_{NU} = 16\ k_{\mathrm{B}}T$), and with values obtained using fluctuation relations (Section 8.4.2). A lower bound for the attempt frequency at zero force can also be estimated by assuming that at large forces the kinetic barrier disappears. Hence, $k_0^{\mathrm{lower}} = 150\ \mathrm{s}^{-1}$. This results is consistent with the estimations of k_m, k_m^f, $\Delta G_{N-\mathrm{TS}}$ and $\Delta G_{U-\mathrm{TS}}$ determined using the DHS model. Additionally, the profile of $B_{NU}(f)$ is observed to satisfy the Leffler-Hammond postulate, since its slope decreases as force increases. This implies that the TS is located close to state U at low values of the forces and approaches state N as force increases. This is in agreement with the structures proposed for the TS mediating mechanical unfolding and folding.

Future work should focus on the direct measurement of the transition path time, which requires an improvement of the instrumental time resolution. This would unravel the existence of fast intermediates on pathway along folding and unfolding transitions. A first approach could be to decrease the relaxation time of the different elements of the experimental setup. For instance, performing experiments with a stiffer optical trap, stiffer molecular handles, or smaller polystyrene beads would decrease the relaxation time of the beads and the handles, and hence the total relaxation time of the system would be determined by the folding time of Barnase.

Part III

Fluctuation relations

Chapter 8

Fluctuation relations

The measurement of free-energy differences ΔG in classical thermodynamics is attained by quasi-statically changing the control parameter of a given experimental system and measuring the net amount of energy exchanged between the system and its environment during the process. When the experiment is carried out under non-equilibrium conditions the average work $\langle W \rangle$ evaluated over different realizations of the same experimental protocol is always larger than the free energy difference, *i. e.* $\langle W \rangle \geq \Delta G$, as stated by the second law of thermodynamics [Zema 97, Liph 12].

At macroscopic scales the experimental output of thermodynamic manipulation does not significantly change over different repetitions, even when experiments are carried out under non-equilibrium conditions. Samples contain a large number N of molecules and fluctuations, which are of the order of $1/\sqrt{N}$, are negligible.

Advances in nanomanipulation carried out during the last 20 years grant access to events occurring at the single molecule level ($N \sim 1$). In this microscopic scale, work measurements are of the same order of magnitude than thermal fluctuations, $\sim k_{\mathrm{B}}T$. These systems are typically known as *small systems*. Here fluctuations are relevant and different repetitions of the same experimental protocol give different outcomes [Rito 06a]. The second law of thermodynamics holds on average, but transient violations can be observed. The need to physically characterize such systems implied a boost in non-equilibrium statistical theories, which favored the development of stochastic thermodynamics and the appearance of fluctuation relations [Jarz 97, Croo 99, Croo 00].

This chapter is focused on work relations, which are those fluctuation theorems that relate non-equilibrium work measurements with free-energy differences. An experimental verification of these relations is provided using results from pulling experiments performed with OT on a DNA hairpin. These results allow us to further measure the molecular free energy of formation at zero force. It will be shown that results depend on the number of measurements performed. In case of large hysteresis effects, as is the case of protein Barnase, fluctuation relations will be combined with results from the Random Energy model in order to obtain accurate estimations of the free-energy difference.

8.1 Crooks and Jarzynski work relations

Fluctuation relations establish fundamental equalities between the work performed in a thermodynamic system and the free-energy difference between the initial and the final state [Alem 11, Jarz 11].

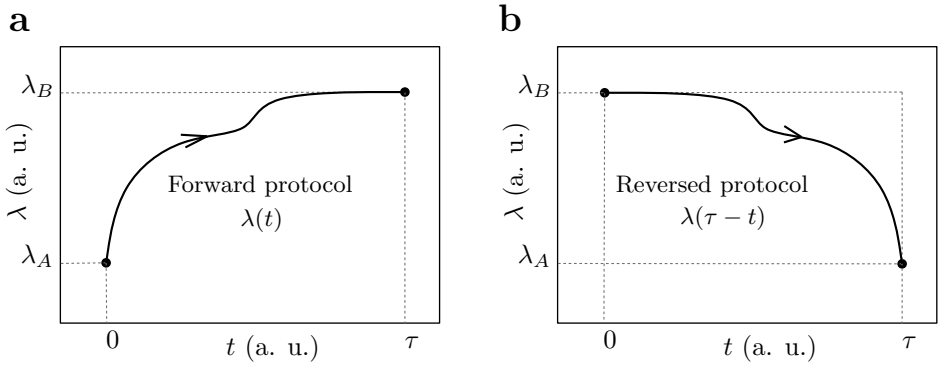

Figure 8.1: Forward and reversed protocols. Schematic representation of a forward protocol (**a**) and its reversed (**b**) in arbitrary units of time: the control parameter λ is externally controlled between values λ_A and λ_B during a time interval τ. The reversed protocol is the time-reversed of the forward protocol.

Consider a system initially in thermal equilibrium at a given state A with a value for the control parameter $\lambda(0) = \lambda_A$. An experimental protocol $\lambda(t)$ is applied on the system during a time interval τ by manipulating the control parameter λ. At the end of the protocol, the system is at state B with a value of the control parameter $\lambda(\tau) = \lambda_B$ (Fig. 8.1a). The work measured along one realization of the experiment is:

$$W = \int_0^\tau dt \frac{d\lambda}{dt} \frac{\partial \mathcal{H}}{\partial \lambda}, \tag{8.1}$$

where $\mathcal{H}$ is the Hamiltonian of the system. When dealing with small systems, the work W measured along the experimental protocol can be significantly different upon different independent realizations of an identical protocol because we cannot access the microscopic configurations explored by the system and consequently the term $\partial\mathcal{H}/\partial\lambda$ in Eq. (8.1) is different for each trajectory.

Now suppose that the time-reversed experimental protocol is performed: the system is in equilibrium in state B at λ_B and the control parameter varies according to $\lambda(\tau - t)$, until reaching state A at λ_A (Fig. 8.1b). An important work relation is the Crooks fluctuation relation (CFR), which reads as [Croo 00]:

$$\frac{P_F(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta G_{AB}}{k_\mathrm{B}T}\right). \tag{8.2}$$

$P_F(W)$ is the probability density function of the work performed along a forward protocol, $P_R(-W)$ is the probability density function of the work (with opposite sign)

performed along the reversed protocol, and $\Delta G_{AB} = G_B(\lambda_B) - G_A(\lambda_A)$ is equal to the Gibbs free-energy difference between the equilibrium states B and A at λ_B and λ_A respectively. A demonstration of equality provided in Eq. (8.2) can be found in Appendix I.1.

The Jarzynski work relation [Jarz 97] can be obtained by multiplying Eq. (8.2) by $P_R(-W)$ and integrating over W. This gives:

$$\left\langle \exp\left(-\frac{W}{k_{\mathrm{B}}T}\right)\right\rangle_F = \exp\left(-\frac{\Delta G_{AB}}{k_{\mathrm{B}}T}\right), \tag{8.3}$$

where $\langle \ldots \rangle_F$ denotes the average over forward trajectories. Even though the Jarzynski relation is a corollary of the CFR, it was derived earlier. Its demonstration is provided in Appendix I.2.

A beautiful consequence of Eq. (8.2) is that the value of the work that is equally probable in the forward and the reversed experimental protocol (*i. e*, $P_F(W) = P_R(-W)$) is equal to the free-energy difference ΔG_{AB}, and it is always the same no matter how far from equilibrium the system is driven during the experimental process. Another interesting consequence of fluctuation relations is that they predict the existence of trajectories where $W < \Delta G_{AB}$. Such trajectories are commonly known as "transient violations of the second law". However, it must be stressed that there is no such violation because thermodynamics do not focus on outcomes from single trajectories but on the resulting average over an infinite number of them. In fact, the work average $\langle W \rangle$ evaluated over multiple independent realizations of the experiment is always larger, or equal if the protocol is applied slow enough, to the free-energy difference, $\langle W \rangle \geq \Delta G_{AB}$.

The recovery of free-energy differences from irreversible work measurements is possible by applying the Jarzynski equality (Eq. 8.3) to unidirectional work measurements or applying the CFR (Eq. 8.2) to bidirectional work measurements (when both the forward and the reversed protocols are feasible). Typically, the combination of information from the forward and reversed protocols provides less biased free-energy estimates. However, when dissipation and hysteresis effects between the forward and the reversed processes are high, the work distributions in Eq. (8.2) separate from each other and a large error is introduced in the free-energy estimate. A theory of bias is then required to improve results [Pala 11]. It will be briefly discussed in Section 8.4.2.

Applications of fluctuation theorems include the measurement of the free energy of formation of RNA and DNA hairpins [Coll 05]; the determination of the stability of native domains in proteins [Shan 10]; the measurement of mechanical torque in rotary motors [Haya 10]; the conversion of information into work in systems under feedback control [Toya 10]; the recovery of free energy landscapes from unidirectional work measurements [Humm 01b, Gupt 11]; the reconstruction of the free-energy branches for the different molecular states of a system as a function of the control parameter [Juni 09, Alem 12]; the determination of the free energies of formation of kinetic states [Alem 12]; and even the measurement of binding free-energies and equilibrium constants in chemical reactions [Koir 11, Yang 12, Camu 14].

8.2 Experimental validation of the Crooks equality

Fluctuation relations have been experimentally tested in different types of small systems, such as one colloidal particle captured in an optical trap that is translated relative to the surrounding water [Wang 02], a periodically excited single defect center in a diamond [Schu 05], or an oscillator in contact with a thermostat under the action of an external force [Doua 06, Joub 07].

A very important verification of the CFR for the course of this thesis was provided in 2005 by Delphine Collin and collaborators [Coll 05]. In their study, two different RNA molecules were pulled using optical tweezers and the free energy of formation of the different structures was determined using Eq. (8.2). This work was fundamental to establish the basis of how to determine the free energies of formation of molecules from irreversible work measurements [Cecc 05, Moss 09c, Moss 09a, Shan 10].

In this section, the CFR will be experimentally tested again. To do so, pulling experiments using the CD4 hairpin (Fig. 4.2a) will be used. As already explained in Chapter 4, the molecule is tethered between two 29-base pairs long dsDNA handles that act as rigid spacers. Using biochemical assays, each end of the whole molecular construct is attached to a micron-sized polystyrene bead. Finally, one bead is immobilized in the tip of a micropipette and the other is captured in the optical trap. In the mini-tweezers setup the control parameter is the distance between the center of the optical trap and the tip of the micropipette, which is denoted by λ.

Pulling experiments consist of unfolding and folding processes (Section 2.3.2). Hereafter, the unfolding process will be identified with the forward protocol, whereas the folding process will be identified with the reversed one.

In the unfolding process, the trap-pipette distance λ is initially set to λ_0, where the molecule is in equilibrium at its folded/native conformation N. Next, λ is increased at a constant pulling speed v during a time interval τ ($d\lambda/dt = \dot{\lambda} = v$). During this period, the mechanical force applied to the DNA hairpin also increases. For a given stochastic value of the force, the hairpin can no longer withstand the force and it unfolds. This is observed as a drop in force that corresponds to the relaxation of the bead into the center of the optical trap due to the release of ssDNA that is associated with the unfolding process (Chapter 3). Interestingly, for CD4 refolding events are occasionally observed along the unfolding trajectory (Fig. 8.2a). These will be referred to as hopping events. Regardless of the molecular state of the hairpin (folded in the native conformation N or unfolded in the stretched conformation U), $d\lambda/dt$ equals v until the value λ_1 is reached at $t = \tau$, $\lambda(\tau) = \lambda_1$, where the protocol stops and the molecule equilibrates at U.

According to Eq. (8.1), the work measured along an unfolding trajectory is:

$$W = \int_0^\tau dt\ vf = \int_{\lambda_0}^{\lambda_1} d\lambda\ f, \tag{8.4}$$

which corresponds to the area below the FDC (Fig. 8.2a). Since the force at which the hairpin unfolds and the number of hopping events changes in each realization of an unfolding protocol, the value of W is different in each trajectory.

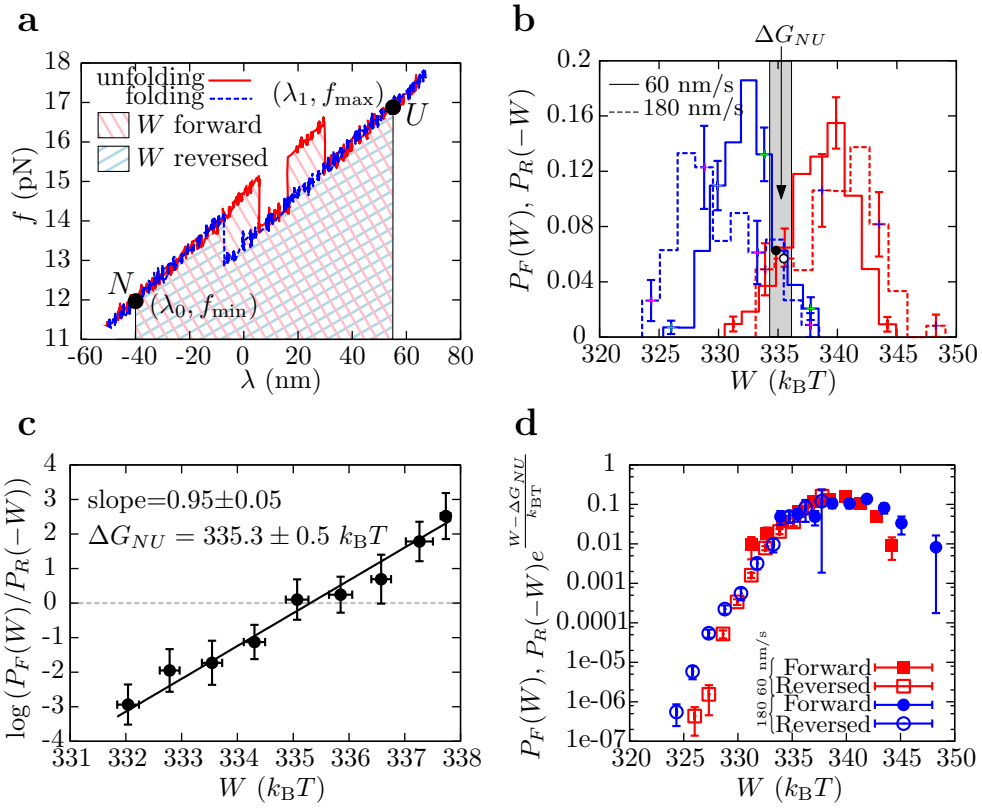

Figure 8.2: Experimental validation of the Crooks fluctuation relation. **a.** Example of an unfolding (red) and folding (blue) trajectory and measurement of the work value as the area below the FDC (dashed areas). **b.** Experimental work distributions, $P_F(W)$ (red) and $P_R(-W)$ (blue), measured at two different pulling speeds. The value of the work at which the forward and reversed work distributions cross is equal to the free-energy difference ΔG_{NU} (gray square). **c.** Logarithm of the ratio between the probabilities $P_F(W)$ and $P_R(-W)$ versus W and fit to a straight line. The theoretical slope is 1 (Eq. 8.6), and the value of the work at which $\log(P_F(W)/P_R(-W)) = 0$ equals ΔG_{NU}. **d.** Matching between the experimental measurement of $P_F(W)$ (solid squares and solid circles) and of $P_R(-W)\exp(W - \Delta G_{NU}/k_BT)$ (empty squares and circles), obtained by pulling CD4 at 60 and 180 nm/s (red and blue points, respectively).

In the refolding process the time-reversed protocol $\lambda(\tau - t)$ is applied. Therefore, the trap-pipette distance λ is initially set to λ_1, where the molecule is equilibrated in state U. Next, λ is decreased at the constant pulling speed $-v$ ($d\lambda/dt = \dot{\lambda} = -v$) during the time interval τ until it reaches the value λ_0, where the protocol ends and the molecule equilibrates in the N state. Along the folding process, the force applied to the DNA hairpin decreases. When it reaches a sufficiently low value, the molecule folds and a jump in force is observed. Hopping events can be observed for a range of forces. The work in a given folding trajectory is measured as:

$$W = \int_0^{\tau} dt\ (-v) f = -\int_{\lambda_0}^{\lambda_1} d\lambda\ f, \tag{8.5}$$

which is equal to the area below the FDC with a negative sign (Fig. 8.2a). Again,

the force at which the molecule folds varies in each independent repetition of the folding process, and so does the number of hopping events. Therefore, the value of W is different for each trajectory.

Figure 8.2b shows the experimental $P_F(W)$ and $P_R(-W)$ measured by pulling the hairpin CD4 at two different pulling speeds. It can be observed that, even though hysteresis effects (and therefore dissipation) increase with the pulling speed, the work value at which the two distributions cross each other does not depend on v. According to the CFR, such value is equal to $\Delta G_{NU} = G_U(\lambda_1) - G_N(\lambda_0)$, since state A in Eq. (8.2) corresponds to state N at λ_0, whereas state B corresponds to U at λ_1. The measurement of the crossing point of work distributions obtained at 60 and 180 nm/s gives $\Delta G_{NU} = 335 \pm 1$ and 336 ± 1 $k_\mathrm{B}T$, respectively (Fig. 8.2b).

A validation of the CFR is shown in Fig. 8.2c, where the logarithm of the ratio between the probabilities $P_F(W)$ and $P_R(-W)$ versus W is represented [Moss 09b]. Results from the linear fit to the experimental data gives a slope equal to 0.95 ± 0.05, which is in excellent agreement with the theoretical prediction provided by Eq. (8.2),

$$\log\left(\frac{P_F(W)}{P_R(-W)}\right) = \frac{W}{k_\mathrm{B}T} - \frac{\Delta G_{NU}}{k_\mathrm{B}T}, \tag{8.6}$$

that implies that the slope equals 1. In addition, from this linear fit ΔG_{NU} can also be measured as the value of the work at which $\log\left(P_F(W)/P_R(-W)\right) = 0$, which is essentially the equivalent as determining the work value where $P_F(W) = P_R(-W)$. In this case, $\Delta G_{NU} = 335.3 \pm 0.5$ $k_\mathrm{B}T$. This verification is only carried out for data measured at the lowest pulling speed (60 nm/s), because the range of work values at which the quantity $\log\left(P_F(W)/P_R(-W)\right)$ can be measured is broader.

Yet another verification can be obtained by rewriting the CFR as:

$$P_F(W) = P_R(-W)\exp\left(\frac{W - \Delta G_{NU}}{k_\mathrm{B}T}\right). \tag{8.7}$$

Here it is stated that by multiplying the reversed work distribution $P_R(-W)$ by the term $\exp((W - \Delta G_{NU})/k_\mathrm{B}T)$ one should get the forward work distribution $P_F(W)$ [Pala 11, Ribe 14]. This is shown in Fig. 8.2d for the work distributions measured by pulling CD4 at 60 and 180 nm/s. The respective values of ΔG_{NU} that provide a good match between $P_F(W)$ and $P_R(-W)\exp(W - \Delta G_{NU}/k_\mathrm{B}T)$ are 335.7 ± 0.5 and 335.1 ± 0.5 $k_\mathrm{B}T$. Both values are in good agreement with the two previous estimators.

8.3 Free-energy estimators

In the previous section two different free-energy estimators were used: (1) the crossing point between the forward and reversed work histograms, which is also determined by finding the value of the work at which $\log\left(P_F(W)/P_R(-W)\right) = 0$, and (2) the value of ΔG_{NB} at which Eq. (8.7) is satisfied (Fig. 8.2b-c, respectively).

In general, the estimation of ΔG_{AB} using fluctuation relations is strongly limited by the number of work measurements. If there is not enough data, the tail of the work

distributions may not be well sampled and a systematic error, defined as bias, in the estimation of the free-energy difference may be introduced. Several procedures have been presented in the literature to overcome this limitation.

To begin with, the different terms of the CFR (Eq. 8.2) are rearranged in order to get:

$$P_F(W)\exp\left(-\frac{W}{k_{\mathrm{B}}T}\right) = P_R(-W)\exp\left(-\frac{\Delta G_{AB}}{k_{\mathrm{B}}T}\right). \tag{8.8}$$

By multiplying both sides of Eq. (8.8) by a generic function $\phi(W)$ and integrating over the work one gets a particular corollary of the CFR:

$$\left\langle \exp\left(-\frac{W}{k_{\mathrm{B}}T}\right)\phi(W)\right\rangle_F = \langle\phi(-W)\rangle_R \exp\left(-\frac{\Delta G_{AB}}{k_{\mathrm{B}}T}\right), \tag{8.9}$$

where $\langle\cdots\rangle_{F(R)}$ denotes the average over all possible forward (reversed) trajectories. The choice of the function $\phi(W)$ determines new work relations and provides different free-energy estimators. For instance, by taking $\phi(W) = 1$ the Jarzynski equality is recovered (Eq. 8.3).

8.3.1 Bennett acceptance ratio method

The Bennett acceptance ratio method [Benn 76] is the standard procedure used in this thesis to determine free-energy differences using the CFR. In the Bennett acceptance ratio method the following function is proposed:

$$\phi(W) = \frac{1}{1+\frac{n_F}{n_R}\exp\left(\frac{W-\Delta G_{AB}}{k_{\mathrm{B}}T}\right)}, \tag{8.10}$$

where n_F and n_R stand for the number of forward and reversed work measurements. The use of this function into Eq. (8.9) is known to minimize the statistical variance of the estimation of ΔG_{AB} [Benn 76, Shir 03]. Therefore, given a set of $n_F(n_R)$ forward(reversed) work measurements $\{W_i\}_{i=1,\ldots,n_F}$ ($\{W_i\}_{i=1,\ldots,n_R}$), the free-energy difference ΔG_{AB} can be estimated by solving the following transcendental equation for u:

$$\frac{u}{k_{\mathrm{B}}T} = -z_F(u) + z_R(u), \tag{8.11}$$

where

$$z_R(u) = \log\left[\frac{1}{n_R}\sum_{i=1}^{n_R}\left(\frac{1}{1+\frac{n_F}{n_R}e^{-\beta(W_i+u)}}\right)\right] \tag{8.12a}$$

$$z_F(u) = \log\left[\frac{1}{n_F}\sum_{i=1}^{n_F}\left(\frac{e^{-\beta W_i}}{1+\frac{n_F}{n_R}e^{\beta(W_i-u)}}\right)\right]. \tag{8.12b}$$

The value of u at which Eq. (8.11) is satisfied equals ΔG_{AB}. The right-hand side of Eq. (8.11) is expected to be constant near the solution of the transcendental equation [Moss 10].

In Fig. 8.3 it is plotted the result of the Bennett acceptance ratio method used to determine ΔG_{NU} in pulling experiments performed at two different pulling speeds for hairpin CD4. Red-squares and blue-circles are the values of the function $-z_F(u)+z_R(u)$ determined using non-equilibrium work measurements obtained at 60 and 180 nm/s, respectively. As expected, the function is almost constant for values of u around ΔG_{NU}. The crossing point between $-z_F(u) + z_R(u)$ and the dashed-black line (equal to u) is equal to ΔG_{NU}. Therefore, $\Delta G_{NU} = 335.1 \pm 0.4\ k_{\rm B}T$ as seen in Fig. 8.3, which is in excellent agreement with the different estimations carried out in the previous section.

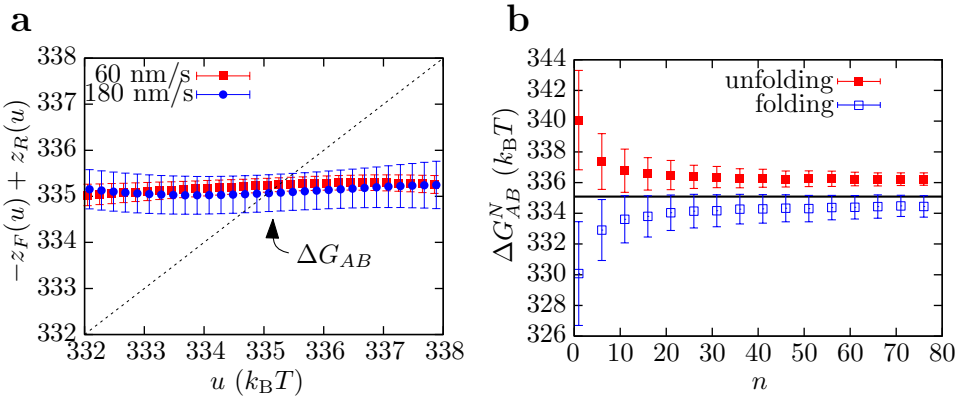

Figure 8.3: Free-energy estimators. a. Result of Bennett acceptance ratio method used in pulling experiments performed at two different pulling speeds. **b.** Convergence of the Jarzynski estimator with the number of unidirectional work measurements obtained from pulling experiments at 180 nm/s. Solid squares use unfolding work values to determine ΔG_{AB}, whereas open squares have been obtained using folding work values.

8.3.2 Unidirectional free-energy estimators and bias

The Jarzynski equality is recovered when the function $\phi(W) = 1$ is introduced into Eq. (8.9). Along the same lines, from a collection $\{W_i\}_{i=1,\ldots,n}$ of n independent work measurements obtained under a given protocol, the following free-energy estimator is obtained:

$$\Delta G_{AB} = -k_{\rm B}T \log \left[\frac{1}{n} \sum_{i}^{n} \exp\left(-\frac{W_i}{k_{\rm B}T} \right) \right]. \tag{8.13}$$

The main difference between this free-energy estimator and the ones previously introduced, such as direct application of the Crooks equality or the Bennett acceptance ratio method (Eqs. 8.2 of 8.11, respectively), is that in this case only work measurements from a unique experimental protocol are required and the reversed protocol is not needed. For this reason, Eq. (8.13) is usually known as a unidirectional free-energy estimator, whereas previously introduced methods to determine ΔG_{AB} are referred to as bidirectional free-energy estimators.

In practice, unidirectional free-energy estimators have a slower convergence to the value of ΔG_{AB} than bidirectional work measurements. That is, a large number n of

trajectories is required for the estimated value of ΔG_{AB} to be trustworthy in unidirectional methods. This is because the exponential average that appears in Eq. (8.13) is dominated by the lowest work values. Therefore, it is crucial to get a good experimental sampling of those rare trajectories with $W_i < \Delta G_{AB}$. Otherwise, estimated values of ΔG_{AB} are larger than expected.

This effect is shown in Fig. 8.3b. The free energy ΔG_{AB} is obtained using Eq. (8.13) for different number of work measurements n using only data from unfolding (red-solid squares) or folding (blue-open squares) experiments. The black-straight line is the value of ΔG_{AB} recovered using the Bennett acceptance ratio method. It can be seen how the estimator of ΔG_{AB} using unidirectional work measurements gets closer to the one obtained from bidirectional estimators as the number of measurements increase, for both unfolding and folding data.

The difference between the correct value of ΔG_{AB} and the one estimated from n work measurements is known as bias, B. When hysteresis effects between the forward and the reversed protocols are too large and work distributions do not cross each other, it is important to get enough experimental data to correctly sample the shape of the tails of the work distributions. Otherwise, the estimation of ΔG_{AB} can have a huge bias. For such situations, a more complex theory is needed to reduce the uncertainty in the recovery of free-energy differences. As an example, in reference [Pala 11], Palassini and Ritort use the Random Energy model to predict the dependence with the number of work measurements of the bias of the free-energy estimator provided by the Jarzynski equality. An application of this method will be shown in Section 8.4.2.

8.4 Recovery of the free energy of formation of molecules

The value of ΔG_{NU} measured in pulling experiments using the CFR gives the difference between the free energy of the molecular setup stretched at λ_1, where the molecule is unfolded and subjected to the maximum force of the protocol $f_{\max}$, and the free energy of the molecular setup stretched at λ_0, where the minimum force $f_{\min}$ of the protocol is applied and the molecule is folded,

$$\Delta G_{NU} = G_U(\lambda_1) - G_N(\lambda_0). \tag{8.14}$$

ΔG_{NU} contains energetic contributions from the different components of the experimental setup, which are the molecule under study, the molecular handles and the bead captured in the optical trap [Coll 05, Moss 09b]:

$$\Delta G_{NU} = \Delta G^0_{NU} + \Delta W^{\rm st}_{NU} + \Delta W^{\rm handles}_{NU} + \Delta W^{\rm bead}_{NU}. \tag{8.15}$$

The term ΔG^0_{NU} is the free energy of formation of the molecule. The rest of the contributions, ΔW^i_{NU} ($i =$ st, handles, bead), are reversible work differences $W^i_U - W^i_N$ related to the different parts of the experimental setup.

$\Delta W^{\rm st}_{NU}$ is the reversible work needed to stretch the molecule under study,

$$\Delta W^{\rm st}_{NU} = \int_0^{x_U(f_{\max})} f_U(x')dx' - \int_0^{x_N(f_{\min})} f_N(x')dx', \tag{8.16}$$

being $f_U(x)$ and $f_N(x)$ the equilibrium force-extension curves (or equation of states) between the force and the end-to-end distance of molecule in state U or N, respectively. The first term in the right-hand size of Eq. (8.16) is the reversible work needed to stretch the unfolded molecule from its random coil state at zero force to its equilibrium end-to-end distance at force $f_{\max}$, $x_U(f_{\max})$. In the case of DNA hairpins and proteins, the equilibrium dependence between the end-to-end distance and the mechanical force, $x_U(f)$ and $f_U(x)$, is typically modeled using the WLC model with proper values of both the persistence and the contour length. The second term in Eq. (8.16) accounts for the reversible work needed to orientate the folded molecule at $f_{\min}$. For the evaluation of such contribution, the folded molecule is modeled as a dipole of extension d with a configurational energy equal to $-fd\cos\theta$, being θ the angle between the force axis and the direction of orientation of the dipole [Forn 11]. As a result, the equilibrium projection of the dipole along the force axis, $x_N(f)$, can be modeled using the FJC elastic model for a single monomer with identical Kuhn and contour lengths (Chapters 3 and 7).

The two terms $\Delta W_{NU}^{\text{handles}}$ and $\Delta W_{NU}^{\text{bead}}$ are the reversible work needed to stretch the handles and to displace the bead in the optical trap from $f_{\min}$ to $f_{\max}$:

$$\Delta W_{NU}^{\text{handles}} + \Delta W_{NU}^{\text{bead}} = \int_{x_{\text{handles}}(f_{\min})}^{x_{\text{handles}}(f_{\max})} f(x')dx' + \int_{x_{\text{bead}}(f_{\min})}^{x_{\text{bead}}(f_{\max})} f(x')dx'. \tag{8.17}$$

Both contributions can also be written as:

$$\Delta W_{NU}^{\text{handles}} + \Delta W_{NU}^{\text{bead}} = \int_{f_{\min}}^{f_{\max}} f' \left| \frac{\partial f}{\partial x_{\text{handles}}} \right|^{-1} df' + \int_{f_{\min}}^{f_{\max}} f' \left| \frac{\partial f}{\partial x_{\text{bead}}} \right|^{-1} df', \tag{8.18}$$

which, by identifying the term $\partial f/\partial x_i$ as the stiffness k_i of each element (i = handles, bead), becomes:

$$\Delta W_{NU}^{\text{handles}} + \Delta W_{NU}^{\text{bead}} = \int_{f_{\min}}^{f_{\max}} f' \left(\frac{1}{k_{\text{handles}}} + \frac{1}{k_{\text{bead}}} \right) df' \tag{8.19a}$$

$$\simeq \int_{f_{\min}}^{f_{\max}} f' \left(\frac{1}{k_{\text{eff}}^F} \right) df' \tag{8.19b}$$

$$= \frac{f_{\max}^2 - f_{\min}^2}{2k_{\text{eff}}^F}. \tag{8.19c}$$

In order to obtain Eq. (8.19b) from Eq. (8.19a), it was used the general expression provided by Eq. (3.2b) for the total stiffness of the molecular setup for the folded and the unfolded branches k_{eff}^F and k_{eff}^U. It is here assumed that the stiffness of the molecule when it is in the folded state is very large and consequently it does not contribute to the stiffness of the folded branch, $k_{\text{eff}}^F \simeq \left(k_{\text{handles}}^{-1} + k_{\text{bead}}^{-1}\right)^{-1}$. Finally, Eq. (8.19c) is obtained by assuming that k_{eff}^F does not depend on force, which is a reasonable approximation for short ranges of integration.

To conclude, in order to recover from pulling experiments the free energy of formation of a given molecule at zero force with respect to the random coil state, ΔG_{NU}^0, the

contributions $\Delta W_{NU}^{\text{handles}}$, $\Delta W_{NU}^{\text{bead}}$ and $\Delta W_{NU}^{\text{st}}$ need to be subtracted from the free-energy difference ΔG_{NU} obtained using fluctuation relations [Coll 05, Moss 09b]. Examples of this method applied to a two-state DNA hairpin and a protein are provided in what follows.

8.4.1 Nucleic acid hairpins

Pulling experiments were performed on CD4, which is the two-state DNA hairpin shown in Fig. 4.2a. In Fig. 8.2a an unfolding and a folding FDC are plotted, and the measurement of the work as the area below the curve is represented. In this example, the interval of integration is given by $\lambda_0 = -40$ nm and $\lambda_1 = 55$ nm, which corresponds to $f_{\text{min}} = 12$ pN and $f_{\text{max}} = 17$ pN, respectively.

The value of ΔG_{NU} for these experiments was already obtained in Sections 8.2 and 8.3, and in average it was found to be equal to $335 \pm 1\ k_{\text{B}}T$. In order to extract the free energy of formation of CD4 it is required to estimate the contributions coming from the different parts of the experimental setup (Eq. 8.15).

First, the value of the reversible work needed to stretch the released ssDNA from 0 to the maximum force of the protocol, $f_{\text{max}} = 17$ pN is calculated. To do so, the WLC model is used with a persistence length P equal to 1.35±0.05 nm and an interphosphate distance 0.59 ± 0.02 nm/base, which were shown in Chapter 3 to properly describe the elastic properties of short ssDNA molecules. One gets:

$$W_U^{\text{st}} = \int_0^{x_U(f_{\text{max}})} f_U(x')dx' = 18.8 \pm 0.2\ k_{\text{B}}T, \tag{8.20}$$

where the error bar is obtained from the propagation of the error of the elastic parameters.

Second, the reversible work needed to orientate the hairpin stem at a value of the force equal to $f_{\text{min}} = 12$ pN (Fig. 8.2a) is estimated. As explained before, the hairpin stem is modeled as a monomer of length $d = 2$ nm that is oriented along the force axis [Forn 11, Wood 06b]. The energetic contribution of this element is calculated using the FJC elastic model with identical Kuhn and contour lengths d. Therefore:

$$W_N^{\text{st}} = \int_0^{x_N(f_{\text{min}})} f_N(x')dx' = 1.8 \pm 0.1\ k_{\text{B}}T. \tag{8.21}$$

As before, the error bar is obtained from error propagation. Thus, the value of $\Delta W_{NU}^{\text{st}} = 17.0 \pm 0.2\ k_{\text{B}}T$ is obtained by subtracting W_U^{st} and W_N^{st} (Table 8.1).

In Fig. 8.2a it can be seen that the force-dependence of the folded branch with force is practically linear, with a slope $k_{\text{eff}}^F = 0.065$ pN/nm. In this case, the energetic contribution of the handles and the bead captured in the optical trap can be calculated using Eq. (8.19c):

$$\Delta W_{NU}^{\text{handles}} + \Delta W_{NU}^{\text{bead}} = \frac{f_{\text{max}}^2 - f_{\text{min}}^2}{2k_{\text{eff}}^F} = 268 \pm 4\ k_{\text{B}}T \tag{8.22}$$

Finally, the free energy of formation of the hairpin obtained using the fluctuation theorem and subtracting the contributions coming from different parts of the experimental setup is equal to $50 \pm 4\ k_{\mathrm{B}}T$ (Table 8.1). This number is in very good agreement with theoretical predictions obtained using the NN model and the UO set of base-pair free energies ($\Delta G^0_{NU} = 51.9\ k_{\mathrm{B}}T$) [Sant 98], or the ones derived from UNZ experiments ($\Delta G^0_{NU} = 48.0\ k_{\mathrm{B}}T$).

$\Delta G_{NU}\ (k_{\mathrm{B}}T)$	$\Delta W^{\mathrm{st}}_{NU}\ (k_{\mathrm{B}}T)$	$\Delta W^{\mathrm{handles}}_{NU} + \Delta W^{\mathrm{bead}}_{NU}\ (k_{\mathrm{B}}T)$	$\Delta G^0_{NU}\ (k_{\mathrm{B}}T)$
335±1	17.0±0.2	268±4	50±4

Table 8.1: Fluctuation theorem applied on a two-state DNA hairpin (CD4). The free energy ΔG_{NU} is obtained from the application of the CFR on pulling experiments carried out with CD4 (Fig. 8.2). Once the different energetic contributions from the experimental setup are removed (Eq. 8.15), the free energy of formation of CD4 is obtained. The result is in very good agreement with theoretical predictions.

8.4.2 Proteins with large hysteresis

Pulling experiments performed on Barnase, which is a small protein made of 110 amino-acids, were already introduced in Chapter 7. It was shown that the determination of the free energy of formation of this protein is a challenge using DFS methods because the folding transition towards the native conformation is not always clearly observed. The situation is also complex when trying to determine its free energy of formation using fluctuation theorems due to the strong hysteresis between unfolding and folding processes.

In what follows, ΔG^0_{NU} for Barnase is determined by directly applying the fluctuation relations and by using an extended theory that takes into account the effect of the bias. Limitations present in both approaches are discussed.

Determination of the free-energy difference ΔG_{NU} using the Crooks relation

As for the two-state DNA hairpin, unfolding and folding processes in pulling experiments are identified with the forward and the reversed protocols, respectively. The work performed in each protocol is calculated as the area below the experimental FDC, as represented in Fig. 8.4a. Extreme values for the control parameter in this case are chosen at $\lambda_0 = -200$ and $\lambda_1 = 200$ nm, where $f_{\min} = 1.0 \pm 0.2$ pN and $f_{\max} = 23.2 \pm 0.1$ pN, respectively. The resulting forward and reversed work histograms are plotted in Fig. 8.4b, and in contrast to the CD4 hairpin, here they do not cross each other. In the example depicted, the average work over forward trajectories is $\langle W \rangle_F = 1117 \pm 1\ k_{\mathrm{B}}T$, whereas the average work over reversed trajectories is $\langle W \rangle_R = 1024.7 \pm 0.6\ k_{\mathrm{B}}T$. Therefore, $\langle W \rangle_F - \langle W \rangle_R \simeq 90\ k_{\mathrm{B}}T$. Such a large value for hysteresis hinders the determination of the free-energy difference using fluctuation theorems for three main reasons:

- First, when the work distributions do not cross each other it is not possible to experimentally measure the quantity $\log\left(P_F(W)/P_R(-W)\right)$ for any work value

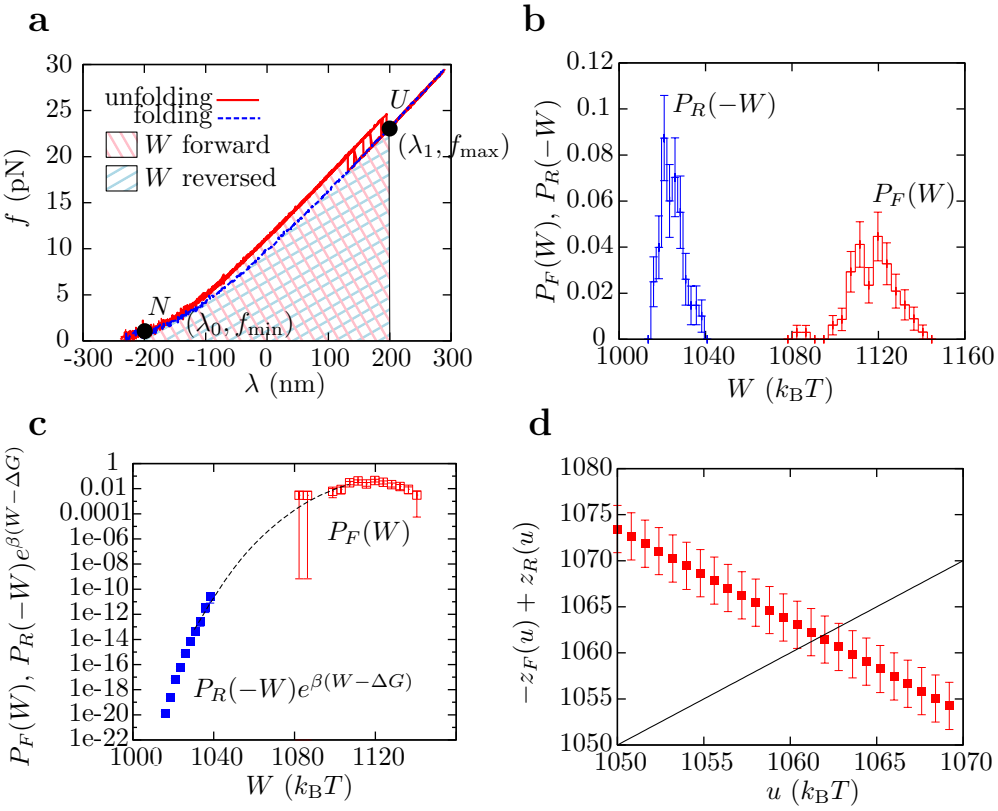

Figure 8.4: Fluctuation theorems applied to pulling experiments with Barnase. a. Example of experimental unfolding and folding FDC and measurement of the work as the area below the FDC between $\lambda_0 = -200$ nm and $\lambda = 200$ nm. Unfolding (folding) processes are identified with forward (reversed) protocols. **b.** Forward and reversed work distributions measured in pulling experiments with Barnase. It can be seen that the distributions are wide apart and do not cross each other. **c.** Match of the forward work distribution, $P_F(W)$ (red-open squares), and the reversed work distribution multiplied by the exponential factor, $P_R(-W)\exp(W - \Delta G_{NU})/k_BT$ (blue-solid squares). Tails are assumed to be a smooth extension of the measured work distributions (dashed line). **d.** Results of the Bennett acceptance ratio method (Eq. 8.11). Error bars are obtined using the Bootstrap method.

and ΔG_{NU} cannot be determined using the linear fit to Eq. (8.6).

- Second, the match between the experimentally measured $P_F(W)$ and $P_R(-W)\exp(W - \Delta G_{NU})/k_BT$ is highly inaccurate. In the example provided in Fig. 8.4c, the shape of the tails of the two work distributions are inferred by eye in a range of work values from 1040 k_BT to 1080 k_BT in order to get a value for ΔG_{NU}. By extrapolating the tails of the measured work distributions (dashed line in Fig. 8.4c),the value of $\Delta G_{NU} \simeq 1058 \pm 5\ k_BT$ is inferred.

- Third, the Bennett acceptance ratio method is not reliable as the profile of $-z_F(u) + z_R(u)$ is not flat around the solution of the transcendental equation given by Eq. (8.11) (Fig. 8.4d). In any case, the value of ΔG_{NU} determined with this

approach is $1062 \pm 5\ k_{\mathrm{B}}T$, which is not far from the previous estimate obtained by matching forward and reversed work distributions.

Hence, the value of ΔG_{NU} cannot reliably be determined using standard application of the CFR (Eq. 8.2) neither the Bennett acceptance ratio method (Eq. 8.11), even though the two approaches provide compatible numerical values.

Alternatively, a free-energy measurement can be obtained by expanding the Jarzynski estimator by the cumulants ω_i [Jarz 97, Humm 01a, Coll 05, Dell 13], that is:

$$\Delta G_{NU} = -k_{\mathrm{B}}T \log \left\langle \exp - \left(\frac{W}{k_{\mathrm{B}}T} \right) \right\rangle = \sum_{i=1}^{\infty} \left(-\frac{1}{k_{\mathrm{B}}T} \right)^{i-1} \frac{\omega_i}{i!}. \tag{8.23}$$

For Gaussian work distributions only the first two terms in Eq. (8.23) survive and a new unidirectional estimator for the free-energy differences is obtained:

$$\Delta G_{NU} = \langle W \rangle - \frac{\sigma_W^2}{2k_{\mathrm{B}}T}, \tag{8.24}$$

where $\sigma_W^2 = \langle W^2 \rangle - \langle W \rangle^2$ is the variance of the work values. For unfolding (forward) measurements $\langle W \rangle = 1117 \pm 1\ k_{\mathrm{B}}T$ and $\sigma_W^2 = 122 \pm 23\ k_{\mathrm{B}}T^2$, whereas for folding (reverse) measurements $\langle W \rangle = 1024.7 \pm 0.6\ k_{\mathrm{B}}T$ and $\sigma_W^2 = 29 \pm 5\ k_{\mathrm{B}}T^2$. Hence, using Eq. (8.24) and work measurements from either unfolding or folding processes, the free-energy values $\Delta G_{NU} = 1102 \pm 3\ k_{\mathrm{B}}T$ and $\Delta G_{NU} = 1021.1 \pm 0.7\ k_{\mathrm{B}}T$ are respectively recovered. Noticeably, the average between these two estimators is $1061 \pm 3\ k_{\mathrm{B}}T$, in agreement with values obtained by matching the forward and reversed work distributions ($1058 \pm 5\ k_{\mathrm{B}}T$) and from the Bennett acceptance ratio method ($1062 \pm 5\ k_{\mathrm{B}}T$).

Determination of the free energy difference ΔG_{NU} using a theory for Bias

The huge hysteresis measured between unfolding and folding curves when pulling Barnase suggests the use of the Random Energy model [Derr 81] to determine the free energy ΔG_{NU} using unidirectional fluctuation theorems, as it is done in [Pala 11]. The key idea is to find an analytical expression for the bias (*i. e.*, the difference between the estimated free energy ΔG_{NU} determined using Eq. (8.13) and the "true" free energy, hereafter referred to as ΔG_{NU}^{*}) as a function of the number of measurements, $B_n = \Delta G_{NU} - \Delta G_{NU}^{*}$. Consequently,

$$\Delta G_{NU}^{*} = -\log \left[\frac{1}{n} \sum_{i=1}^{n} \exp \left(-\frac{W_i}{k_{\mathrm{B}}T} \right) \right] - B_n, \tag{8.25}$$

should be a proper estimate of the free-energy difference in systems with high dissipation, being n the number of unidirectional work measurements W_i ($i = 1 \ldots n$).

In reference [Pala 11] it is shown that the analytical expression for the bias satisfies:

$$B_n = \mu + \log n - \Omega \left(\log n \right)^{1/\delta} - \lambda^{(1-\delta)/\delta} \left[\gamma_{\mathrm{E}} + \frac{1-\delta}{\delta} \log \Big(\log n \Big) + \log \frac{q}{\delta} \right], \tag{8.26}$$

where $\gamma_{\mathrm{E}} = 0.577215665$ is the Euler-Mascheroni constant; Ω and δ are parameters determined by fitting the work distribution left-most tail to the following expression:

$$P(W) \sim \frac{q}{\Omega} \exp\left(\frac{|W - W_{\max}|}{\Omega}\right)^{\delta}; \tag{8.27}$$

and μ and λ are defined as:

$$\mu = (\delta - 1)\left(\frac{\Omega}{\delta}\right)^{\frac{\delta}{\delta-1}}, \qquad \lambda = \log n \left(\frac{\delta}{\Omega}\right)^{\frac{\delta}{\delta-1}}. \tag{8.28}$$

In Fig. 8.5a the fits of the left-most tails of both the forward and reversed work distributions to Eq. (8.27) are shown. In each case, the term $W_{\max}$ is fixed at the value of the work at which the experimental distribution has the maximum, Ω is related to the variance of the distribution, q is a normalization constant, and δ is related to the shape of the tail and must be larger than 1. In Table 8.2 the different numerical values are summarized.

In Fig. 8.5b the slow convergence of the Jarzynski unidirectional free-energy estimator ΔG_{NU} is shown as a function of the number of work measurements n (Eq. 8.13) for both forward and reversed data (red-open squares and blue-open circles, respectively). In addition, the respective free-energy estimators ΔG^{*}_{NU} obtained by correcting ΔG_{NU} with the bias B_n calculated according to Eq. (8.26) are also shown (solid symbols). Remarkably, the value of ΔG^{*}_{NU} obtained from forward and reversed work measurements converge to the identical value of $1048 \pm 6\ k_{\mathrm{B}}T$ for large enough values of n, which is almost 20 $k_{\mathrm{B}}T$ lower than the value measured by directly applying the Bennett acceptance ratio method (black-dashed line), by matching the work distributions according to the CFR (Eq. 8.7), or by using the cumulant expansion of the Jarzynski estimator (Eq. 8.24).

To sum up, the use of a proper theory to model the bias of the free-energy estimators provided by fluctuation theorems in systems with large hysteresis is very important to get reliable values for free-energy differences. For the case of the protein under consideration, results from the Random Energy model [Derr 81, Pala 11] have been used in order to correct the bias of the free energies estimated from non-equilibrium forward and reversed work measurements.

	Forward	Reversed
$W_{\max}$ ($k_{\mathrm{B}}T$)	1121±3	1022±1
Ω ($k_{\mathrm{B}}T$)	15±1	7±1
q ($k_{\mathrm{B}}T$)	0.57±0.05	0.55±0.06
δ	1.90±0.05	1.80±0.05

Table 8.2: Tails of forward and reversed work distributions for Barnase. Numerical values obtained by fitting Eq. (8.27) to the tails of the forward and reversed work distributions measured by pulling Barnase. Results are plotted in Fig. 8.5a.

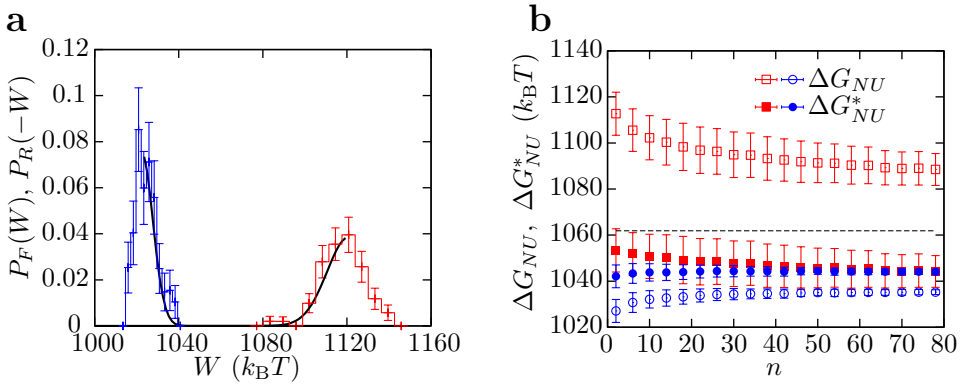

Figure 8.5: Bias theory applied to pulling experiments with Barnase. a. Forward (red) and reversed (blue) work distributions, and results of the fit to Eq. (8.27) of the respective tails (black). **b.** Convergence of Jarzynski free-energy estimator, obtained by direct application of Eq. (8.13) with the number of work measurements for both forward (red-open squares) and reversed (blue-open circles) work data, and respective estimation of ΔG^*_{NU} obtained from Eq. (8.25) by estimating the bias according to Eq. (8.26) (solid symbols). The dashed line indicates the value of the free energy measured using the Bennett acceptance ratio method (Fig. 8.4d). Error bars are obtained using the Bootstrap method.

Determination of the free energy of formation ΔG^0_{NU}

In order to estimate the free energy of formation of Barnase from mechanical denaturation, it is required to subtract the energetic contributions from stretching the handles, the peptide chain and the globular conformation, and displacing the bead in the optical trap, from the value of ΔG_{NU} measured using fluctuation relations (Eq. 8.15).

First, the value of the reversible work needed to stretch the released peptide chain from 0 to $f_{\max} = 23.3 \pm 0.1$ pN is calculated. The elastic response of Barnase was measured in Chapter 7, where the persistence length $P = 0.60 \pm 0.05$ nm and inter-aminoacid distance $d_{\text{aa}} = 0.34 \pm 0.01$ nm/aa was obtained using the WLC model. Hence, in this case one gets:

$$W_U^{\text{st}} = \int_0^{x_U(f_{\max})} f_U(x')dx' = 40.0 \pm 0.5 \ k_{\text{B}}T. \tag{8.29}$$

The error bar is obtained from propagation of the error of the experimentally determined elastic parameters.

Second, the reversible work needed to align the folded protein along the force axis, *i. e.* the protein in the globular conformation, from 0 to $f_{\min} = 1.0 \pm 0.2$ pN, is estimated using the FJC with identical Kuhn and contour length, $d = 3$ nm. Hence:

$$W_N^{\text{st}} = \int_0^{x_N(f_{\min})} f_N(x')dx' = 0.10 \pm 0.03 \ k_{\text{B}}T. \tag{8.30}$$

Thus, according to Eq. (8.16) the value of $\Delta W_{NU}^{\text{st}}$ is equal to $\sim 40.0 \pm 0.5 \ k_{\text{B}}T$ (Table 8.3), since the contribution for orienting the folded protein is negligible in this case.

The curvature observed in the experimental FDC of Barnase prevents the use of Eq. (8.19c) to estimate the energetic contributions of the handles and the bead in the optical trap. Here, both contributions must be computed separately according to Eq. (8.17). In this case, the two handles are dsDNA molecules made of 500 base pairs each. To estimate their energetic contributions, the extensible WLC model with the Bouchiat interpolation formula is used [Bouc 99]. The elastic parameters are taken from previous studies, where numerical values were obtained at similar ionic concentrations than the ones used here [Ribe 12]. Hence, the persistence length is equal to $P = 34 \pm 5$ nm, the Young modulus $S = 850 \pm 100$ pN and the interphosphate distance $d_b = 0.34$ nm/base. From Eq. (8.17):

$$\Delta W_{NU}^{\text{handles}} = \int_{x_{\text{handles}}(f_{\min})}^{x_{\text{handles}}(f_{\max})} f(x')dx' = 80 \pm 4 \ k_{\text{B}}T \tag{8.31}$$

To finish, in order to estimate the energetic contribution due to the displacement of the bead in the optical trap along the pulling protocol, the force-dependent strap stiffness determined by N. Forns and collaborators for the mini-tweezers is used [Forn 11]: $k_{\text{bead}}(f) = 0.062 + 0.00059f$ pN/nm. Thus, from Eqs. (8.17) and (8.18):

$$\Delta W_{NU}^{\text{bead}} = \int_{x_{\text{bead}}(f_{\min})}^{x_{\text{bead}}(f_{\max})} f(x')dx' \tag{8.32a}$$

$$= \int_{f_{\min}}^{f_{\max}} \frac{f'}{k_{\text{bead}}(f')} df' = 914 \ k_{\text{B}}T \tag{8.32b}$$

Once the different energetic contributions of the molecular setup have been estimated, they must be subtracted from the value of $\Delta G_{NU}*$ determined using fluctuation relations. Results are summarized in Table 8.3. The value recovered for $\Delta G_{NU}^0 = 14 \pm 7$ is in reasonable agreement within error bars with the one determined in Chapter 7 using DFS methods ($20 \pm 5 \ k_{\text{B}}T$), and the one determined using a protein-engineering approach ($16 \ k_{\text{B}}T$). However, it must be stressed that the error in the estimation of the folding free energy of the molecule at zero force ΔG_{NU}^0 is relatively large (50%), even though it is obtained by propagating the experimental errors of ΔG_{NU}^* and $\Delta W_{NU}^{\text{handles}}$, which are relatively small (0.5% and 5%, respectively). It might be that the error bars for ΔG_{NU}^* and $\Delta W_{NU}^{\text{handles}}$ are not statistically independent and hence propagation is not the best choice to determine error bars for ΔG_{NU}^0. This could decrease the estimation of relative error for ΔG_{NU}^0 without having an effect on the recovered mean value.

ΔG_{NU}^* ($k_{\text{B}}T$)	$\Delta W_{NU}^{\text{st}}$ ($k_{\text{B}}T$)	$\Delta W_{NU}^{\text{handles}}$ ($k_{\text{B}}T$)	$\Delta W_{NU}^{\text{bead}}$ ($k_{\text{B}}T$)	ΔG_{NU}^0 ($k_{\text{B}}T$)
1048±6	40.0±0.5	80±4	915	14±7

Table 8.3: Fluctuation theorem applied to Barnase. The free energy ΔG_{NU}^* is obtained from the application of Jarzynski equality and corrected by an estimation of the bias (Eqs. 8.13 and 8.26, respectively). Once the different energetic contributions from the experimental setup are removed (Eq. 8.15) the free energy of formation of Barnase is obtained.

8.5 Conclusions

Fluctuation relations are mathematical identities that allow us to extract free-energy differences from a collection of irreversible work measurements [Alem 11, Jarz 11]. In this chapter, the most important relations in the field of single molecule experiments –which are the Crooks fluctuation relation and the Jarzynski equality– are presented [Jarz 97, Croo 99, Croo 00]. The only requirements that must be satisfied for the fluctuation relations to hold are that at the beginning of both forward and reversed protocols the system under study is at equilibrium.

Using non-equilibrium work measurements obtained by pulling a two-state DNA hairpin, the validity of the Crooks fluctuation relation has been experimentally tested along the chapter using different approaches. Furthermore, limitations of fluctuation relations to determine free energies in presence of hysteresis between forward and reversed work values has been explored. In such cases, it has been shown that the use of statistical theories for the bias can be useful to improve free-energy estimations [Pala 11].

To sum up, fluctuation relations are extremely useful to determine the folding free energy at zero force, *i. e.* free energy of formation, of molecular structures from non-equilibrium pulling experiments. Until now, most studies have focused on the determination of the free energy of formation of native states, which can be easily accessed under non-equilibrium conditions. In the next chapter it will be shown how to extend fluctuation relations in order to characterize intermediate and misfolded states, which are of crucial importance for many biological processes.

Chapter 9

The extended fluctuation relation

In the previous chapter the power of fluctuations relations was explored: it is possible to obtain free-energy differences from irreversible work measurements.

However, in order to measure free-energy differences with the CFR an important condition must be satisfied in experiments: the initial state of both the forward and the reversed protocol, states A and B respectively, must be in full equilibrium. Therefore, the measured value of ΔG_{AB} using Eq. (8.2) is the free-energy difference between the final state and the initial state of the system along the forward protocol. This is a limitation if one wants to measure free-energy differences between states which are difficult to observe in equilibrium conditions and that are only transiently sampled under non-equilibrium experiments [Liph 12], such as intermediates or misfolded states.

In this chapter it is shown how to extend the CFR in order to recover the free-energy branches of the different states that can be observed along a non-equilibrium protocol as a function of the order parameter. This makes it possible to measure free-energy differences of states that are sampled under partial equilibrium conditions. This extension makes it possible to measure free-energies of metastable states, including native, intermediate and misfolded molecular states, which usually play a key role in the fate of many molecular reactions, such as protein and peptide-nucleic acid binding, specific cation binding, antigen-antibody interactions, transient states in enzymatic reactions or the formation of transient intermediates and non-native structures in molecular folders.

9.1 Partial equilibrium and kinetic states

A thermodynamic system is considered to be in total equilibrium conditions when, for a fixed value of the control parameter λ, the probability distribution $P_\lambda^{\mathrm{eq}}(x)$ of the system to be in the configuration x satisfies:

$$P_\lambda^{\mathrm{eq}}(x) = \frac{\exp\left(-\frac{E_\lambda(x)}{k_{\mathrm{B}}T}\right)}{Z_\lambda}, \qquad Z_\lambda = \sum_{x\in\mathcal{S}} \exp\left(-\frac{E_\lambda(x)}{k_{\mathrm{B}}T}\right), \tag{9.1a,b}$$

i. e., it is Gibbsian over the phase space $\mathcal{S}$ of the system (Fig. 9.1a). $E_\lambda(x)$ is the energy of the configuration x for a given λ and Z_λ is the partition function of the system at λ [Path 96].

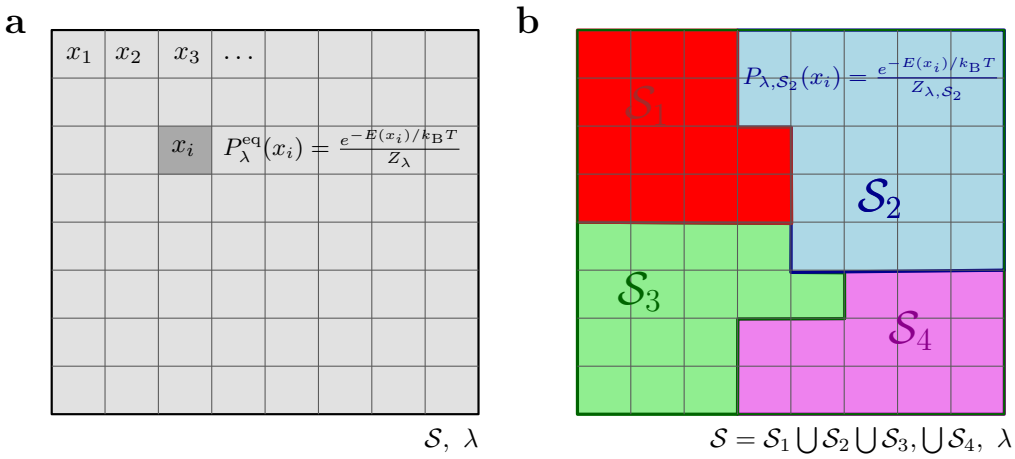

Figure 9.1: Total and partial equilibrium. a. Representation of the phase space $\mathcal{S}$ equilibrated for a given value of the control parameter λ. Squares represent different configurations of the system x_i. Under equilibrium conditions, each configuration x_i is sampled according to the Boltzmann distribution. **b.** Representation of the partially equilibrated phase space $\mathcal{S}$ for a given value of the control parameter λ. The phase space can be divided into different regions $\mathcal{S}_i$ $(i = 1, 2, 3, 4)$. Configurations x_i inside each region are sampled according to the Boltzmann distribution restricted to such region, but the statistical weights of the different regions are not required to satisfy the equilibrium distribution.

Under partial equilibrium conditions, the whole phase space $\mathcal{S}$ is divided into regions $\mathcal{S}'$ that are partially equilibrated [Palm 82]. In contrast, the statistical weights of the different regions $\mathcal{S}'$ do not necessarily follow an equilibrium distribution. This means that any configuration x inside the region $\mathcal{S}'$ is sampled according to a Boltzmann–Gibbs distribution restricted to configurations contained in that region (Fig. 9.1b). It can be mathematically described as:

$$P_{\lambda,\mathcal{S}'}(x) = P_\lambda^{\mathrm{eq}}(x)\chi_{\mathcal{S}'}(x)\frac{Z_\lambda}{Z_{\lambda,\mathcal{S}'}}, \qquad Z_{\lambda,\mathcal{S}'} = \sum_{x\in\mathcal{S}'} \exp\left(-\frac{E_\lambda(x)}{k_\mathrm{B}T}\right), \tag{9.2a,b}$$

where $\chi_{\mathcal{S}'}(x) = 1$ if $x \in \mathcal{S}'$ and zero otherwise.

In what follows, a kinetic state is defined as a partially equilibrated region of the configurational space. Kinetic states play important roles in many physical processes, such as phase transitions or molecular reactions (protein binding, enzymatic reactions, protein folding, ...) [Zema 97].

In the case of biomolecules that are mechanically unfolded, the configurational space can be considered to be partitioned into different molecular kinetic states, such as the native conformation, intermediate and misfolded states, or the unfolded conformation. As a result, during an experiment the molecule follows a sequence of kinetic states that determines its trajectory. Because of thermal fluctuations and the stochastic nature

of small systems, each independent realization of a pulling experiment may result in a different trajectory. That is, the molecule does not necessarily repeat the same sequence of kinetic states among a trajectory for different realizations of an identical protocol.

9.2 Extended fluctuation relations

The CFR (Eq. 8.2) can only be applied to obtain free-energy differences between states sampled under equilibrium conditions. However, it is possible to extend the relation to measure not only the free-energy differences between states that are observed under partial equilibrium conditions, but also to measure their free-energy branches as a function of the control parameter that is manipulated during the experiment [Mara 08, Juni 09].

Let A and B denote any two kinetic states of a thermodynamic system and λ denote a control parameter. In a forward process the system starts in partial equilibrium at A and λ varies from λ_0 to λ_1 during a time τ according to a predetermined protocol $\lambda(t)$. In the time-reversed process the system is initially in partial equilibrium at B and λ varies from λ_1 to λ_0 according to the time-reversed protocol $\tilde{\lambda}(t) = \lambda(\tau - t)$. The extended fluctuation relation (EFR) reads as [Juni 09, Alem 12]:

$$\frac{\phi_F^{A\to B}}{\phi_R^{A\leftarrow B}} \frac{P_F^{A\to B}(W)}{P_R^{A\leftarrow B}(-W)} = \exp\left(\frac{W - \Delta G_{AB}}{k_\mathrm{B}T}\right), \tag{9.3}$$

where $\Delta G_{AB} = G_B(\lambda_1) - G_A(\lambda_0)$ is the free energy difference between kinetic states B at λ_1 and A at λ_0; $P_F^{A\to B}(W)$ and $P_R^{A\leftarrow B}(-W)$ denote the partial work distributions for the forward and reversed processes that start and end at A and B respectively; $\phi_F^{A\to B}$ and $\phi_R^{A\leftarrow B}$ are the fraction of paths starting in A (or B) at λ_0 (or λ_1) and ending in B (or A) at λ_1 (or λ_0); k_B is the Boltzmann constant and T the temperature of the environment. A mathematical demonstration of the EFR can be found in appendix I.3. The EFR implies that the point at which the forward and the revered work histograms cross each other ($P_F^{A\to B}(W) = P_R^{A\leftarrow B}(-W)$) is no longer equal to ΔG_{AB} but it is equal to $\Delta G_{AB} + k_\mathrm{B}T \log\left(\phi_F^{A\to B}/\phi_R^{A\leftarrow B}\right)$. Moreover, by setting $W = 0$ into Eq. (9.3) detailed balance is recovered.

Equation (9.3) is a generalization of the CFR (Eq. 8.2), since equilibrium is a particular case of partial equilibrium. In fact, if at the beginning of the forward and the reversed processes the system is equilibrated at states A and B respectively, the two fractions $\phi_F^{A\to B}$ and $\phi_R^{A\leftarrow B}$ are equal to 1 and hence the CFR is recovered from Eq. (9.3).

If Eq. (9.3) is multiplied by $P_R^{A\leftarrow B}(-W)$ and the resulting expression is integrated over W one gets an extended version of the Jarzynski equality for kinetic states,

$$\left\langle \exp\left(-\frac{W - \Delta G_{AB}}{k_\mathrm{B}T}\right)\right\rangle_F = \frac{\phi_R^{A\leftarrow B}}{\phi_F^{A\to B}}, \tag{9.4}$$

where $\langle \ldots \rangle_F$ denotes the average over forward trajectories.

There are two main differences between the EFR in Eq. (9.3) and the CFR in Eq. (8.2). First, the use of partial work distributions, which implies that from all the

measured forward (reversed) trajectories, only those starting in state A (B) and ending in state B (A) are selected. Second, the presence of the prefactor $\phi_F^{A\to B}/\phi_R^{A\leftarrow B}$, which introduces the additional correction $-k_{\mathrm{B}}T\log(\phi_F^{A\to B}/\phi_R^{A\leftarrow B})$ into the Crooks estimation of the free-energy difference between kinetic states. The omission of such correction yields to systematically biased results for the free-energy differences between different kinetic states [Juni 09]. Moreover, for the case of kinetic structures that apparently behave reversibly under the protocol, ΔG_{AB} is not just equal to the measured work during the experiment, which is apparently reversible, since the term $k_{\mathrm{B}}T\log(\phi_R^{A\leftarrow B}/\phi_F^{A\to B})$ must be added. Although this correction might be small, it is important in many situations. For instance, ignoring the prefactor $k_{\mathrm{B}}T\log(\phi_R^{A\leftarrow B}/\phi_F^{A\to B})$ even if very low hysteresis is obtained between forward and reversed processes in a case where $\phi_R^{A\leftarrow B}=0.01$ and $\phi_F^{A\to B}=1$ would underestimate by 4.5 $k_{\mathrm{B}}T$ the free energy ΔG_{AB}.

9.3 Experimental validation of the extended fluctuation relation

In order to prove the validity of the EFR, pulling experiments performed at two different puling speeds (60 and 180 nm/s) with CD4 will be performed. The unfolding process is identified with the forward protocol, whereas the folding process with the reversed one. In Fig. 9.2a, the measured FDC at the two different conditions are shown. As before, the work is calculated as the area below the FDC along an interval of integration (λ_0,λ_1). Here λ_0 is set at -40 nm (Fig. 9.2a), where the molecule is always folded in the N state either in the forward (unfolding) and reversed (folding) protocols. The EFR is applied to the experimental data for different values of λ_1 in the interval $\lambda_1\in[-40,55]$ nm. At any value of λ_1 the molecule is in partial equilibrium, and its phase space is divided into two regions identified with states N and U. For each value of λ_1 the following steps are performed:

1. The set of forward trajectories is classified into two subsets according to the state of the molecule at the end of the protocol, where λ equals λ_1. As a result, in one of the subsets there are all the trajectories where the molecule starts at state N at λ_0 and ends at state N at λ_1 (hereafter referred to as "$N\to N$" trajectories), and in the second subset there are all the trajectories where the molecule starts at N and ends at U (trajectories "$N\to U$"). Then, the fractions of forward trajectories $\phi_F^{N\to N}$ and $\phi_F^{N\to U}$ are computed as the fraction of the number of trajectories $N\to N$ and $N\to U$, respectively, and the total number of forward trajectories. Note that $\phi_F^{N\to N}+\phi_F^{N\to U}=1$. Results at different values of λ_1 obtained at two different pulling speeds are shown in Fig. 9.2b.

2. Reversed trajectories are classified according to the state of the hairpin, N or U, at the beginning of the reversed protocol at λ_1. At the end of the reversed protocol, at λ_0, the hairpin is always at N. Therefore, there are two types of forward trajectories which are "$N\leftarrow N$" and "$N\leftarrow U$". Under this circumstances,

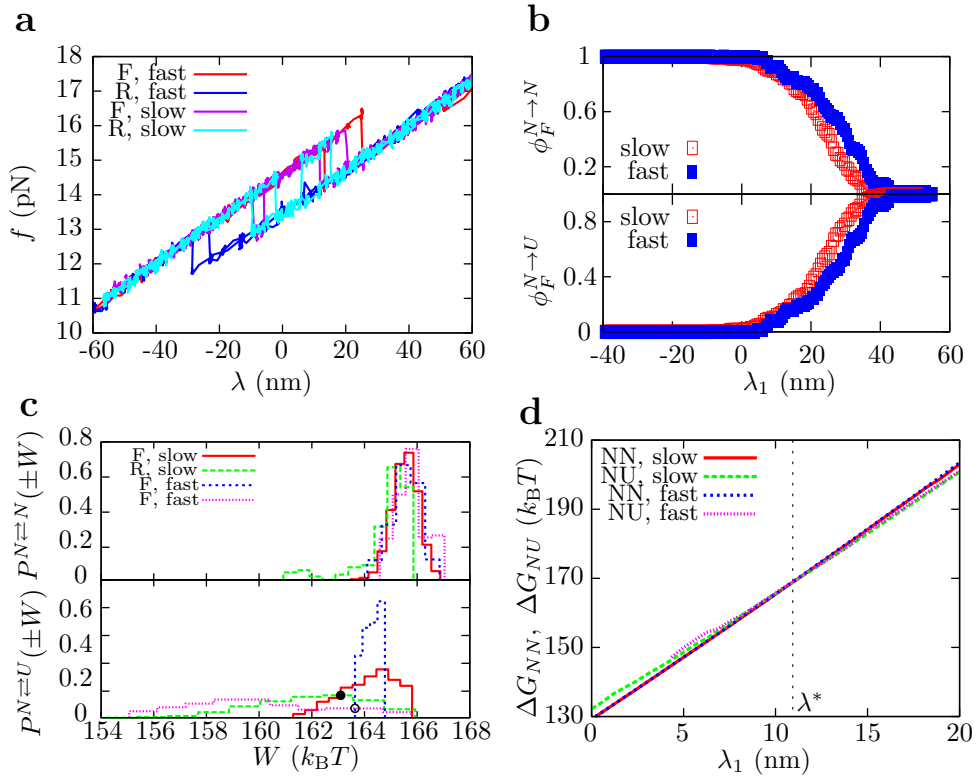

Figure 9.2: Application of the extended fluctuation relation to pulling experiments with CD4. **a.** FDC measured by pulling CD4 at two at different pulling speeds (60 and 180 nm/s, slow and fast respectively). **b.** Fraction of forward trajectories $\phi_F^{N\to N}$ (top) and $\phi_F^{N\to U}$ (bottom) as a function of λ_1 obtained at the two different pulling speeds. Note that $\phi_F^{N\to N} + \phi_F^{N\to U} = 1$. **c.** Partial forward and reversed work distributions $P_F^{N\to N}(W)$, $P_R^{N\leftarrow N}(-W)$ (top) and $P_F^{N\to U}(W)$, $P_R^{N\leftarrow U}(-W)$ (bottom) for work values obtained by integrating the FDC between $\lambda_0 = -40$ nm and $\lambda_1 = 10$ nm. **d.** Free-energy branches for N and U relative to $\Delta G_N(\lambda_0)$ obtained by direct application of the EFR (Eq. 9.3) setting $\lambda_0 = -40$ nm and varying λ_1. The vertical dotted line indicates the value of λ^* at which states N and U have the same free energy, $\Delta G_{NN}(\lambda^*) = \Delta G_{NU}(\lambda^*)$. Results are shown for two different pulling speeds.

$\phi_R^{N\leftarrow B} = 1$ and $\phi_R^{U\leftarrow B} = 0$ always, being B either N or U, since the final state of the molecule at the selected value of λ_0 along a forward protocol is always N.

3. Partial work distributions $P_F^{N\to N}(W)$, $P_F^{N\to U}(W)$, $P_R^{N\leftarrow N}(-W)$ and $P_R^{N\leftarrow U}(-W)$ are calculated for each corresponding set of forward and reversed trajectories. In Fig. 9.2c results obtained at $\lambda_1 = 10$ nm are shown. Note that now the point at which $P_F^{N\to U}(W)$ and $P_R^{N\leftarrow U}(-W)$ cross each other does actually depend on the pulling speed.

4. ΔG_{NN} and ΔG_{NU} are determined using the EFR once the partial work distributions and the prefactors $\log \phi_F^{N\to B}/\phi_R^{N\leftarrow B} = \log \phi_F^{N\to B}$ ($B = N,\ U$) are known. The representation of free-energy differences ΔG_{NN} and ΔG_{NU} as a function of the control parameter λ_1 are known as the free-energy branches of states N and

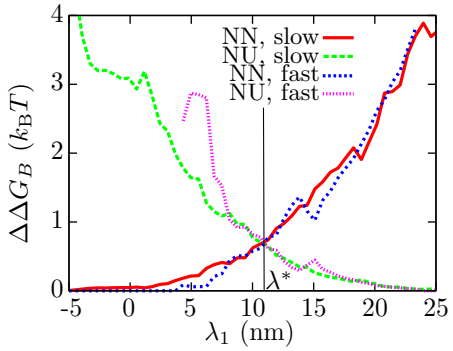

Figure 9.3: Free-energy branches. Free-energy branches for N and U relative to the full equilibrium free energy, $\Delta\Delta G_B = \Delta G_{NB} - \Delta G$, where $B = N,\ U$, for each value of λ_1 (Eq. 9.5). The vertical line indicates the value of λ^* at which states N and U have the same free energy, $\Delta\Delta G_N(\lambda^*) = \Delta\Delta G_U(\lambda^*)$.

U, respectively (Fig. 9.2d).

In Fig. 9.2d the free-energy branches $\Delta G_{NN}(\lambda_1)$ and $\Delta G_{NU}(\lambda_1)$ obtained using the EFR for the two different pulling speeds are shown. In both cases, the free energy of state N at $\lambda_0 = -40$ nm is taken as the reference energy. As expected, the profile of the free-energy branches does not depend on the speed of the pulling protocol.

The full equilibrium free energy of the system at each value of λ_1 can be defined as:

$$\Delta G = -k_{\mathrm{B}}T \log\left[\sum_B \exp\left(-\frac{\Delta G_{NB}}{k_{\mathrm{B}}T}\right)\right] \tag{9.5}$$

$$= -k_{\mathrm{B}}T \log\left[\exp\left(-\frac{\Delta G_{NN}}{k_{\mathrm{B}}T}\right) + \exp\left(-\frac{\Delta G_{NU}}{k_{\mathrm{B}}T}\right)\right] \tag{9.6}$$

In Fig. 9.3, the free-energy branches for states N and U are represented taking as the reference state the full free energy at each value of λ_1, that is $\Delta\Delta G_B = \Delta G_{NB} - \Delta G$ (for $B = N,\ U$). At low values of λ_1 the stability of the DNA hairpin is governed by state N, whereas at large values of λ_1 the most stable state is U. It can be observed that at $\lambda_1 = \lambda^* \simeq 10$ nm states N and U coexist as $\Delta\Delta G_N(\lambda^*) \simeq \Delta\Delta G_U(\lambda^*)$. Force values of states N and U along the FDC at λ^* are approximately 15.2 ± 0.2 pN and 14.0 ± 0.2 pN, respectively. This gives an average force value equal to 14.6 ± 0.1 pN (Fig. 9.2a). Remarkably, this value is in good agreement with the coexistence force of CD4 predicted by the NN model with the UO set or the UNZ set of base-pair free energies (f_c =15.3 and 14.6 pN, respectively, Fig. 4.3b).

In Fig. 9.2c it is observed that the value of the work at which the forward and the reversed partial work histograms (of type $N \rightleftarrows U$) cross each other does depend on the pulling speed (*i. e.*, it does depend on the pulling protocol), in contrast to the free-energy branches ΔG_{NN} and ΔG_{NU} (or $\Delta\Delta G_N$ and $\Delta\Delta G_U$), which are independent of the protocol (Fig. 9.2d and Fig. 9.3). This indicates that the presence of the prefactor $\phi_F^{A\to B}/\phi_F^{A\leftarrow B}$ in the EFR is key to properly estimate free-energy differences under partial equilibrium conditions. Hence, this result provides a verification of the EFR.

There are alternative ways to experimentally visualize the validity of the EFR. A first approach is given by the the Bennett acceptance ratio method (Section 8.3.1), used to determine ΔG_{NU} from the set of work values obtained by integrating the FDC from

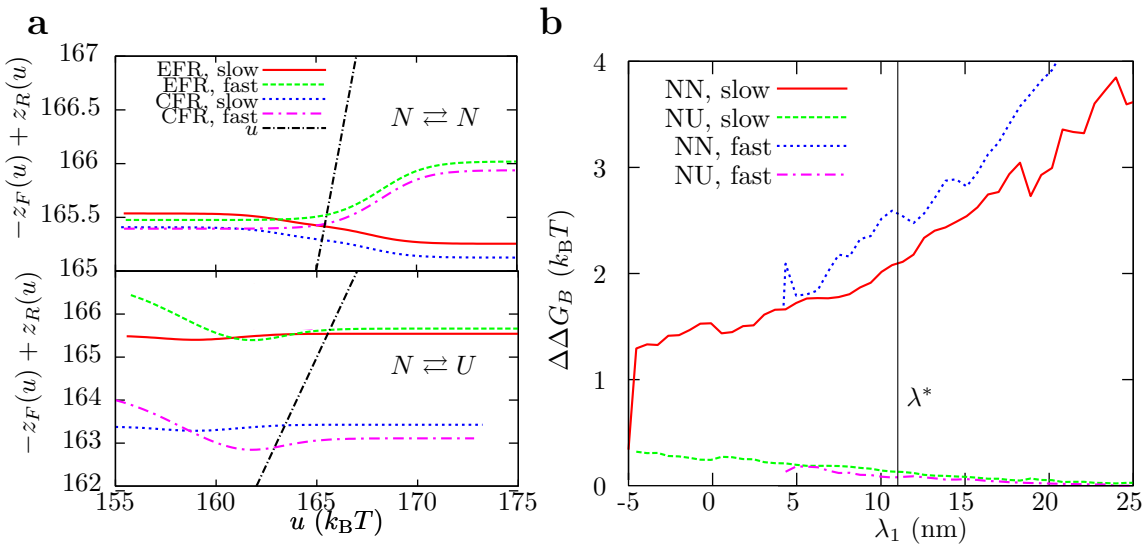

Figure 9.4: Experimental validation of the EFR. **a.** Bennett acceptance ratio method applied to two different sets of work values calculated by the integrating experimental FDC between $\lambda_0 = -40$ nm and $\lambda_1 = 10$ nm obtained at different pulling speeds, for trajectories of type $N \rightleftarrows N$ (top) and $N \rightleftarrows U$ (bottom). The transcendental Eq. (9.7) is used according to the EFR, whereas the transcendental Eq. (8.11) is used according to the CFR. **b.** Free-energy branches of states N and U relative to the full equilibrium free energy ΔG (Eq. 9.5), *i. e.* $\Delta\Delta G_B = \Delta G_{NB} - \Delta G$ ($B = N,\ U$), measured using the CFR for pulling experiments performed at two different pulling speeds. The vertical line indicates the value of λ^* at which states N and U have the same free energy according to the EFR.

$\lambda_0 = -40$ nm to $\lambda_1 = 10$ nm (Fig. 9.2a). The molecule is in full equilibrium at λ_0, but it is only partially equilibrated at λ_1. Under these circumstances, the correct application of the Bennett acceptance ratio method must take into account the presence of the prefactor $\phi_F^{A\to B}/\phi_R^{A\leftarrow B}$ in the EFR in order to determine free-energy differences. In this case, the following transcendental equation has to be solved:

$$\frac{u}{k_\mathrm{B}T} = -z_F(u) + z_R(u), \tag{9.7}$$

where

$$z_R(u) = \log\left[\frac{1}{n_R}\sum_{i=1}^{n_R}\left(\frac{1}{1+\frac{n_F}{n_R}\frac{\phi_F^{A\to B}}{\phi_R^{A\leftarrow B}}e^{-\beta(W_i+u)}}\right)\right] \tag{9.8a}$$

$$z_F(u) = \log\left[\frac{1}{n_F}\sum_{i=1}^{n_F}\left(\frac{e^{-\beta W_i}}{1+\frac{n_F}{n_R}\frac{\phi_F^{A\to B}}{\phi_R^{A\leftarrow B}}e^{\beta(W_i-u)}}\right)\right]. \tag{9.8b}$$

Figure 9.4a shows the results of the Bennett acceptance ratio method obtained using the transcendental equation (9.7), that is, using the EFR (solid-red and dashed-green lines) or the transcendental equation (8.11), *i. e.*, using the CFR (blue-dotted and magenta-dashed-dotted line) for the trajectories of type $N \rightleftarrows N$ and $N \rightleftarrows U$. Results are not

illuminating for trajectories of type $N \rightleftarrows N$ (Fig. 9.4a, top). In contrast, for trajectories of type $N \rightleftarrows U$ it is clearly observed that the solution of the transcendental equation depends on the pulling speed if the prefactor $\phi_F^{A\to B}/\phi_R^{A\leftarrow B}$ is not included (Fig. 9.4a, bottom).

A second approach to prove the validity of the EFR is to determine the free-energy branches of states N and U obtained without the prefactor $\phi_F^{A\to B}/\phi_R^{A\leftarrow B}$. These are shown in Fig. 9.4b relative to the full equilibrium free energy ($\Delta\Delta G_B = \Delta G_{NB} - \Delta G$, $B = N,\ U$, Eq. 9.5). In this case, free-energy branches depend on the pulling speed, specially for state U. Moreover, these results suggest that the stability of CD4 is always dominated by state U under pulling experiments (*i. e.*, the free-energy branches for N and U do not cross at any value of λ_1). Hence, it is observed that the use of the EFR is required to properly recover the thermodynamic stabilities of the two states.

9.4 Free-energy measurements of kinetic molecular states

The EFR (Eq. 9.3) makes it possible to recover the free energy of formation of kinetic molecular states. In this chapter, two types of kinetic states are investigated: molecules I1 and I2, which have intermediate kinetic states on-pathway to the native state, and molecules M1 and M2, which exhibit misfolded kinetic states off-pathway to the native state.

9.4.1 Intermediate states

Molecular intermediate states behave as kinetic states on-pathway under non-equilibrium pulling experiments. When a force is applied to the ends of a molecule with an intermediate state on-pathway to the unfolded state, such intermediate is revealed as a new force branch before the molecule reaches the unfolded-stretched conformation. In some occasions, intermediates can be observed and thermodynamically characterized using equilibrium experiments. However, this is not always the case. The use of the EFR paves the way to determine the thermodynamic stability of intermediate states that can only be generated dynamically, that is, that can only be observed under non-equilibrium conditions.

In what follows, it is shown how to extract the free-energy of formation of intermediate states for two different DNA hairpins.

Hairpin with one intermediate (I1)

The EFR (Eq. 9.3) is first applied to the DNA hairpin I1, which was already introduced in Section 5.3. This hairpin is characterized by the presence of an internal loop in the middle of the stem (Fig. 5.18a) that favors the existence of an intermediate state I along the molecular unfolding pathway. As it can be observed in Fig. 5.18b, the FEL evaluated at the coexistence force f_c reveals three main conformational states, which are the native N ($n = 0$), the intermediate I ($n = 11-14$) and the unfolded U ($n = 23$). Experimental

hopping and pulling traces exhibit three force branches corresponding to the three states (Fig. 5.18d-e and Fig. 9.5a, respectively).

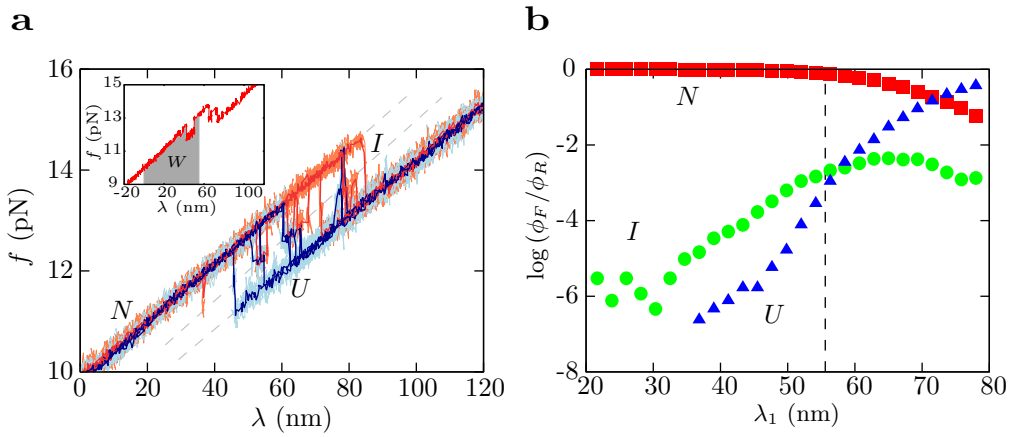

Figure 9.5: Pulling experiments with hairpin I1. a. Unfolding (red) and folding (blue) FDCs recorded at 60 nm/s exhibiting three branches of force corresponding to three states N, I and U. Inset: the work along a trajectory is calculated as the area below the FDC. **b.** Contribution of the prefactor $\log(\phi_F^{N\to B}/\phi_R^{N\leftarrow B})$, $B = N$ (red squares), $B = I$ (green circles), $B = U$ (blue triangles) to the free energy of each kinetic state as a function of λ_1. Note that for I this contribution is larger than 2 k_BT (in absolute value) over the entire range of λ_1 values. The vertical dashed line indicates the value of λ^* at which states N and U have the same free energy, $\Delta G_{NN}(\lambda^*) = \Delta G_{NU}(\lambda^*)$.

The free-energy branches of states N, I and U can be recovered by applying the recipe described in the previous section. As for hairpin CD4 and Barnase (Section 8.4), the work along a given trajectory is calculated as the area below the FDC. Here, λ_0 is set to 0 nm, where I1 is equilibrated at N as observed in Fig. 9.5a. λ_1 varies from 45 to 65 nm, where the molecule is partially equilibrated between states N, I and U. For each value of λ_1, forward and reversed trajectories are classified according to the initial and final states of the molecule at λ_0 and λ_1, respectively. Therefore, there are three types of forward trajectories, which are $N \to N$, $N \to I$ and $N \to U$, and also three types of reversed trajectories, which are $N \leftarrow N$, $N \leftarrow I$ and $N \leftarrow U$. The fractions $\phi_F^{N\to B}$, being $B = N, I$, or U, are calculated as the ratio between the number of forward trajectories of type $N \to B$ and the total number of forward trajectories. On the other hand, $\phi_R^{N\leftarrow B} = 1$ for any B (N, I or U). From $\phi_F^{N\to B}$ and $\phi_R^{N\leftarrow B}$ the effect of the prefactor $\log\left(\phi_F^{N\to B}/\phi_R^{N\leftarrow B}\right) = \log \phi_F^{N\to B}$ on the free energy estimated using the EFR can be obtained. It is observed in Fig. 9.5b that such contribution can be as large as 6 k_BT for the molecule under consideration (see $\lambda_1 = 20$ nm for state U), and it is larger that 2 k_BT for I along the entire range of λ_1 values. Next, the forward and reversed partial work distributions are measured (as shown for $\lambda_1 = 55.6$ nm in Fig. 9.6a). Finally, the EFR is applied using the Bennett acceptance ratio method to get the free-energy differences ΔG_{NN}, ΔG_{NI} and ΔG_{NU} (example shown for $\lambda_1 = 55.6$ nm in Fig. 9.6b).

The free-energy branches, *i. e.* the values of ΔG_{NN}, ΔG_{NI} and ΔG_{NU} as a function of λ_1, are plotted in Fig. 9.6c. The inset shows the free energy of each state measured relative to the full equilibrium free energy of the system, that according to Eq. (9.5)

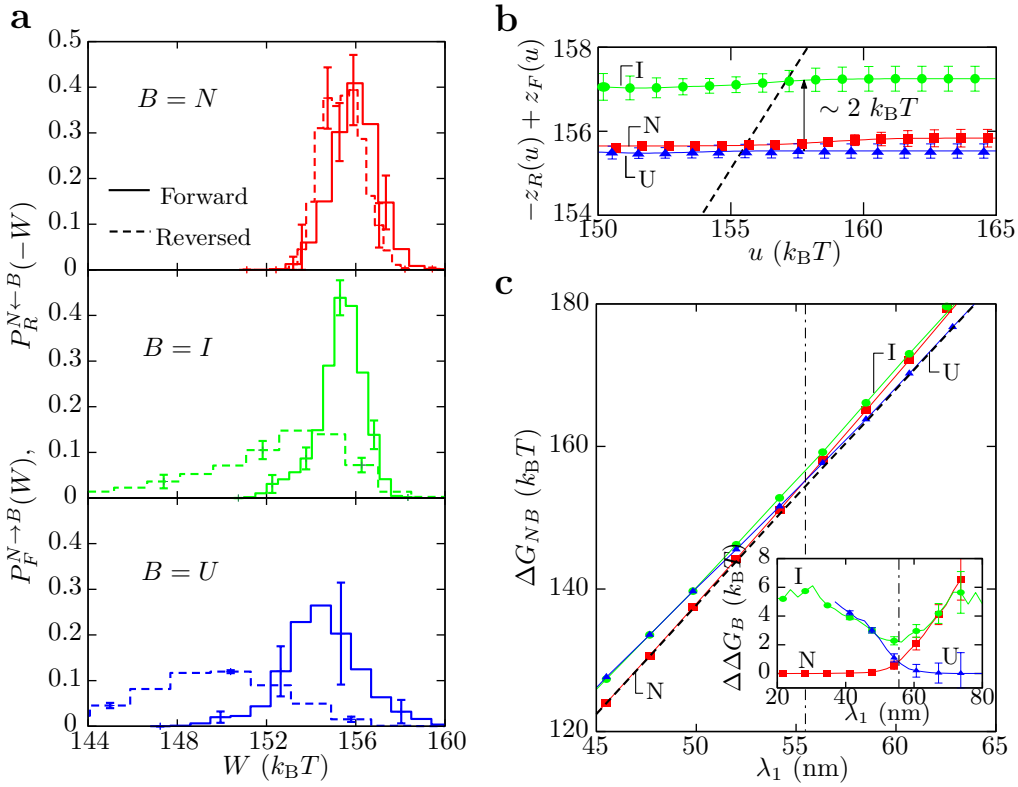

Figure 9.6: EFR applied to hairpin I1. a. Partial forward (solid lines) and reversed (dashed lines) work distributions measured for λ_0=0, where $A = N$, and λ_1=55.6 nm, where $B = N$ (top), $B = I$ (middle) or $B = U$ (bottom). **b.** Bennett acceptance ratio method applied to the work measurements shown in panel b to obtain the free-energy differences ΔG_{NB} for $B = N$ (red squares), I (green circles) and U (blue triangles). **c.** Reconstruction of the free-energy branches, relative to $G_N(\lambda_0)$, for states N, I and U obtained by fixing λ_0 and letting λ_1 change between 45 and 65 nm (color code as in b). The black-dashed curve is the full free energy of the system, $\Delta G = -k_BT \log \sum_{B=N,I,U} e^{-\Delta G_{NB}/k_BT}$. Inset: free-energy branches relative to the full equilibrium free energy, $\Delta\Delta G_B = \Delta G_{NB} - \Delta G$ for $B = N,\ I,\ U$.

is equal to $\Delta G = -k_BT \log\left(e^{-\Delta G_{NN}/k_BT} + e^{-\Delta G_{NI}/k_BT} + e^{-\Delta G_{NU}/k_BT}\right)$. The vertical dashed-dotted line at λ^*=55.6 nm indicates the coexistence point of N and U. For $\lambda_1 < \lambda^*$ ($\lambda_1 > \lambda^*$) N (U) is the most stable state, while I is never the absolute free-energy minimum for any λ_1. This is in agreement with the behavior of the theoretical FEL, which predicts that I is never the most stable state of the molecule.

In order to emphasize the validity of the EFR, the logarithm of the ratio of the forward and reversed work distributions obtained at λ^* (Fig. 9.6a) plus the logarithm of the prefactor $\phi_F^{N\to B}/\phi_F^{N\leftarrow B}$ (Fig. 9.5b) is plotted as a function of work in k_BT units for the three states N, I and U. As predicted by Eq. (9.3), the slope is approximately 1 for all three cases.

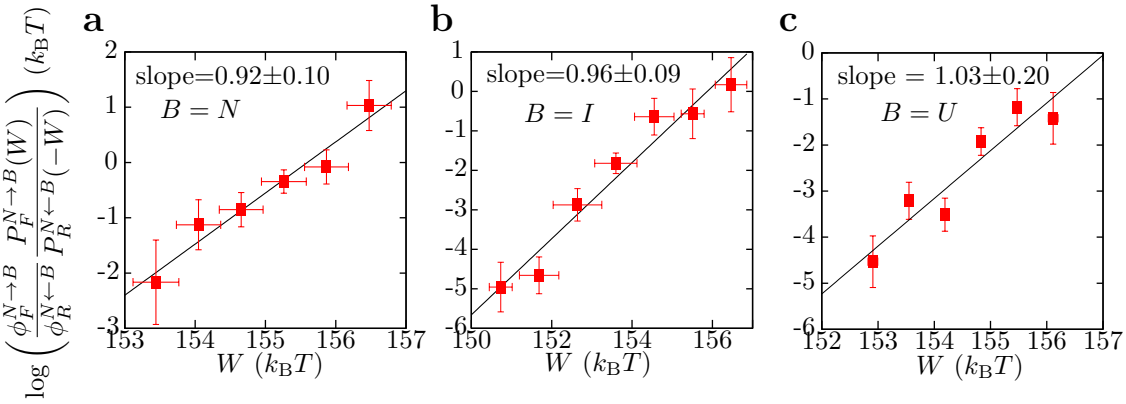

Figure 9.7: Validation of the EFR using hairpin I1. Representation of the logarithm of the ratio of the forward and reversed work distributions plus the logarithm of the prefactor $\phi_F^{N\to B}/\phi_F^{N\leftarrow B}$ obtained at λ^* for $B = N$ (**a**), $B = I$ (**b**) and $B = U$ (**c**).

(k_BT)	$\log\left(\phi_F^{N\to B}/\phi_R^{N\leftarrow B}\right)$	ΔG_{NB} (EFR)	ΔG_{NB} (CFR)	$\Delta W_{NB}^{\rm st}$
$B = N$	-0.11± 0.02	156±1	156±1	0±0.2
$B = I$	-2.77± 0.02	157±1	154±1	9.3±0.2
$B = U$	-3.12± 0.02	156±1	152±1	18.4±0.2

$\Delta W_{NB}^{\rm bead+handles}$	ΔG_{NB}^0 (EFR)	ΔG_{NB}^0 (eq.)	ΔG_{NB}^0 (UO)	ΔG_{NB}^0 (UNZ)
156±3	0±3	0±0	0	0
115±3	30±3	31±2	30.5	27.1
83±3	55±3	61±2	60.1	56.1

Table 9.1: Experimental measurement of ΔG_{NN}^0, ΔG_{NI}^0 and ΔG_{NU}^0 for hairpin I1. Numerical estimates for $\log\left(\phi_F^{N\to B}/\phi_R^{N\leftarrow B}\right)$ ($B = N,\ I,\ U$), the free energy ΔG_{NB} obtained using both the EFR and the CFR, and the reversible work differences $\Delta W_{NB}^{\rm st}$ and $\Delta W_{NB}^{\rm bead+handles}$ (Eqs. 8.16 and 8.19c) obtained by integrating the FDC between $\lambda_0 = 0$ and $\lambda_1 = 55.6$ nm are provided. In addition, the free energy of formation of each states experimentally determined using the EFR and equilibrium-based hopping experiments (Appendix J), and theoretical predictions obtained using the NN model with the UO or UNZ set of base-pair free energies (Table 4.1) are provided. Error bars contain statistical and systematic errors. All the magnitudes are given in k_BT units.

By subtracting the energetic contributions at λ^* due to the displacement of the bead captured in the optical trap, and to the stretching of the handles and the released ssDNA it is possible to extract the free energies of formation of the different structures with respect to the random coil state at zero force (Section 8.4). The values $\Delta G_{NU}^0 = 55 \pm 3\ k_BT$ and $\Delta G_{NI}^0 = 30 \pm 3\ k_BT$ are recovered ($\Delta G_{NN} \simeq 0\ k_BT$), in agreement with free energy predictions obtained using the NN model. In Table 9.1 the most important contributions to properly get the free energy of formation of the three states from pulling experiments is summarized. The estimated values for the free energies ΔG_{NB} ($B = N$, I and U) obtained using the CFR (that is, by applying the CFR into the partial work distributions and neglecting the contribution of the prefactor $\phi_F^{A\to B}/\phi_R^{A\leftarrow B}$ in the EFR) are also given. Note that a discrepancy of 4 k_BT can be obtained in the estimation

of ΔG_{NU}.

Equilibrium-based hopping experiments can also be carried out with hairpin I1. For a given range of values of the trap-pipette distance λ the three states N, I and U can be observed, as seen in Fig. 9.5d. Applying detailed balance or hidden Markov models to analyze the experimental data it is possible to measure the free energy of formation for the three states (Appendix J). Results, given in Table 9.1, are also in good agreement with the values obtained using the EFR and theoretical predictions.

Hairpin with two intermediates (I2)

Now, to illustrate the power of the EFR to thermodynamically characterize kinetic states on-pathway, the fluctuation theorem is applied to hairpin I2. Hairpin I2, which was already introduced in Section 5.3, consists of a three-way junction (Fig. 5.21a). The theoretical FEL predicts the existence of two intermediate states (Fig. 5.21b), referred to as I_1 and I_2, along the molecular unfolding pathway. State I_1 ($n = 21$) corresponds to the structure where the stem is fully open and the two hairpins of the bifurcation are formed. I_2 ($n = 37$) is made of two identical structures, where the full stem and one hairpin in the bifurcation are open. As these two configurations have identical molecular extensions, I_2 is considered as a single state. In pulling experiments four force branches, corresponding to states N, I_1, I_2 and U are distinguished (Fig. 5.22c). Four levels of force can also be observed under equilibrium conditions. However, hopping between the four conformations was never observed for any value of λ in the accessible experimental timescales precluding the use of equilibrium hopping experiments to determine free-energy differences: at high forces, once the molecule leaves state N, hopping between I_1, I_2 and U (but not N) is observed; at low forces, once the molecule reaches N it never escapes out (Fig. 5.22a-b).

To determine the free-energy branches of states N, I_1, I_2 and U, λ_0 was set to 0 nm, where the molecule is always folded at N (Fig. 5.22c) and λ_1 was varied between 110 and 190 nm. Due to the large hysteresis exhibited by this molecule between unfolding and folding FDC, the standard pulling protocol only generates reversed trajectories of type $N \leftarrow U$. This implies that it is impossible to find a value of λ_1 for which forward trajectories of type $N \rightarrow N$, $N \rightarrow I_1$ and $N \rightarrow I_2$ and reversed trajectories of type $N \leftarrow N$, $N \leftarrow I_1$ and $N \leftarrow I_2$ are simultaneously sampled.

In order to measure forward and reversed partial work distributions for the four states, I2 is pulled back and forth between $\lambda_0 = 0$ nm, where the molecule is at state N for both unfolding and folding trajectories, and $\lambda_1 = 183$ nm, where the molecule is partially equilibrated at states N, I_1, I_2 or U for unfolding trajectories (Fig. 5.22c). Hence, when the molecule reaches λ_1 along a forward process, it can be partially equilibrated in any of the four states N, I_1, I_2 or U. If the folding process starts before the molecule completely unfolds to state U, reversed trajectories of type $N \leftarrow B$ ($B = N, I_1$ and I_2) will be generated. Examples of such pulling experiments and the measured forward and reversed work distribution are shown in Fig. 9.8 and Fig. 9.9a, respectively. Interestingly, along reversed trajectories $N \leftarrow N$, the molecule may or may not unfold during the reversed process. This generates trajectories with extremely different work values, which results

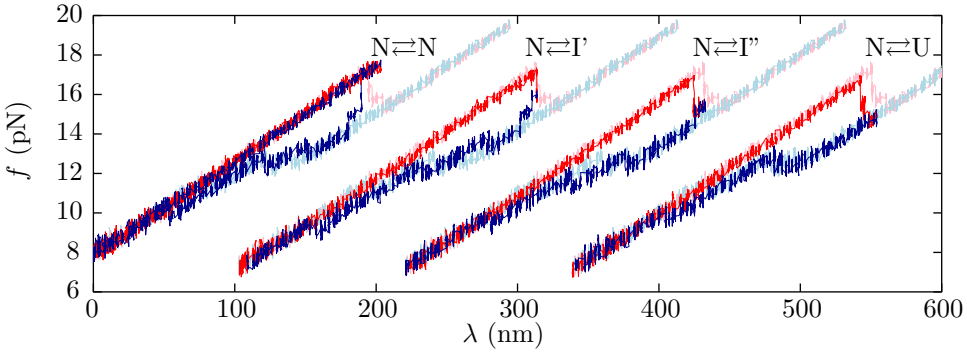

Figure 9.8: FDC for I2 obtained with a new pulling protocol. Unfolding (red) and folding (blue) FDCs recorded at 60 nm/s: the four branches of force correspond to the four states N, I_1, I_2 and U. Curves plotted using light colors are measured by setting extreme values of λ such that the molecule equilibrates at N and U at the initial and final pulling conditions respectively. In curves plotted using dark colors λ_1 was chosen so that at the end of the forward process and at the beginning of the reversed one the molecule can be found in any state: N (two patterns can be observed in the reverse trajectories with different degrees of dissipation), I_1, I_2 or U. Curves corresponding to different transitions are shifted along the horizontal axis for the sake of clarity.

in a work distribution with two characteristic peaks (Fig. 9.9a, top).

Once the partial work distributions are known, the free-energy differences ΔG_{NN}, ΔG_{NI_1}, ΔG_{NI_2} and ΔG_{NU} can be extracted by determining the corresponding fraction of forward and reversed trajectories $\phi_F^{N\to B}$ and $\phi_R^{N\leftarrow B} = 1$ ($B = N,\ I_1,\ I_2,\ U$) and applying the EFR. In Table 9.2 there are summarized the different contributions to the measured free energies for the different states obtained from work values calculated by integrating the FDC between $\lambda_0 = 0$ nm and $\lambda_1 = 183$ nm. The energetic contributions coming from the experimental setup and predicted values for the free energy of formation determined using the NN model and the UO and UNZ set of base-pair free energies are also provided. Note that the size of the error bars is comparable to the discrepancy observed between the free energy predictions [Sant 98, Zuke 03, Hugu 10b] (Table 4.1).

To get the free-energy branches of the different states, pulling experiments setting the final value of λ at different positions should be carried out in order to properly sample the four type of forward and reversed trajectories at different positions of the control parameter ($A \rightleftarrows B,\ A, B = N,\ I_1,\ I_2,\ U$). For simplicity, the extended version of the Jarzynski equality (Eq. 9.4) is used, together with standard unfolding experiments where the molecule is always either folded or unfolded at the minimum and maximum values of λ of the pulling protocol (Fig. 5.22c). To extract work values from forward trajectories, the value of λ_0 is set to 0 nm, and λ_1 varies between 110 and 190 nm. At each value of λ_1, the fraction of forward/unfolding trajectories of type $N \to N$, $N \to I_1$, $N \to I_2$ and $N \to U$ is computed, and $\phi_R^{N\leftarrow B}$ is taken equal to 1. Then, the extended Jarzynski equality (EJE) is applied to unfolding trajectories in order to get the free energies ΔG_{NN}, ΔG_{NI_1}, ΔG_{NI_2} and ΔG_{NU} at each value of λ_1.

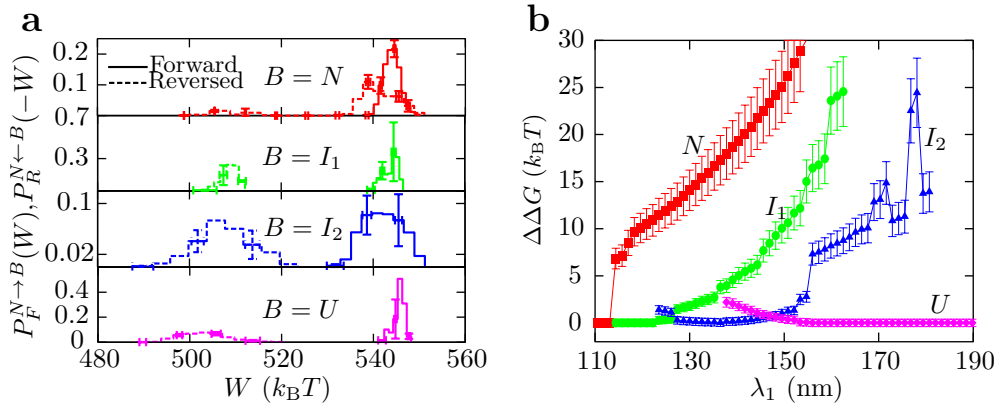

Figure 9.9: FER applied to hairpin I2. a. Partial work histograms for work values measured between λ_0=0 nm, where $A = N$, and $\lambda_1 = 183$ nm, where $B = N$ (red, first panel), $B = I_1$ (green, second panel), $B = I_2$ (blue, third panel) or $B = U$ (purple, bottom panel). Dark colors refer to forward work distributions and light colors to reversed work distributions. **b.** Free-energy branches of states N (red squares), I_1 (green circles), I_2 (blue triangles) and U (purple diamonds) measured relative to the full free energy of the system, $\Delta G = -k_BT \log \sum_{B=\mathrm{N,I_1,I_2,U}} e^{-\Delta G_{NB}/k_BT}$. Error bars in panel b indicate the standard statistical deviation computed over 400 cycles for a given molecule. These were obtained using the bootstrap method.

Similarly to the standard Jarzynski estimator (Eq. 8.13), the free-energy estimator provided by the EJE (Eq. 9.4) is strongly biased too [Pala 11]. Instead of using the Random Energy model approach to determine the magnitude of the bias (Eq. 8.26), here it is estimated by taking the difference between the free energy values ΔG_{NB} ($B = N,\ I_1,\ I_2,\ U$) determined at $\lambda_1 = 183$ nm either using standard pulling experiments and Eq. (9.4) or using the Bennett acceptance ratio method in pulling experiments where kinetic states are partially equilibrated at λ_1 (Fig. 9.8). Table 9.3 shows the magnitude of the free energies measured for each state using the EFR or the EJE, and the corresponding bias. Henceforth, the bias is assumed to be independent of λ_1. Hence, reliable free-energy branches are obtained by subtracting the magnitude of the bias shown in Table 9.3 to the free energies ΔG_{NB} ($B = N,\ I_1,\ I_2,\ U$) determined using the EJE for the different values of λ_1.

Finally, the free-energy branches of states N, I_1, I_2 and U for hairpin I2 are shown in Fig. 9.9b. In contrast to I1, kinetic intermediates present in I2 become the most stable states at a given range of λ. At low values of λ, the thermodynamic stability is dominated by N, and as λ increases stability shifts to I_1, I_2 and finally to U. This behavior is also predicted from the calculation of the molecular FEL using the NN model (Fig. 5.21b).

$(k_{\rm B}T)$	$\log\left(\phi_F^{N\to B}/\phi_R^{N\leftarrow B}\right)$	ΔG_{NB} (EFR)	ΔG_{NB} (CFR)	$\Delta W_{NB}^{\rm st}$
$B=N$	-0.2± 0.2	543±3	544±3	0±0
$B=I_1$	-3.0± 0.2	527±1	525±1	18.9±0.2
$B=I_2$	-2.4± 0.2	526±1	525±1	30.5±0.2
$B=U$	-3.7± 0.2	528±1	526±1	41.7±0.2

$\Delta W_{NB}^{\rm bead+handles}$	ΔG_{NB}^0 (EFR)	ΔG_{NB}^0 (UO)	ΔG_{NB}^0 (UNZ)
543±4	0±4	0	0
469±6	40±6	41.9	39.4
417±6	80±7	83.8	78.8
365±6	125±7	138.0	129.1

Table 9.2: Experimental measurement of ΔG_{NN}^0, $\Delta G_{NI_1}^0$, $\Delta G_{NI_2}^0$ and ΔG_{NU}^0 for hairpin I2. Numerical estimates for $\log\left(\phi_F^{N\to B}/\phi_R^{N\leftarrow B}\right)$ ($B=N,\ I_1,\ I_2\ ,\ U$), the free energy ΔG_{NB} obtained using both the EFR and the CFR, and the reversible work differences $\Delta W_{NB}^{\rm st}$ and $\Delta W_{NB}^{\rm bead+handles}$ (Eqs. 8.16 and 8.19c) obtained by integrating the FDC between $\lambda_0=0$ and $\lambda_1=183$ nm are provided. In addition, the free energy of formation of each states experimentally determined using the EFR, and theoretical predictions obtained using the NN model with the UO or UNZ set of base-pair free energies (Table 4.1) are provided. Error bars contain statistical and systematic errors. All the magnitudes are given in $k_{\rm B}T$ units.

$(k_{\rm B}T)$	ΔG_{NB} (EJE)	ΔG_{NB} (EFR)	Bias
$B=N$	543±3	543±3	0±4
$B=I_1$	545±1	527±1	18±2
$B=I_2$	539±1	526±1	13±2
$B=U$	547±1	528±1	19±2

Table 9.3: Estimation of bias for I2. Free energy difference ΔG_{NB} obtained at $\lambda_0=0$ nm and $\lambda_1=183$ nm from the EFR and the EJE, and estimation of the bias, for states $N,\ I_1,\ I_2$ and U.

9.4.2 Misfolded states

Molecular misfolded states behave as kinetic states off-pathway under non-equilibrium pulling experiments. When the force applied to the ends of an unfolded-stretched molecule decreases, different folding pathways may take place in presence of dynamically accessible misfolded states. The formation of those molecular states competes with the formation of the native conformation of the molecule. In non-equilibrium pulling experiments, the presence of misfolded states is revealed through the existence of different unfolding and folding patterns in the FDC. In some occasions, misfolded states can be identified and thermodynamically characterized under equilibrium conditions. However, this is not always the case. The use of the EFR opens the possibility to determine the thermodynamic stability of misfolded states that can only be generated dynamically, that is, that can only be observed and identified under non-equilibrium conditions.

In what follows, it is shown how to extract the free-energy of formation and the free-energy branches of misfolded states for two different DNA hairpin sequences.

Hairpin with one misfolded state (M1)

Hairpin M1 can fold into two unrelated structures. The hereafter referred to as native N structure consists of a single-stem hairpin, whereas the misfolded state M is made of two hairpins serially connected by seven T (Figs. 9.10a-b, respectively). The molecular FEL of both states N and M (Fig. 9.10c-d) at the coexistence force show an intermediate state located at $n = 13$ and $n = 10$, respectively. For N, this intermediate (I_N) is surrounded by a very low kinetic barrier to the folded state ($\sim 1\ k_{\rm B}T$). For M, the intermediate state (I_M) might have a relevant kinetic role since it is surrounded by high kinetic barriers ($\sim 7\ k_{\rm B}T$). Configurations that mostly contribute to I_M are those where one of the hairpins is unzipped and the other is folded.

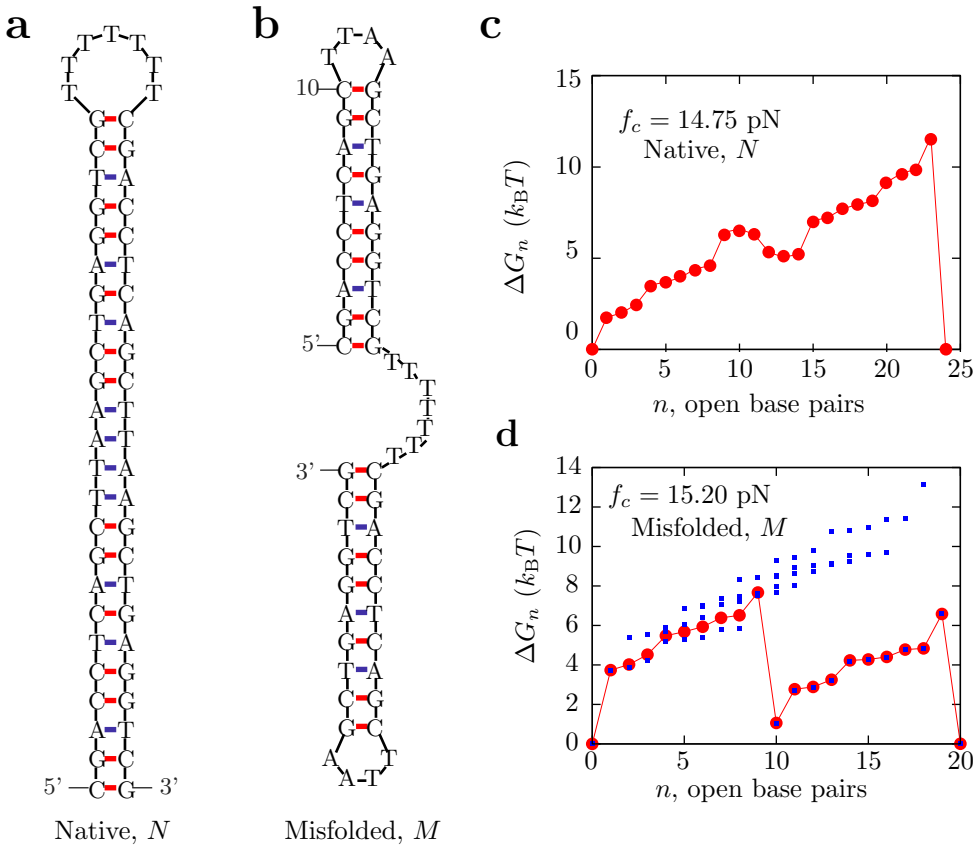

Figure 9.10: Hairpin M1, with one misfolded state. Sequence and structure of state N (**a**) and M (**b**). FEL of N (**c**) and M (**d**) evaluated at their respective coexistence forces. In panel d, blue squares are the free energies of all the possible sequential configurations of the structure that have a given number n of unzipped base pairs distributed between the two hairpins. Red circles are the potential of mean force, calculated as the exponential Boltzmann average over configurations constrained by a given number of unzipped base pairs n.

In non-equilibrium experiments, two different FDC patterns are identified corresponding to the unfolding and folding of structures N and M according to the predicted

structures, since M is larger due to the linker made of seven T (Fig. 9.11a). At low forces, M and N are distinguished because M has a larger molecular extension (inset in Fig. 9.11a). In contrast to the unfolding/folding cycles that start and end in N, those that start and end in M show almost no hysteresis, indicating low kinetic barriers between M and U. Moreover, for this hairpin M has a larger basin of attraction than N during folding, since 80% of folding trajectories end in M. This phenomenon is attributed to the kinetic competition of loop formation between M and N.

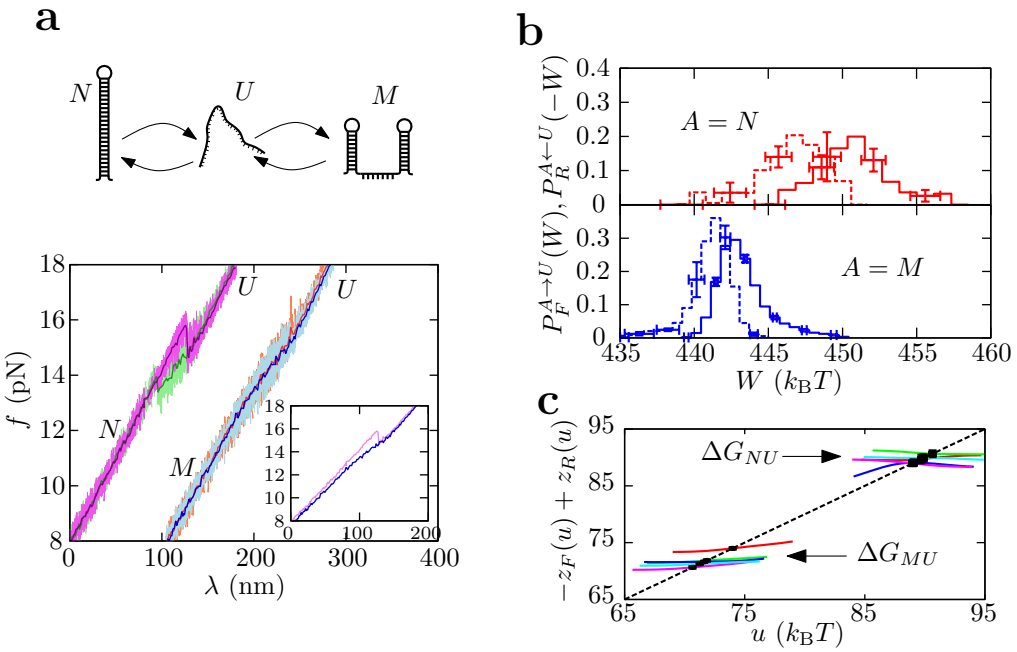

Figure 9.11: Pulling experiments and EFR with hairpin M1. a. Schematic illustration of the mechanical unfolding/folding pathways (top) and patterns identified in the FDC (bottom). For the sake of clarity curves are shifted. Left: unfolding (purple) and folding (green) FDC for the N state show a force rip ($\simeq$ 0.5 pN) around 15 pN. Right: unfolding (red) and folding (blue) curves are the FDC for the M state. Inset: Unfolding curves for N and M without shift (data filtered). At low forces M has larger extension than N, as expected from Fig. 9.10a-b. **b.** Partial work distributions for work values measured at 125 nm/s between $\lambda_0 = 0$ nm, where $A = N$ (red, top panel) or M (blue, bottom panel) and λ_1=130 nm, where $B = U$. Dark colors refer to unfolding and light colors to folding work distributions. **c.** Bennett acceptance ratio method used to extract ΔG_{NU} and ΔG_{MU}. Different colors are results obtained for different molecules pulled at different speeds.

The classification of forward trajectories into type $N \to U$ or $M \to U$ and of reversed trajectories into type $N \leftarrow U$ or $M \leftarrow U$ is straightforward from the pattern of the FDCs. However, kinetics are very fast and even though there is a hint of the intermediate states I_N and I_M being transiently visited by the molecule in both unfolding and folding processes, it is very challenging to identify which is the molecular state of the hairpin for intermediate values of λ along pulling experiments. For the sake of simplicity, neither I_N or I_M will be characterized using the EFR and the free-energy branches will not be recovered for any of the relevant states of hairpin M1.

The following steps are performed in order to get the free-energy of formation of N and M, ΔG^0_{NU} and ΔG^0_{MU} respectively, from the EFR. First, the classification of trajectories is carried out. For M1, both $\phi_F^{N\to U}$ and $\phi_F^{M\to U}$ are equal to 1 as all forward trajectories $A \to U$ ($A = N$ or M) end at U independently of the initial state. $\phi^{A\leftarrow U}$ are calculated as the ratio between the number of reversed trajectories of type $A \leftarrow U$ and the total number of measured reversed trajectories. Therefore, the prefactor $\phi_F^{A\to U}/\phi_R^{A\leftarrow U}$ is equal to $1/\phi_R^{A\leftarrow U}$, for both $A = N$ or M. Next, in order to get the work performed along each trajectory, FDCs are integrated over $\lambda_0 = 0$ nm and $\lambda_1 = 130$ nm. The resulting partial work distributions are shown in Fig. 9.11b. The free-energy differences between state U at λ_1=130 nm and N or M at λ_0=0 nm, ΔG_{NU} and ΔG_{MU} respectively, can be recovered from the application of the EFR. It can be seen in Table 9.4 that neglecting the prefactor $\phi_F^{A\to U}/\phi_R^{A\leftarrow U}$ from the estimation of the free energy ΔG_{AU} implies an error of 4 $k_{\rm B}T$. In Fig. 9.11c there are shown the results of the Bennett acceptance ratio method applied to different sets of data (each color represents a different molecule). The lower set of measurements corresponds to ΔG_{MU}, whereas the upper set to ΔG_{NU}.

The free energy of formation of each structure is finally extracted by subtracting the energetic contributions of the handles, the bead captured in the optical trap, and the released ssDNA for each conformation to the free-energy value determined using the EFR (Section 8.4). This gives $\Delta G^0_{MU} = 47 \pm 2$ $k_{\rm B}T$ and $\Delta G^0_{NU} = 62 \pm 3$ $k_{\rm B}T$. The difference between both free energies is in agreement with predictions based on the NN model (Table 9.4).

The free energies of formation of both M and U (relative to U) can also be obtained from equilibrium experiments (Appendix J), using the detailed balance condition or using Hidden Markov models. Both methods give identical results, which are also in good agreement with theoretical predictions obtained from the NN model (Table 9.4).

($k_{\rm B}T$)	$\log\left(\phi_F^{A\to U}/\phi_R^{A\leftarrow U}\right)$	ΔG_{AU} (EFR)	ΔG_{AU} (CFR)	$\Delta W^{\rm st}_{AU}$
$A = N$	2.2± 0.2	464±2	466±2	22.4±0.5
$A = M$	0.1± 0.1	459±2	460±1	21.0±0.5

$\Delta W^{\rm bead+handles}_{AU}$	ΔG^0_{AU} (EFR)	ΔG^0_{AU} (eq.)	ΔG^0_{AU} (UO)	ΔG^0_{AU} (UNZ)
381±3	61±3	58±3	60.2	57.2
392±3	47±3	46±3	49.6	46.9

Table 9.4: Experimental measurement of ΔG^0_{NU}, ΔG^0_{MU} for hairpin M1. For each state, it is shown the contribution to the free energy introduced by the prefactor $\log\left(\phi_F^{A\to B}/\phi_R^{A\leftarrow B}\right)$, the free energy ΔG_{AU} obtained using the EFR and the CFR, the reversible work $\Delta W^{\rm st}_{AU}$ needed to stretch the ssDNA and orient of the hairpin stem. and the reversible work $\Delta W^{\rm bead+handles}_{AU}$ performed to stretch the handles and to displace the bead in the optical trap. Furthermore, estimations of the free energy of formation ΔG^0_{AB} determined using the EFR, from equilibrium experiments, and from theoretical predictions obtained with the NN model and different sets of base-pair free energies (Table 4.1) are provided. These numbers have been extracted from the set of work values obtained by integrating the FDC between $\lambda_0 = 0$ nm and $\lambda_1 = 130$ nm. Error bars contain statistical and systematic errors. All the magnitudes are given in $k_{\rm B}T$ units.

Hairpin with two misfolded states (M2)

Hairpin M2 can fold into one native structure N, and two misfolded structures M_1 and M_2 following several pathways (Fig. 9.12). N consists of a single-stem hairpin; M_1 is made of four small hairpins that do not share any common base pair with N; and M_2 is made of three hairpins, one having 21 base pairs in common with N.

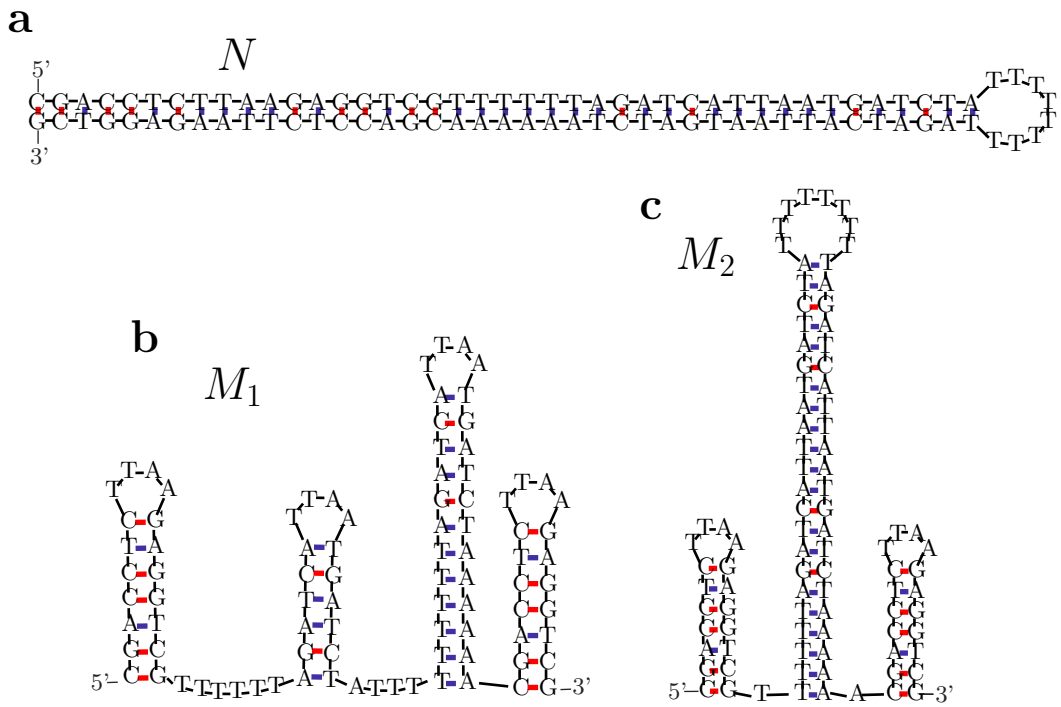

Figure 9.12: Hairpin M2, with two misfolded states. Sequence and structure for N (**a**), M_1 (**b**) and M_2 (**c**).

The FEL of each folded structure is shown in Fig. 9.13. Structure N has a typical two-state FEL profile. Since states M_1 and M_2 are made of several hairpins, all sequential configurations are represented (blue dots), and the potential of mean force obtained from the Boltzmann average over configurations restricted to the number of open base pairs n is shown in red circles. In both cases several intermediates are revealed, with identical end-to-end molecular extensions. Therefore, in this case it is difficult to get accurate information about the kinetics of the molecule from equilibrium experiments, as shown in the experimental trace obtained from a hopping-based experiment (Fig. 9.14), where it is unfeasible to identify different states.

In non-equilibrium pulling experiments it is easy to identify trajectories where hairpin M2 unfolds from and folds into N, *i. e.* trajectories of type $N \to U$ or $N \leftarrow U$, since these show a jump in force characteristic of a two-state DNA hairpin (red/blue FDC in Fig. 9.15, left). On the other hand, all folding trajectories in which the molecule misfolds look practically the same making it impossible to distinguish between M_1 and M_2 at low forces.

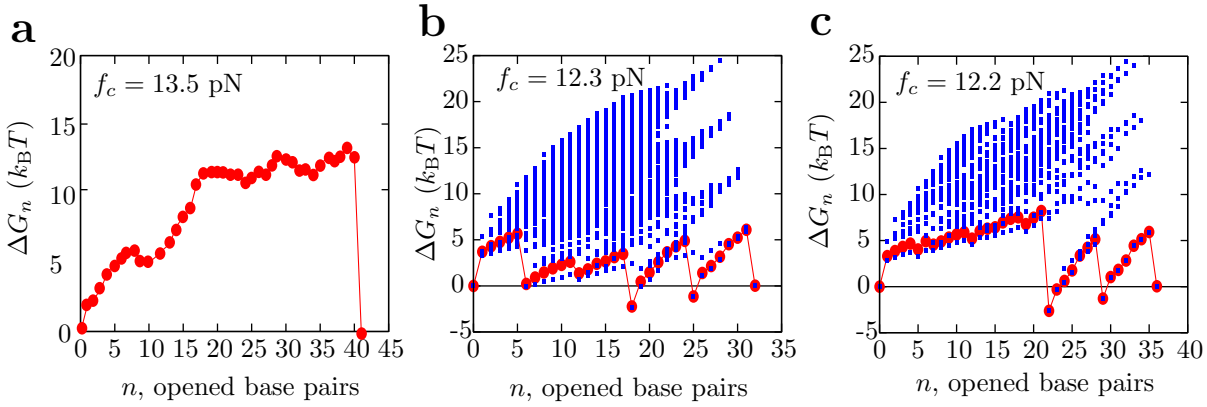

Figure 9.13: Free-energy landscapes for the most stable structures of hairpin M2. FEL calculated at the coexistence force for N (**a**), M_1 (**b**) and M_2 (**c**).

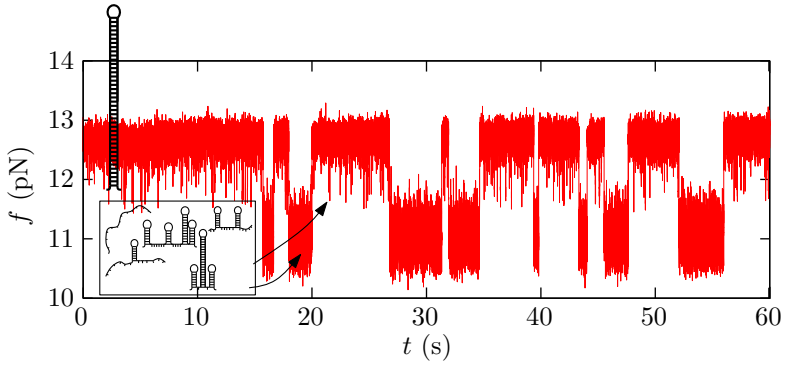

Figure 9.14: Hopping-based equilibrium experiment. At an intermediate range of forces (10-13pN) the hairpin shows hopping between N (mean force 12.7 pN), an intermediate state (mean force 11.7 pN) and a mixture several states (mean force 11pN). The latter mixture is potentially composed of$U$, M_1, M_2 and other intermediates (boxed region that exhibit nearly the same molecular extension being very difficult to distinguish.

Careful inspection reveals two different patterns of unfolding curves that start from a misfolded state: either the molecule unfolds quasi-reversibly (purple FDC in Fig. 9.15 middle, ∼30% of trajectories), or it folds back to N before it unfolds (cyan FDC in Fig. 9.15 right, ∼20% of trajectories) [Li 07]. It is assumed that unfolding curves starting from a misfolded state and showing rescue to N around 9.5 pN (Fig. 9.15 right) are of type $M_2 \rightarrow U$. Furthermore, folding curves preceding unfolding curves that are rescued by N are assigned to fold into M_2, and they are classified as $M_2 \leftarrow U$. Consequently, unfolding curves that do not show any rescue and their preceding folding trajectories are assigned to be of the type $M_1 \rightarrow U$ and $M_1 \leftarrow U$ respectively. This is supported by two facts. First, M_1 consists of four small hairpins that confer low mechanical stability to the structure and, as a result, it apparently reversibly unfolds under tension. Second, M_2 has a large stem in common with N (Fig. 9.12a-c) which is surrounded by two small hairpins with low mechanical stability. Once these two hairpins unfold around 9-10 pN,

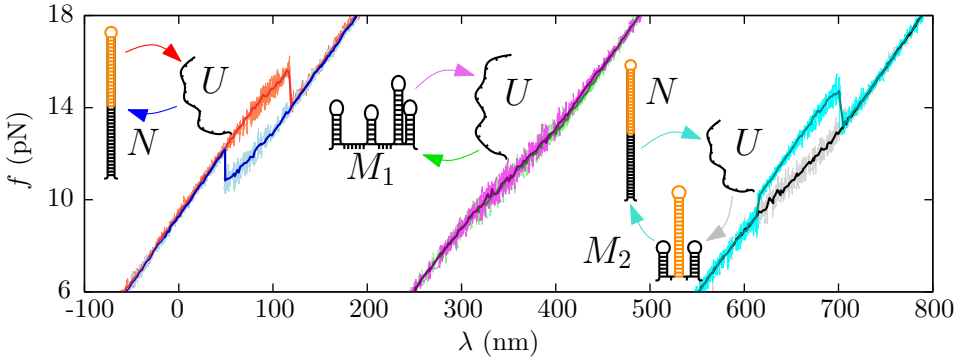

Figure 9.15: Pulling experiments with hairpin M2. FDC with their corresponding unfolding/folding pathways. Left: unfolding (red) and folding (blue) FDCs for state N show a force rip ($\simeq$2 pN) around 15 pN. Middle: unfolding (purple) and folding (green) FDCs for state M_1 show no hysteresis. Right: folding (gray) FDCs for state M_2 are identical to the ones measured for M_1, whereas unfolding (cyan) FDCs show a rescue to the N state.

force remains low enough for the molecule to fold back to N before complete unfolding.

A crucial question to justify this assumption is how far from equilibrium the system is: under quasi-static conditions (slow pulling speeds), the molecule has more time to explore the FEL and to overcome high kinetic barriers, preferentially folding into N. On the other hand, under far from equilibrium conditions (fast pulling speeds), the molecule will mostly fold into M_1 because its folding pathway encounters lower kinetic barriers (Fig. 9.13b). However, at an intermediate value of the pulling speed the molecule will occasionally fold into M_2. Subsequent unfolding events from M_2 at intermediate pulling speeds will show rescue to N. To conclude, it is crucial to be far from of equilibrium to favor misfolded states, but not too far. Otherwise, rescue to N would not be observed and those trajectories might be erroneously attributed to M_1.

Free-energy branches for states N, M_1, M_2 and U of hairpin M2 are recovered. Results are not completely accurate because the molecule visits several intermediate states along unfolding and folding experiments, as revealed by the FEL of each structure. However, they are valid from a qualitative point of view. In this case, λ_1 is set at 230 nm and λ_0 varies between 0 and 150 nm. For each value of λ_0, trajectories are classified according to the initial state, as the final state is always U. Hence, forward trajectories may be of type $A \to U$ and reversed trajectories may be of type $A \leftarrow U$, where $A = N$, M_1, M_2 or U. As for hairpin M1, the fraction $\phi_F^{A\to U}$ is always 1, whereas $\phi_R^{A\leftarrow U}$ is calculated as the ratio between the number of reversed trajectories of type $A \leftarrow U$ and the total number of reversed trajectories. Then, the work is calculated as the area below the FDC between λ_0 and λ_1. The resulting partial work distributions, obtained at $\lambda_0 = 0$ nm, are shown in Fig. 9.16a. Note that at this value of λ_0, state U is never observed. Finally, the EFR is applied in order to get the free energies ΔG_{NU}, ΔG_{M_1U}, ΔG_{M_2U} and ΔG_{UU}. In Table 9.5 it can be seen that neglecting the prefactor $\phi_F^{A\to U}/\phi_R^{A\leftarrow U}$ in the determination of the free energy ΔG_{AU} leads to wrong estimated values up to 3 $k_{\mathrm{B}}T$. Figure 9.16b shows the reconstruction of the four thermodynamic branches relative to the

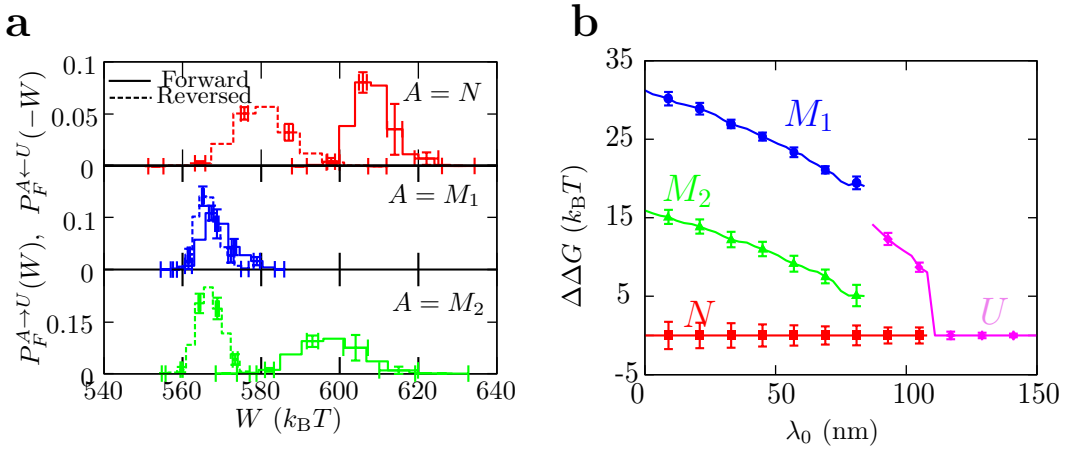

Figure 9.16: Partial work distributions and free-energy branches for M2. a. Partial work histograms for work values measured between λ_0=0, where $A = N$ (red, top panel), M_1 (blue, middle panel) or M_2 (green, bottom panel), and λ_1=230 nm, where $B = U$. Solid lines are forward work distributions whereas dashed lines correspond to reversed work distributions. **b.** Free energy branches of states N (red squares), M_1 (blue circles), M_2 (green triangles) and U (purple diamonds) measured relative to the full free energy of the system, $\Delta G = -k_BT \log \sum_{A=N,M_1,M_2,U} e^{-\Delta G_{AU}/k_BT}$.

full equilibrium free energy of the system, $\Delta G = -k_BT \log \sum_{A=\mathrm{N,M_1,M_2,U}} e^{-\Delta G_{AU}/k_BT}$. For this molecule, the most stable state is either N or U, depending on the value of λ_0. Moreover, M_2 is more stable than M_1, as expected from its structure (one long hairpin plus two small hairpins is expected in general to be energetically more favorable than four small hairpins).

Finally, from the values of ΔG_{NU}, ΔG_{M_1U}, ΔG_{M_2U} obtained using the EFR at $\lambda_0 = 0$ nm and $\lambda_1 = 230$ nm, the energetic contributions from the handles, the bead in the optical trap, and the released ssDNA can be subtracted in order to get the free energy of formation of each configurations. Results, summarized in table 9.5 are again in agreement with theoretical predictions.

9.5 Conclusions

Summarizing, it has been shown how the EFR can be used to extract free energies of non-equilibrium kinetic structures in DNA hairpins exhibiting intermediate and misfolded states. The method accurately works in far from equilibrium situations and when equilibrium experiments are insufficient to characterize non-native states. There are two main differences between the EFR in Eq. (9.3) and the Crooks relation [Croo 00]: (1) the partial work distributions and (2) the prefactor $\phi_F^{A\to B}/\phi_R^{A\leftarrow B}$, which introduces the additional correction $-k_BT \log(\phi_F^{A\to B}/\phi_R^{A\leftarrow B})$ into the Crooks estimation of the free energy difference between kinetic states. Despite the fact that the correction is generally small, its omission yields wrong relative thermodynamic stabilities for the free energy branches of the different kinetic states [Juni 09]. Moreover, for the case of misfolded structures that apparently unfold/misfold reversibly, ΔG_{MU} is not just equal to the

	$(k_{\mathrm{B}}T)$	$\log\left(\phi_F^{A\to U}/\phi_R^{A\leftarrow U}\right)$	ΔG_{AU} (EFR)	ΔG_{AU} (CFR)	$\Delta W_{AU}^{\mathrm{st}}$
$A=$	N	0.7± 0.2	612±1	613±1	37.5±0.1
$A=$	M_1	1.8± 0.1	579±1	581±1	31.5±0.1
$A=$	M_2	1.1± 0.2	590±1	592±1	35.5±0.1

$\Delta W_{AU}^{\mathrm{bead+handles}}$	ΔG_{AU}^0 (EFR)	ΔG_{AU}^0 (UO)	ΔG_{AU}^0 (UNZ)
480±2	95±3	92.9	87.4
487±3	60±3	62.0	57.4
485±3	69±3	72.3	67.9

Table 9.5: Experimental measurement of ΔG_{NU}^0, $\Delta G_{M_1U}^0$ and $\Delta G_{M_2U}^0$ for hairpin M2. For each state, it is shown the contribution to the free energy introduced by the prefactor $\log\left(\phi_F^{A\to B}/\phi_R^{A\leftarrow B}\right)$, the free energy ΔG_{AU} obtained using the EFR and the CFR, the reversible work $\Delta W_{AU}^{\mathrm{st}}$ needed to stretch the ssDNA and orient of the hairpin stem. and the reversible work $\Delta W_{AU}^{\mathrm{bead+handles}}$ performed to stretch the handles and to displace the bead in the optical trap. Furthermore, estimations of the free energy of formation ΔG_{AB}^0 determined using the EFR, and from theoretical predictions obtained with the NN model and different sets of base-pair free energies (Table 4.1) are provided. These numbers have been extracted from the set of work values obtained by integrating the FDC between $\lambda_0 = 0$ nm and $\lambda_1 = 130$ nm. Error bars contain statistical and systematic errors. All the magnitudes are given in $k_{\mathrm{B}}T$ units.

measured reversible work during unfolding since the term $k_{\mathrm{B}}T\log\phi_R^{M\leftarrow U}$ must be added (here $\phi_F^{M\to U}=1$ as forward processes always end at U). Although this correction is small for states M (hairpin M1) and M_2 (hairpin M2) ($\sim 0.2\ k_{\mathrm{B}}T$ and $1.2\ k_{\mathrm{B}}T$ respectively), it is important in situations where $\phi_R^{M\leftarrow U}\ll 1$, even if very low hysteresis is obtained between the forward and reversed processes. For example, in cases where misfolding occurs with a probability of 1% (that is, $\phi_F^{M\to U}=1$ and $\phi_R^{M\leftarrow U}=0.01$), the exclusion of the prefactor $k_{\mathrm{B}}T\log\phi_R^{M\leftarrow U}$ would underestimate by $4.5\ k_{\mathrm{B}}T$ the estimation of the free energy of formation of the misfolded structure using fluctuation relations.

The main limitation of the method is the identification of kinetic states from the measured signal. In this regard, a combination of fluorescence techniques, such as FRET, with force measurements, and the application of advanced statistical methods (e.g. hidden Markov models or Bayesian inference) might be very useful. Our methodology should find many applications that range from molecular biophysics to condensed matter physics. Any situation where equilibrium experiments are unpractical should be treatable with different versions of Eq. (9.3). To start with, the method can be employed for measuring free energies of kinetic structures that appear in many molecular reactions, such as RNA, proteins, and many kinetic states related to intermolecular binding.

Part IV

Conclusions and future perspectives

Conclusions

Codification of genetic information; regulation of gene expression; transport of nutrients inside cells; immune protection against infectious agents; transduction of external signals... These are some of the crucial processes that take place in living organisms at the molecular level. The more we learn about how biomolecules perform their assigned biological task, the more we are surprised by their high degree of accuracy and efficiency. A deep understanding of how these phenomena occur is vital from at least two different point of views. (1) From the biological perspective, the origin of many diseases could be unraveled. For instance, neurodegenerative disorders such as Alzheimer are thought to have an origin on the misfolding and oligomerization of otherwise soluble proteins. Hence, the study of the mechanisms of protein folding can be decisive to not only discover the origin, but also prevent and even find cures for many illnesses. (2) From the view of theoretical physics, a precise knowledge of the laws governing the microscopic world is interesting in itself. Additionally, it could also boost a technological revolution with designs of extremely efficient microscopic motors or a DNA-based computation techniques.

Single-molecule experiments emerge as a powerful and versatile tool that allows us to investigate molecular systems at the level of individual molecules with unprecedented spatial and temporal resolution. This has several advantages, one of the most important being the possibility to observe intermediate and metastable molecular states, in both equilibrium and non-equilibrium situations. In contrast, in traditional bulk techniques experimental results are the average over all the molecules (in the order of the Avogadro's number) and transient states are very difficult to observe. Another limitation of bulk techniques was found in this thesis when attempting to measure the free energy of formation in presence of monovalent ions of a small RNA hairpin using UV absorbance, since the sample started boiling and evaporating before any relevant signal could be obtained (Appendix F). The same measurement can be easily carried out using single-molecule techniques such as optical tweezers (OT) (Section 5.2).

In this thesis, OT are used to perform single-molecule experiments. This instrument makes it possible to manipulate with nanometric precision a biomolecule and exert forces on it in the range of [0-100] pN. The diversity of systems being studied using optical tweezers increases every day. In the particular case of this thesis, OT are used to unravel the mechanisms of unfolding and folding of several small nucleic acid hairpins and a protein when a force is applied to their ends. Elastic, kinetic and thermodynamic properties are accessed mostly using non-equilibrium pulling experiments. Moreover, single bonds made of an antigen and an antibody are investigated by qualitatively measuring the correlation between bond affinity and bond elasticity.

All these experiments have in common the relevant role played by thermal fluctuations. In any of the cases, measured distances and applied forces are in the range of nanometers and picoNewtons, respectively. Hence, the energies explored at room temperature are of the order $E \sim \text{nm} \times \text{pN} \sim 0.25\ k_{\text{B}}T$, *i. e.* of the order of thermal fluctuations, and consequently independent repetitions of an identical experimental pro-

tocol may lead to different outputs. Therefore, independent trajectories differ either in the value of the force at which the molecular system undergoes a particular transition (unfolding, folding or rupture forces), or even in the followed reaction pathway (folding towards the native or the misfolded conformation).

The study of the characteristic forces at which molecular transitions occur, such as molecular unfolding and folding, is known as dynamic force spectroscopy (DFS). In this thesis DFS combined with Markov models are widely used to characterize the unfolding/folding reaction pathway, the transition states present in the molecular free-energy landscape (FEL), and the elastic, kinetic and thermodynamic properties of small nucleic acid hairpins under different conditions (Chapters 4 and 5). A general result is that non-equilibrium DFS methods provide an excellent platform to extract thermodynamic properties of molecular states that can only be observed under dynamical conditions (that is, they are never observed in equilibrium at reasonable time scales), such as intermediate or bounded configurations. Next, non-equilibrium DFS techniques are also applied to characterize the correlation between bond affinity and bond elasticity in the antibody-antigen interactions (Chapter 6). Results here are qualitative but appealing, since they suggest that during the maturation line of the immune systems antibodies evolve in order to establish stiffer bonds with the antigens as the recognition becomes stronger and more specific. To end, the power of both equilibrium and non-equilibrium DFS techniques combined with Markov models is shown to characterize the FEL and the thermodynamic and kinetic properties of protein Barnase (Chapter 7). Remarkably, our results are in good agreement with previous bulk observations despite the hysteresis effects observed between unfolding and folding processes in DFS experiments.

An alternative method to extract thermodynamic information is the use of fluctuation relations (FR). These are mathematical identities that allow us to obtain free-energy differences from non-equilibrium work measurements. As it was the case for the values of characteristic forces, work values may also significantly differ in each independent repetition of an experimental protocol due to the effect of thermal fluctuations. FR use the whole distribution of non-equilibrium work measurements obtained in forward and reversed non-equilibrium protocols between an initial equilibrated state and a final state (that is not necessarily equilibrated at the end of the each protocol) and determine the free-energy difference between the two states (Chapter 8). In this thesis, an extension of FR is used that no longer requires the system to be initially at equilibrium for any of the protocols, but only requires the system to be partially equilibrated (Chapter 9). This new condition makes it possible to measure free-energy differences between states that are hardly observed at equilibrium, and also allows us to reconstruct the free-energy branches of molecular states along a given experimental protocol.

In sum, in this thesis single-molecule experiments with OT are performed in order to extract information of the molecular FEL of different molecular systems. To this end, both DFS and FR are powerful to obtain equilibrium information, such as free-energy differences, from non-equilibrium measurements. These results are promising since some molecular states can only be generated dynamically and because most processes that take place in living systems occur under non-equilibrium conditions.

Future perspectives

OT are an excellent single-molecule technique to investigate the behavior of molecular systems with an great accuracy. In this thesis, the molecular FEL of several molecular systems (nucleic acid hairpins, antigen-antibody bonds, protein) is characterized using both non-equilibrium DFS techniques combined with Markov models and an extended version of FR. Results presented open the door to the quantitative study of complex molecular processes that take place in living organisms (and hence under non-equilibrium conditions).

Combination of OT with other experimental techniques would be greatly beneficial. For instance, the inclusion of a temperature controller in the mini-tweezers setup would allow us to determine enthalpic and entropic barriers, which play an important role in many biophysical processes. In a second example, a combination of fluorescence techniques, such as FRET, with OT might be very useful to identify kinetic states from measured signals.

With regard to DFS techniques, systems investigated always show cooperative transitions between molecular states (native, unfolded, intermediates, bounded, unbounded) at characteristic ranges of force values. It would be interesting to extend this kind of analysis to address situations where transitions are not cooperative, as it is the case for misfolding events. Hence, kinetic and thermodynamic properties of misfolded states could be also unraveled using DFS and Markov models. Complex situations where native, intermediate, misfolded and even bounded states occur along a non-equilibrium process could be then investigated using DFS experiments.

On the other hand, the extension of FR paves the way to characterize kinetic molecular states (*i. e.*, states that are dynamically observed). An example of these is given by bound states. Therefore, it will be possible to characterize the thermodynamic properties of chemical reactions of binding/unbinding (that is, binding free energies and equilibrium constants). Additionally, the scientific community has shown a growing interest in the last few years in the understanding of the conversion between free energy and information. In this context, protocols with feedback mechanisms have been proposed in order to find efficient processes where the minimum amount of work is performed. New FR have been derived that include the cost in free energy of such feedback mechanisms and, interestingly, these are formally equivalent to the extended FR. The realization of a Szilard-based engine turns out to be feasible using two-state nucleic acid hairpins pulled with OT, and the extended FR can provide a novel theoretical framework to investigate the conversion between free energy and information.

To end, the study of the variability of unfolding, folding, misfolding and binding kinetic and thermodynamic properties over different members of an ensemble of mutants could be performed to carry out molecular evolution studies.

Part V

Appendices

Appendix A

Synthesis of molecular constructs

A.1 DNA hairpins with short handles

Following, the steps to synthesize DNA hairpins with short 29-basepairs long dsDNA handles are summarized. Details of the synthesis are also provided in reference [Forn 11]. Figure A.1 sketches how does the molecular construct look like. The short handles consist of a 29-basepairs long dsDNA chain. Such handles are the same for any DNA hairpin. Moreover, the sequence of both handles, right and left, is the same: The oligonucleotide

Left & right handle: 5' - AGT TAG TGG TGG AAA CAC AGT GCC AGC GC - 3'

that makes up one strand of the handle double helix is hereafter denominated as "Splint". The sequence of the Splint is:

Splint: 5' - GCG CTG GCA CTG TGT TTC CAC CAC TAA CT - 3'

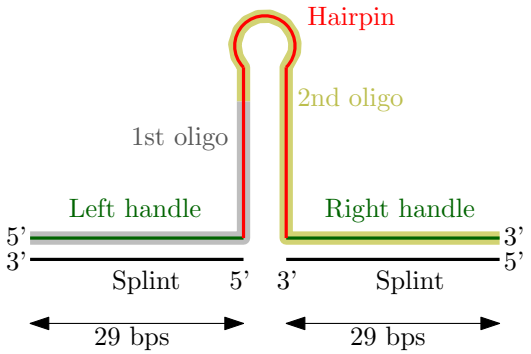

Figure A.1: Sketch of the molecular construct made of a DNA hairpin inserted between two short handles.

The DNA hairpins with short handles are synthesized using the hybridization of three different oligonucleotides. The sequences used to synthesize the hairpins used in

this thesis are provided in Table A.1. These oligonucleotides are supplied by a company (for example *Sigma-Aldrich* or *Invitrogen*), and consist of:

1. The sequence of the ssDNA left handles plus a part of the sequence of the desired DNA hairpin. This oligonucleotide has a biotin at its 5' end (gray-thick line in Fig. A.1).

2. The rest of the sequence and the ssDNA right handle (yellow-thick line in Fig. A.1). This oligonucleotide is modified at its 3' end with a digoxigenin tail (see below).

3. The Splint oligonucleotide, that will hybridize to the right and the left ssDNA handles contained in the two previous oligonucleotides.

The steps of the synthesis are:

1. Digoxigenin tailing of 2nd oligonucleotide

Using the *Oligonucleotide Tailing Kit (Roche)*, a tail of an average of 50 nucleotides randomly is added to the 3' end of an oligonucleotide. Some of these nucleotides are Digoxigenin-dUTP. The steps of the digoxigenin tailing are:

1. Spin down the 2nd oligonucleotide tube (part of the hairpin plus right handle). If it is not dissolved, dissolve the oligonucleotide with sterile distilled water at 100 μM. Mix.

2. Mix the following components in a sterile eppendorf tube in the following order:

8 μl sterile distilled water
1 μl oligonucleotide 100 M (100 pmol)
4 μl Reaction Buffer x 5
4 μl $CoCl_2$
1 μl dATP
1 μl Dig-dUTP
1 μl Terminal transferase (enzyme)
20 μl

3. Incubate during 15 min at 37ºC.

4. Add 2 μl of 2 M EDTA PH 8.0.

5. Oligonucleotide purification using the *Qiaquick Nucleotide Purification Kit (QUIAGEN)*. Follow the instructions of the kit.

6. 50 μl of the modified oligonucleotide (at 2 pmol/μl if no oligonucleotide is lost) is obtained. Keep the Kit and the oligonucleotide at -20ºC.

2. Annealing and ligation of the three oligonucleotides

At this step you want to mix the three different oligonucleotides so they can hybridize to obtain the correct construction (Figure A.1).

1. Spin down the oligonucleotide tubes (if not dissolved yet, dissolve the 1st oligonucleotide with sterile distilled water at 100 μM).

2. Do a dilution of the 1st and the 3rd oligonucleotide at 5 pmol/μl (for example, mix 95 μl of sterile distilled water with 5 μl of the oligonucleotide 100 μM.

3. Mix at two different eppendorf tubes:

1 μl	1st oligonucleotide 5 pmol/μl ($\simeq$ 5 pmol)
1 μl	Splint 5 pmol/μl ($\simeq$ 5 pmol)
0.5 μl	Tris 1 M pH7.5
0.5 μl	NaCl 5 M
12 μl	sterile distilled water
15 μl	

5 μl	2nd oligonucleotide, with digoxigenin tail ($\simeq$ 5 pmol)
1 μl	Splint 5 pmol/μl ($\simeq$ 5 pmol)
0.5 μl	Tris 1 M pH7.5
0.5 μl	NaCl 5 M
8 μl	sterile distilled water
15 μl	

4. Incubate for 4 hours at 42℃.

5. Mix the two annealing mixtures.

6. Dialysis of the mixture for 30 min in 50 ml of 5 mM NaCl, 20 mM Tris pH 7.5. Recover the maximum amount of the annealing mixture, hereafter refereed to as $\mathcal{X}$.

7. Mix the following components. These products are kept at -20℃. Melt them before starting with the exception of the enzyme.

$\mathcal{X}$ μl	DNA annealing mixture
2 μl	ATP 10 mM
3 μl	Ligation Buffer x 10
2 μl	T4 DNA ligase (enzyme)
$23 - \mathcal{X}$ μl	sterile distilled water
30 μl	

8. Incubate overnight at 16℃.

9. Dialysis of the mixture for 30 min in 50 ml of 10 mM NaCl, 10 mM Tris pH 7.5, 1 mM EDTA. Recover the maximum amount of the ligation mixture. Keep the DNA at 4℃.

Streptavidin-coated polystyrene microspheres (1.87 μm; Spherotech, Libertyville, IL) and protein G microspheres (3.0-3.4 μm) coated with anti-digoxigenin polyclonal antibodies were used for specific attachments to the DNA molecular construct. Attachment to the anti-digoxigenin microspheres was achieved first by incubating the beads with the tether DNA. The second attachment was achieved in the fluidics chamber and was accomplished by bringing a trapped anti-digoxigenin and streptavidin microspheres close to each other.

A.2 RNA hairpins with long-hybrid handles

The RNA molecule was prepared as previously described in reference [Coll 05]. The main steps are:

1. Oligonucleotides CD4F and CD4R (Table A.2) were annealed and cloned into the pBR322 DNA plasmid (*GenBank J01749*) digested with EcoRI (position 4360) and HindIII (position 30). The annealed oligonucleotides contain the sequence that codes for a modified version of CD4-42F class I hairpin that targets the mRNA of the CD4 receptor of the human immunodeficiency virus [McMa 02].

2. Oligonucleotides T7F and T7R (Table A.2) were used as primers to amplify by PCR a product of 1201 bp from the recombinant clone containing the CD4 insert. This amplicon contains the T7 RNA Polymerase promoter at one end, and was used as a template to synthesize an RNA containing the RNA hairpin (20 bp stem sequence and tetraloop GAAA) and the RNA components of handles A (527 bp) and B (599 bp).

3. The DNA components of handles A and B were obtained by PCR from the pBR322 vector (positions 3836-1 for handle A and positions 31-629 for handle B). Handle A was 30 biotinylated while handle B was tagged with a 50 digoxigenin molecules. Hybridization reactions were performed in a formamide-based buffer with a step-cool temperature program: denaturation at 85℃ for 10 min, followed by 1.5 h incubation at 62℃, 1.5 h incubation at 52℃ and finished with a cooling to 10℃ within 10 min.

A.3 Antigen-antibody constructs

Carboxy-derivatized polystyrene beads of 3 μm and 2 μm diameter (*Kisker Biotech GmbH & Co.*, Germany) were used for conjugation of antibody and antigen, respectively.

Antigens consist of low-weight/size estrogens (methyl-boldenone, 17β-boldenone and testosterone) linked to bovine serum albumin (BSA). Hapten design was carried out preserving the chemical, electronic and conformational properties of estrogens, while introducing a spacer arm for the covalent coupling to the BSA protein (Fig. 6.4a). Haptens were linked to BSA using the active ester method as previously described [Este 05]. The use of BSA is mandatory to reduce electrostatic interactions between polystyrene beads.

The steps followed to conjugate biomolecules to polystyrene beads are:

1. Wash beads (250 μl/sample) twice with 1 ml of 10 mM NaH_2PO_4, pH 6 (buffer A).
2. Centrifugate at 10000 rpm for 5 min, and re-suspend beads in 1 m1 of buffer A.
3. Prepare in buffer A 2.5 ml of a 90 mM solution of N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide and 2.5 ml of a 40 mM solution of N-hydroxysuccinimide and add it to bead solution.
4. Stir the mixture for 20 min at room temperature.
5. Centrifugate activated beads at 4000 rpm for 5 min and washed once with 1 ml of 2 mM HCl.
6. Re-suspend beads in 1.5 ml of 20 mM Na_2HPO_4/NaH_2PO_4, pH 7.5 (buffer B).
7. In buffer B, add 100 μl of serial dilutions of antibody or antigen. (In order to avoid multiple binding, the amount of coating antigen/antibody on the beads was tested at different concentrations until most unbinding events showed single rupture events. 100 μg of antibody and 250 μg of coating antigen per reaction showed the best results.)
8. Leave reaction for 3 h at room temperature.
9. Wash conjugated beads three times with 1 ml of buffer B (4000 rpm for 5 min) and store in 250 μl of 20 mM Na_2HPO_4/NaH_2PO_4, 100 mM glycine, pH 7.5, with 3.3 mg/ml BSA and 0.02% NaN_3.

The coupling efficiency of the immunoreagents to the polystyrene beads was checked by measuring the protein content of the supernatant after bioconjugation. In all cases the bioconjugation yield was between 85 and 90%, indicating a similar antibody/antigen density on the different biofunctionalized particles.

A.4 Barnase with long dsDNA handles

Following, the steps to synthesize the protein Barnase an insert it between to 500-basepairs long dsDNA handles are summarized. Details of the synthesis are provided in reference [Cecc 11]. Figure A.2a sketches how does the molecular construct looks like. The protein is inserted between two 500-basepairs long 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 aminoacids of Barnase have been mutated to cysteins. On the other hand, 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 in order to attach the molecular setup to polystyrene beads. Furthermore, a tail of histidines is included at the C- terminal of barnase in order to simplify its purification.

The steps of the synthesis consist of expression and purification of the protein, the synthesis of the DNA handles, and the final reaction to get the molecular setup depicted in Fig. A.2. Buffer contents are summarized in Table A.3.

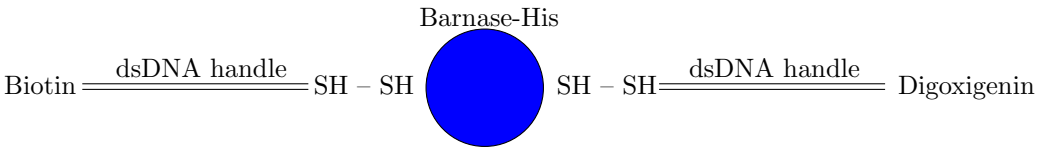

Figure A.2: Synthesis of Barnase. Sketch of the molecular setup made by inserting barnase between two dsDNA handles.

1. Barnase expression and purification

Bacterial growth

1. Prepare a pre-culture of BL21 [pET-Bar]: working on ice under sterile conditions, touch the BL21 pET-Bar glycerol from -80℃ with a toothpick and inoculate 10 ml of LB+100μM ampicillin media. Grow at 250 rpm, 37℃, overnight.

2. Inoculate 1 l of LB+100μM ampicillin media with 1 ml of pre-culture. Grow at 250 rpm, 37℃, until an O.D. of 0.6-0.8 is reached.

3. Induce with 0.5 mM IPTG. Let the bacterial cells grow and express the protein during 5h at 37℃.

4. Transfer the culture to 80 ml centrifuge tubes and centrifuge at 3000 rpm for 15 min at room temperature. Discard the supernatant (*i. e.*, the liquid lying above the solid residue) and repeat the centrifugation until all cells are collected.

Bacterial lysis

1. Resuspend the cells with the Wash-Equilibration Buffer.

2. Sonicate on ice.

3. Transfer the cell suspension to several 1.5 ml eppendorf tubes. Centrifuge with the microcentrifuge for 10min at maximum speed. Collect the supernatant (*i. e.*, the lysate containing the soluble over-expressed protein) and discard the cell debris.

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.

1. Load the lysate into the affinity column (follow instructions from the commercial product for its preparation).
2. Wash resin with Wash-Equilibration Buffer. At this point, Barnase-His and other histidine-rich proteins are attached to the resin.
3. Elute the protein from the column in 10 ml fractions of Elution buffer supplemented with 500 mM imidazole.
4. Analyze the fractions by SDS-PAGE and pool those containing the protein.
5. Concentrate the selected fractions with the protein until 200-400 μl.
6. Load the concentrated protein to a size exclusion. Carry out purification using 00 mM NaPi Buffer+1 mM DTT.
7. Analyze the resulting fraction with SDS-PAGE.
8. Concentrate the sample again reaching 2 ml or less. Add 20 mM DTT and wait 20 minutes.
9. Carry out a Bradford test to determine concentration.

Barnase-His Activation

1. Unfreeze a DTDP stock aliquot and add 8.5 ml of 0.1 M NaPi.
2. Add 200 μl of freshly prepared DTDP to 200 μl of protein. Leave the reaction overnight at room temperature in the dark.
3. Carry out a dialysis against 500 ml 100 mM NaPi buffer for 2 hours. Repeat 3 times.

2. Preparation of the handles

The two handles are 500 basepairs long and have identical sequence. 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 1 μmol and are resuspended in water at a final concentration of 100 mM. 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

1. Synthesize the Bio-SH 500 basepairs handle. Prepare a 9 ml PCR reaction mixing all the components in a 50 ml Falcon tub. Transfer 100 μl aliquots to PCR tubes and start amplification. Use the standard polymerase from *Biotools*.

2. Synthesize the Dig-SH handle following the protocol described in step 1.

3. Transfer the PCR fractions to a 50 ml Falcon tub. Check the PCR with a 1.5% agarose gel.

4. Use the Hispeed Plasmid Maxi Kit (*Qiagen, Ref. 12663*) starting at step 6 and follow instructions. Perform final elution with 1 ml of 15 mM NaPi and 3 mM DTT.

5. Measure DNA concentration by Abs260. It is important to obtain a concentration of at least 300 μg/ml.

3. Reaction between protein and handles

1. Mix equal amounts of Bio-SH and Dig-SH handles, typically 0.5 ml, trying to get a final concentration of 150μg/ml for each handle ($\sim$1 μM total DNA).

2. Concentrate handles mixture with Microcon to a final volume of 50-60 μl.

3. Equilibrate 3 Bio-Spin columns with 0.1 NaPi/EDTA (5 x 0.5 ml, 1 minute at 1000 g). Exchange the buffer by passing through 3 consecutive Bio-Spin columns.

4. Prepare reaction:

   4.1. Place $\sim$100 pmol of protein in the reaction eppendorf.

   4.2. Add 10 μl of 500 basepairs DNA handles (just after DTT removal). Reaction starts meanwhile final concentrations are adjusted.

   4.3. Determine the 500 basepairs DNA handles concentration by Abs260. Use a 1:50 dilution. Concentration should be around 20 μM.

   4.4. Adjust the concentration of the 500 basepairs handles and DTDP-protein to $\sim$12 μM and $\sim$4 μM respectively. The reaction ratio should be 4:1.

5. Let the reaction run overnight.

6. Analyze the reaction by 4% SDS-PAGE stained with silver nitrate to follow DNA.

	CD4 hairpin
1st oligo	**Left -** GCG AGC CAT AAT
2nd oligo	CTC ATC TGG AAA CAG ATG AGA TTA TGG CTC GC **- Right**
	M15 hairpin
1st oligo	**Left -** CCC GCA CGG CCG GCC ATT ATT AAA
2nd oligo	CAA TTT GAA AAA ATT GTT TAA TAA TGG CCG GCC GTG CGG G **- Right**
	N08 hairpin
1st oligo	**Left -** CCG CCC GCA ATT TAA TAA T
2nd oligo	TTA TTA ATA AAG AAA TTT ATT AAT AAA TTA TTA AAT TGC GGG CGG **- Right**
	U20 hairpin
1st oligo	**Left -** CCG CGA CCG TAC CTCG ACC
2nd oligo	GGA TCC GCA TAG AAA TAT GCG GAT CCG GTC GAG GTA CGG TCG CGG **- Right**
	I1 hairpin
1st oligo	**Left -** GCG TCG CAG CGC CAAA AGG CAG GCG GAA AGA GCG CCT GCC TTT TCG CTG
2nd oligo	CGA CGC **- Right**
	I2 hairpin
1st oligo	**Left -** GCG TCG CAG CGC GAT ATC TAC CAC GCG GCG TTA AAT AGA AAT ATT TAA CGC CG
2nd oligo	CGT GCA CGC CGC GTT AAT TAG AAA TAA TTA ACG CGG CGT GGT AGA TAT CGC GCT GCG ACG C **- Right**
	LX hairpin
1st oligo	**Left -** GCG AGC CAT AAT
2nd oligo	CTC ATC TGG A_{X-1} CAG ATG AGA TTA TGG CTC GC **- Right**
	M1 hairpin
1st oligo	**Left -** CGA CCT CAG CTT AAG CTG AG
2nd oligo	GTC GTT TTT TTC GAC CTC AGC TTA AGC TGA GGT CG **- Right**
	M2 hairpin
1st oligo	**Left -** CGA CCT CTT AAG AGG TCG TTT TTT AGA TCA TTA AT
2nd oligo	**- Right** GAT CTA TTT TTT TTT AGA TCA TTA ATG ATC TAA AAA ACG ACC TCT TAA GAG GTC G

Table A.1: Oligonucleotides used to synthesize different DNA hairpins with short handles. Left and right labels indicate the presence of the handle sequence and are marked in bold. Sequences are written in the 5'→3' direction. The first oligonucleotide has a biotin in its 5' end. The second oligonucleotide has a phosphate group in its 3' end, and a digoxigenin tail to be added in its 5' end. Loops in native conformations are underlined. The term A_{X-1} in LX indicates the presence of $X-1$ adenine bases.

CD4F	5'-AAT TCA CAC **GCG AGC CAT AAT CTC ATC TGG AAA CAG ATG AGA TTA TGG CTC GC**A CAC A-3'
CD4R	5'-AGC TTG TGT **GCG AGC CAT AAT CTC ATC TGT TTC CAG ATG AGA TTA TGG CTC GC**G TGT G-3'
T7F	5'-TAA TAC GAC TCA CTA TAG GGA CTG GTG AGT ACT CAA CCA AGT C-3'
T7R	5'-TAG GAA GCA GCC CAG TAG TAG G-3'

Table A.2: Oligonucleotides used to synthesize CD4-RNA with long hybrid DNA/RNA handles. The original CD4 sequence in CD4F and its complementary in CD4R are indicated in bold.

LB	Dissolve 10 g Tryptone, 5 g Yeast extract and 10 g NaCl in 900 ml. Adjust volume to 1 l and autoclave. Store at room temperature.
Ampillcin 00 mg/ml	Dissolve 1 g of ampicillin in 10 ml of sterile water. Filter the solution through 0.22 μm disposable filters under sterile conditions. Store at -20℃ in 1 ml aliquots.
IPTG 100 mM	Dissolve 238 mg of IPTG into 10 ml of sterile water. Filter through a 0.22 μm disposable filter under sterile conditions. Store at -20℃ 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 1 M NaH_2PO_4 with 50 ml of Na_2HPO_4 and add 350 ml of water. Adjust the pH to 7.0 with HCl and bring the final volume to 5000 ml. Autoclave and store at 4℃.
NaPi/EDTA	Mix 19 ml of 1 M NaH_2PO_4 with 10 ml of Na_2HPO_4 and 0.5 ml of 0.2 M EDTA and add 70 ml of water. Adjust pH to 8.0 with HCl and level to 100 ml. Autoclave and store at 4℃.

Table A.3: Buffers required in the synthesis of Barnase.

Appendix B

Simulation of pulling experiments

B.1 Model of the system

In order to simulate pulling experiments we proceed as described in reference [Mano 05, Mano 07a]. The experimental setup consists of three different elements, which are the two-state DNA hairpin (that can be in the native state N or the unfolded state U), the dsDNA handles, and the bead captured in the optical trap. The control parameter is λ, which is the relative distance between the center of the optical trap and the tip of the micropipette (Fig. 3.3). For a given value of λ it is satisfied that:

$$\lambda(f) = x^{\sigma}_{\mathrm{DNA}}(f) + x_{\mathrm{b}}(f) + x_{\mathrm{h}}(f), \tag{B.1}$$

where $x^{\sigma}_{\mathrm{DNA}}(f)$ is the force-dependent end-to-end distance of the hairpin and depends on the state σ of the hairpin ($\sigma = N, U$); $x_{\mathrm{b}}(f)$ is the force-dependent end-to-end distance of the dsDNA handles; and $x_{\mathrm{b}}(f)$ is the position of the captured bead relative to the center of the optical trap. Since only relative variations of λ are important in pulling experiments, the diameter of the trapped bead is not taken into account in Eq. (B.1).

The dsDNA handles are modeled according to the WLC elastic model:

$$f = \frac{k_{\mathrm{B}}T}{P_{\mathrm{h}}}\left[\frac{1}{4\left(1 - x_{\mathrm{h}}/L_c\right)^2} - \frac{1}{4} + \frac{x_{\mathrm{h}}}{L_c}\right], \tag{B.2}$$

where $P_{\mathrm{h}} = 10$ nm and $L_c = 2 \times 29 \times 0.34 = 19.72$ nm [Mano 05, Mano 07a]. More elaborated models can be used, such as the extensible WLC model. However, this is not necessary for the purpose of these simulations, which is to study the quality of the two methods presented in Chapter 3 to obtain the elastic response of ssDNA using simulated data.

The bead captured in the optical trap satisfies:

$$|f| = k_{\mathrm{b}} x_{\mathrm{b}}, \tag{B.3}$$

being k_{b} the trap stiffness (equal to 0.78 pN/nm). Again, a more elaborated model can be used but does not play any significant role for the scope of these simulations. Finally,

the bead behaves as a Brownian particle under the action of two opposing springs, which are the optical trap and the molecular setup. Hence:

$$\gamma \frac{dx_{\mathrm{b}}}{dt} = -k_{\mathrm{b}} x_{\mathrm{b}} + k_x^{\sigma} \left(\lambda - x_{\mathrm{b}}\right) + \eta(t), \tag{B.4}$$

where $\gamma = 2.8 \times 10^{-5}$ pN·s/nm; k_x^{σ} is the the stiffness of the molecular setup, which satisfies:

$$\frac{1}{k_x^{\sigma}} = \frac{1}{k_{\mathrm{h}}} + \frac{1}{k_{\mathrm{DNA}}^{\sigma}}, \tag{B.5}$$

being k_{h} and $k_{\mathrm{DNA}}^{\sigma}$ the stiffnesses of the handles and the DNA hairpin respectively; and $\langle \eta(t)\eta(t') \rangle = 2k_{\mathrm{B}}T\gamma\delta(t-t')$ is the white noise due to thermal fluctuations. From Eq. (B.4) it can be derived that:

$$\left\langle \delta x^2 \right\rangle = \frac{k_{\mathrm{B}}T}{k_x^{\sigma} + k_{\mathrm{b}}}, \qquad \left\langle \delta f^2 \right\rangle = \frac{k_{\mathrm{B}}T k_b^2}{k_x^{\sigma} + k_b}. \tag{B.6a,b}$$

When the hairpin is in state N its presence can be either neglected ($x_{\mathrm{DNA}}^{F} = 0$ accordingly) or it can be assumed to behave as single bond of length $d = 2$ nm that is oriented under the action of a force according to the energy function $E = -\vec{d} \cdot \vec{f}$ (Chapter 3). In this case it is satisfied that:

$$x_{\mathrm{DNA}}^{F} = d \left[\coth\left(\frac{fd}{k_{\mathrm{B}}T}\right) - \frac{k_{\mathrm{B}}T}{fd} \right], \tag{B.7}$$

When the hairpin is in state U it is assumed to satisfy:

$$f = \frac{k_{\mathrm{B}}T}{P} \left[\frac{1}{4\left(1 - x_{\mathrm{DNA}}^{U}/L_c\right)^2} - \frac{1}{4} + \frac{x_{\mathrm{DNA}}^{U}}{L_c} + \sum_{i=2}^{7} a_i \left(\frac{x_{\mathrm{DNA}}}{L_c}\right)^i \right], \tag{B.8}$$

where a_i have different numerical values depending on the interpolating formula used for the WLC model [Bust 94, Bouc 99], P is the persistence length of the ssDNA (taken equal to 1.3 nm from Chapter 3) and $L_c = \mathrm{N}d_b$, being N the total number of bases of the two-state DNA hairpin and d_b the interphosphate distance (taken equal to 0.59 nm/base). For a value of λ, the probabilities of the hairpin to be in the N or U states, p_N and p_U respectively, satisfy the following master equations:

$$\frac{dp_N}{dt} = -k_{N\to U}(\lambda) p_N + k_{N\leftarrow U}(\lambda) p_U \tag{B.9a}$$

$$\frac{dp_U}{dt} = +k_{N\to U}(\lambda) p_N - k_{N\leftarrow U}(\lambda) p_U, \tag{B.9b}$$

where:

$$k_{N\to U}(\lambda) = k_0 \exp\left(-\frac{B(\lambda)}{k_{\mathrm{B}}T}\right) \tag{B.10a}$$

$$k_{N\leftarrow U}(\lambda) = k_0 \exp\left(-\frac{B(\lambda) - \Delta G_{NU}(\lambda)}{k_{\mathrm{B}}T}\right), \tag{B.10b}$$

being $k_0 \simeq 10^4$ the attempt frequency at zero force (Chapter 4); $\Delta G_{NU}(\lambda)$ the free-energy difference between states N and U, and $B(\lambda)$ the kinetic barrier. The term $\Delta G_{NU}(\lambda)$ is calculated according to:

$$\Delta G_{NU}(\lambda) = \Delta G^0_{NU} + \Delta W^{\mathrm{st}}_{NU} + \Delta W^{\mathrm{h}}_{NU} + \Delta W^{\mathrm{b}}_{NU}, \tag{B.11}$$

where ΔG^0_{NU} is the free energy of formation of the hairpin at zero force, and:

$$\Delta W^{\mathrm{st}}_{NU} = \int_0^{x^U_{\mathrm{DNA}}} f^U_{\mathrm{DNA}}(x')dx' - \int_0^{x^F_{\mathrm{DNA}}} f^F_{\mathrm{DNA}}(x')dx' \tag{B.12a}$$

$$\Delta W^{\mathrm{h}}_{NU} = \int_{x^F_{\mathrm{h}}}^{x^U_{\mathrm{h}}} f_{\mathrm{h}}(x')dx' \tag{B.12b}$$

$$\Delta W^{\mathrm{b}}_{NU} = \int_{x^F_{\mathrm{b}}}^{x^U_{\mathrm{b}}} f_{\mathrm{b}}(x')dx', \tag{B.12c}$$

being $(x^U_{\mathrm{DNA}}, x^U_{\mathrm{h}}, x^U_{\mathrm{b}})$ and $(x^N_{\mathrm{DNA}}, x^N_{\mathrm{h}}, x^N_{\mathrm{b}})$ the trio of values that satisfy Eq. (B.1) when the hairpin is in the U or N state given a fixed value of λ; and $f^U_{\mathrm{DNA}}(x)$, $f^F_{\mathrm{DNA}}(x)$, $f_{\mathrm{h}}(x)$ and $f_{\mathrm{b}}(x)$ the equilibrium relation between force and end-to-end distances for the DNA hairpin (in state N and U, Eqs. B.7 and B.8), for the handles (Eq. B.2) and for the bead in the OT (Eq. B.3), respectively.

The kinetic barrier $B(\lambda)$ is calculated according to the Kramers solution (Chapter 4):

$$B(\lambda) = k_{\mathrm{B}}T \log\left(\sum_{n=0}^{\mathrm{N}} \sum_{n'=0}^{n} e^{\frac{\Delta G_n(\lambda) - \Delta G_{n'}(\lambda)}{k_{\mathrm{B}}T}}\right). \tag{B.13}$$

Here, $\Delta G_n(\lambda)$ is the free energy of the configuration of the hairpin determined by the number of open base pairs n at a given value of λ, that is:

$$\Delta G_n(\lambda) = \Delta G^0_n + W^{\mathrm{st}}_n(\lambda) + W^{\mathrm{h}}_n(\lambda) + W^{\mathrm{b}}_n(\lambda), \tag{B.14}$$

where ΔG^0_n is the free energy of formation of configuration n, and:

$$W^{\mathrm{st}}_n(\lambda) = \int_0^{x^n_{\mathrm{DNA}}} f^n_{\mathrm{DNA}}(x')dx' \tag{B.15a}$$

$$W^{\mathrm{h}}_n(\lambda) = \int_0^{x^n_{\mathrm{h}}} f_{\mathrm{h}}(x')dx' \tag{B.15b}$$

$$W^{\mathrm{b}}_n(\lambda) = \int_0^{x^n_{\mathrm{b}}} f_{\mathrm{b}}(x')dx', \tag{B.15c}$$

where the three values $(x^n_{\mathrm{DNA}}, x^n_{\mathrm{h}}, x^n_{\mathrm{b}})$ are the solution of Eq. (B.1) when the DNA hairpin has n open base pairs (and hence its contour length is given by $2nd_b$), and $f^n_{\mathrm{DNA}}(x')$ is the equilibrium relation between force and end-to-end distance for the DNA hairpin at configuration n.

B.2 Simulation

In an unfolding process the hairpin is initialized at N ($n = 0$) at $\lambda \sim 50$ nm. In a folding process, the hairpin is initialized at U ($n = \mathrm{N}$) at $\lambda \sim 400$ nm. During the simulation the relaxation of the handles and the bead is assumed to be instantaneous.

The steps of the algorithm for an unfolding (folding) process are:

1. λ increases (decreases) by an amount $v\Delta t$, where $\Delta t^{-1} = 4{\times}10^3$ Hz is the frequency at which data is collected.

2. Equation (B.1) is solved according to the state of the hairpin and the value of the corresponding force f acting on the experimental setup is determined. To the deterministic value of f a Gaussian noise is added of mean zero and variance given by Eq. (B.6b)

3. If the hairpin is folded, it unfolds with a probability $k_{N\to U}(\lambda)\Delta t$, whereas if it is unfolded it folds with a probability $k_{N\leftarrow U}(\lambda)\Delta t$.

For each DNA hairpin (L06, L12, L16 and L20), four different models were simulated depending on the treatment given to the DNA double helix (neglected or assumed to satisfy Eq. B.7) and elastic properties used to model ssDNA (Eq. B.8 with different values for a_i, $i = 2, \ldots, 7$):

Model 1: **ssDNA elastic properties based on the WLC model [Bust 94] & inclusion of the helix diameter.** The elastic behavior of ssDNA is modeled according to the WLC model, using the interpolation formula proposed in [Bust 94]. The persistence length was set to $P = 1.3$ nm and the interphosphate distance to $d_\mathrm{b} = 0.59$ nm/base. The diameter of the double helix, present when the DNA hairpin is folded, is modeled as a single bond of length $d = 2$ nm that aligns with the applied force as a single dipole does under the action of a magnetic field.

Model 2: **ssDNA elastic properties based on the WLC model [Bust 94] & exclusion of the helix diameter.** The elastic behavior of ssDNA is modeled according to the WLC model, using the interpolation formula proposed in [Bust 94]. The persistence length was set to $P = 1.3$ nm and the interphosphate distance to $d_\mathrm{b} = 0.59$ nm/base. The contribution of the DNA hairpin diameter to the molecular extension and stiffness in the folded state is neglected.

Model 3: **ssDNA elastic properties based on the WLC model [Bouc 99] & inclusion of the helix diameter.** The elastic behavior of ssDNA is modeled according to the WLC model, using the interpolation formula derived in [Bouc 99]. The persistence length was set to $P = 1.3$ nm and the interphosphate distance to $d_\mathrm{b} = 0.59$ nm/base. The diameter of the double helix, present when the DNA hairpin is folded, is modeled as a single bond of length $d = 2$ nm that aligns with the applied force as a single dipole does under the action of a magnetic field.

Model 4: **ssDNA elastic properties based on the WLC model [Bouc 99] & exclusion of the helix diameter.** The elastic behavior of ssDNA is modeled according to the WLC model, using the interpolation formula derived in [Bouc 99]. The persistence length was set to $P = 1.3$ nm and the interphosphate distance to $d_b = 0.59$ nm/base. The contribution of the DNA hairpin diameter to the molecular extension and stiffness in the folded state is neglected.

For each of the four DNA hairpins, 100 unfolding and folding FDC are simulated. In Figure B.1 we show FDC obtained from pulling experiments and from simulation according to model 1 for the four hairpins. Data for simulations performed according to models 2, 3 and 4 is not shown, but results are practically the same. Each model is then simulated for each of the four hairpins (L06, L12, L16 and L20) and the obtained FDC are analyzed *a posteriori* using the four different models according to the two methods presented in Chapter 3 to extract the elastic properties of ssDNA.

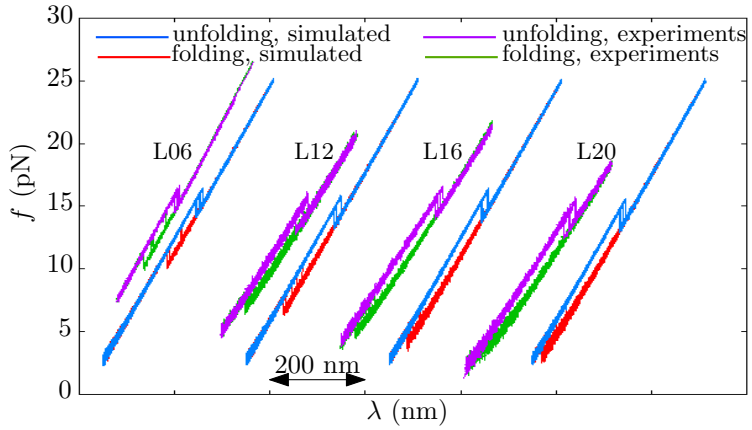

Figure B.1: Comparison between FDC obtained from pulling experiments and from simulation according to model 1 for the four DNA hairpins L06, L12, L16 and L20.

B.3 Recovery of the elastic properties of ssDNA

B.3.1 Force-jump measurement

Simulated data is analyzed according to the force-jump method presented in Section 3.2: the jump in force upon unfolding and folding is extracted, and using Eq. 3.5 the elastic response $x_{\mathrm{ssDNA}}(f)$ is obtained. Simulated data from the four different molecules is merged together in a single $x_{\mathrm{ssDNA}}(f)/L_c$-plot as described in Section 3.2, and the results are fitted to the WLC using both interpolating formulas proposed in [Bust 94] and [Bouc 99], taking into account or neglecting the presence of the hairpin. In this way, each set of simulations was analyzed according to the four different models presented above.

Results obtained for the four simulations analyzed within the assumptions of each model are summarized in the two first rows of Table B.1. It can be seen that values of P and d_b recovered from the data analysis are in agreement, within error bars, with values used to simulate pulling experiments.

Simulations with **Model 1**	P (nm)	d_b (nm/base)
Analysis with Model 1	1.3±0.1	0.60±0.02
Model 2	1.3±0.1	0.56±0.02
Model 3	1.1±0.1	0.60±0.02
Model 4	1.1±0.1	0.57±0.02

Simulations with **Model 2**	P (nm)	d_b (nm/base)
Analysis with Model 1	1.2±0.1	0.66±0.02
Model 2	1.3±0.1	0.60±0.02
Model 3	1.1±0.1	0.66±0.02
Model 4	1.1±0.1	0.61±0.02

Simulations with **Model 3**	P (nm)	d_b (nm/base)
Analysis with Model 1	1.4±0.1	0.60±0.02
Model 2	1.5±0.1	0.55±0.02
Model 3	1.2±0.1	0.60±0.02
Model 4	1.3±0.1	0.56±0.02

Simulations with **Model 4**	P (nm)	d_b (nm/base)
Analysis with Model 1	1.4±0.1	0.65±0.02
Model 2	1.4±0.1	0.60±0.02
Model 3	1.2±0.1	0.65±0.02
Model 4	1.3±0.1	0.61±0.02

Table B.1: Force-jump analysis of simulations performed according to the four different models. Persistence length P and interphosphate distance d_b obtained by analyzing data simulated according to models 1-4 using the force-jump measurement and models 1-4.It can be observed that parameters of the simulation ($P = 1.3$ nm and $d_b = 0.59$ nm/base) are recovered within error bars

B.3.2 Stiffness measurement

Simulated data is also analyzed according to the force-jump method presented in Section 3.3: the force-dependent stiffnesses of the folded and the unfolded branch are extracted from the simulated FDC. Using Eq. 3.9 the stiffness of ssDNA is then obtained. Again, data is analyzed using the WLC using both interpolating formulas proposed in [Bust 94] and [Bouc 99] to model ssDNA, and taking into account or neglecting the presence of the hairpin (models 1–4).

Results obtained for the four simulations are summarized in Tables B.2. It can be seen that values of P and d_b recovered from the data analysis are in agreement, within

error bars, with values used to simulate pulling experiments.

Simulations with **Model 1**	P (nm)	d_b (nm/base)
Analysis with Model 1	1.3±0.1	0.59±0.02
Model 2	1.2±0.1	0.51±0.02
Model 3	0.8±0.1	0.50±0.02
Model 4	0.9±0.1	0.48±0.02

Simulations with **Model 2**	P (nm)	d_b (nm/base)
Analysis with Model 1	1.3±0.1	0.60±0.02
Model 2	1.4±0.1	0.60±0.02
Model 3	1.0±0.1	0.58±0.02
Model 4	1.0±0.1	0.54±0.02

Simulations with **Model 3**	P (nm)	d_b (nm/base)
Analysis with Model 1	1.5±0.1	0.56±0.02
Model 2	1.5±0.1	0.50±0.02
Model 3	1.4±0.1	0.61±0.02
Model 4	1.5±0.1	0.61±0.02

Simulations with **Model 4**	P (nm)	d_b (nm/base)
Analysis with Model 1	1.5±0.1	0.61±0.02
Model 2	1.7±0.1	0.60±0.02
Model 3	1.3±0.1	0.62±0.02
Model 4	1.4±0.1	0.61±0.02

Table B.2: Stiffness analysis of simulations performed according to the four different models. Persistence length P and interphosphate distance d_b obtained by analyzing data simulated according to models 1-4 using the measurement of the stiffness and models 1-4. It can be observed that parameters of the simulation ($P = 1.3$ nm and $d_b = 0.59$ nm/base) are recovered within error bars.

Appendix C

Transition state theory

C.1 Eyring's theory

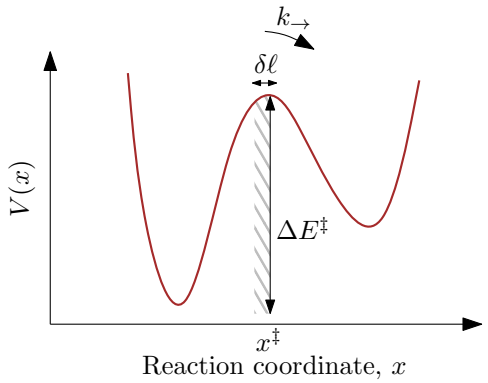

Figure C.1: Potential of mean force for a chemical reaction.

The Eyring's theory [Eyri 35, Zhou 10] assumes that:

i. Molecules energetically located in the barrier region $\delta\ell$ (dashed area in Fig. C.1) are in thermal equilibrium with those molecules found in the energy well of the reactant state (left well in Fig. C.1).

ii. The motion of molecules obeys Newton's equation.

iii. As soon as a reactant molecule crosses the energy barrier located at $x^{\ddagger}$ it becomes a product.

Consider N reactant molecules, δN of whom are located at the barrier maximum (*i. e.*, they are in the activated state, indicated by the dashed area in Fig. C.1). Because such molecules are found at the maximum of the potential of mean force $V(x)$, they are not subjected to any force. For a given time interval dt, the number of activated molecules

that will cross the energy barrier $N_{\rightarrow}$ will be:

$$N_{\rightarrow} = \delta N \frac{v dt}{\delta \ell}, \tag{C.1}$$

being v the mean velocity towards the potential well of products (right well in Fig. C.1), $\delta\ell$ the maximum distance that the activated molecules must travel in order to overcome the barrier, and δN the number of molecules located at the barrier maximum. The kinetic rate $k_{\rightarrow}$ is defined as the fraction of reactant molecules that cross the kinetic barrier per unit time, and therefore, from Eq. (C.1) it can be written:

$$k_{\rightarrow} = \frac{\delta N}{N} \frac{v}{\delta \ell}, \tag{C.2}$$

The fraction $\delta N/N$ can be writen in terms of the partition functions evaluated at the energetic barrier, $Z^{\ddagger}$, and at the reactant well, Z^{R}:

$$\frac{\delta N}{N} = \frac{Z^{\ddagger}}{Z^{R}}, \tag{C.3}$$

where

$$\begin{aligned} Z^{\ddagger} &= \frac{1}{h} \int_{-\infty}^{\infty} dp e^{-\beta \frac{p^2}{2m}} \int_{\delta \ell} dx e^{-\beta V(x)} \\ &\simeq \frac{1}{h} \sqrt{\frac{2\pi m}{\beta}} \delta \ell e^{-\beta V(x^{\ddagger})}, \end{aligned} \tag{C.4}$$

and $\beta = 1/k_{\mathrm{B}}T$.

Moreover, v is defined as the mean velocity pointing to the potential well of the products. Therefore:

$$v = \frac{\int_0^{\infty} dp v e^{-\beta \frac{p^2}{2m}}}{\int_{-\infty}^{\infty} dp e^{-\beta \frac{p^2}{2m}}} = \frac{1}{\sqrt{2\pi m \beta}}, \tag{C.5}$$

where it was used that $v = p/m$.

Using equations (C.3), (C.4) and (C.5), it can be shown that Eq. (C.2) becomes:

$$k_{\rightarrow} = \frac{k_{\mathrm{B}}T}{h Z_R} e^{-\beta V(x^{\ddagger})} = \frac{k_{\mathrm{B}}T}{h Z_R} e^{-\beta \Delta E^{\ddagger}}, \tag{C.6}$$

where $V(x^{\ddagger}) = \Delta E^{\ddagger}$.

C.2 Kramers' theory

Suppose the system sketched in Fig. C.2, where a Brownian particle is subject to the one dimensional potential of mean force $V(x)$. The time evolution of the probability density

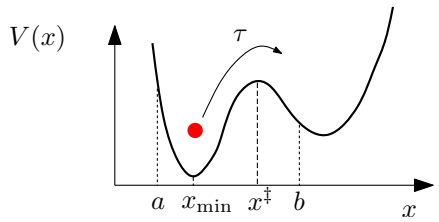

Figure C.2: Brownian particle in a double well potential. The position where the potential has the minimum and the maximum value, $x_{\min}$ and $x^{\ddagger}$ respectively, are indicated. In addition, positions a and b used to model the absorbing and reflecting boundary conditions of the dynamics of the Brownian particle are indicated.

function $p(x,t)$ to find the particle at the position x at time t follows the Fokker-Planck equation [Zwan 01, Hyeo 07]:

$$\frac{\partial p(x,t)}{\partial t} = D\frac{\partial}{\partial x}\left[\frac{\partial}{\partial x} + \frac{1}{k_{\mathrm{B}}T}\frac{dV(x)}{dx}\right]p(x,t), \tag{C.7}$$

where $D = k_{\mathrm{B}}T/\gamma$ is the diffusion coefficient, γ is the friction coefficient, k_{B} is the Boltzmann constant and T is the absolute temperature. Equation (C.7) is rewritten in terms of a Fokker-Plank operator, $\mathcal{L}_{FP} = D\frac{\partial}{\partial x}\left[\frac{\partial}{\partial x} + \frac{1}{k_{\mathrm{B}}T}\frac{dV(x)}{dx}\right]$:

$$\frac{\partial p(x,t)}{\partial t} = \mathcal{L}_{FP}p(x,t) \tag{C.8}$$

By considering that the particle initially is located at x_0 in a region $\mathcal{R}$ (*i. e.* $x_0 \in \mathcal{R} = [a,b]$), then $p(x,0) = \delta(x - x_0)$ and the formal solution of Eq. (C.8) is:

$$p(x,t) = e^{t\mathcal{L}_{FP}}\delta(x - x_0) \tag{C.9}$$

The survival probability is defined as the probability of the Brownian particle to remain inside $\mathcal{R}$ at time t, and can be expressed as:

$$S(t,\mathcal{R}) = \int_{x\in\mathcal{R}} p(x,t)dx \tag{C.10}$$

The time derivative of the survival probability is equal to the probability density function of the survival time $\rho(t)$, that is, the time it takes to the Brownian particle to leave the region $\mathcal{R}$.

$$S(t+dt,\mathcal{R}) - S(t,\mathcal{R}) = -\rho(t)dt \Rightarrow \rho(t) = -\frac{\partial S(t,\mathcal{R})}{\partial t} \tag{C.11}$$

Therefore, the mean first passage time τ can be calculated as:

$$\begin{aligned}\tau &= \int_0^\infty dt\ t\rho(t) = -\int_0^\infty dt\ t\frac{\partial S(t,\mathcal{R})}{\partial t} \\ &= \int_0^\infty dt\ S(t,\mathcal{R}) = \int_0^\infty dt \int_{x\in\mathcal{R}} dx\ p(x,t) \\ &= \int_0^\infty dt \int_{x\in\mathcal{R}} dx\ e^{t\mathcal{L}_{FP}}\delta(x-x_0) = \int_0^\infty dt \int_{x\in\mathcal{R}} dx\ \delta(x-x_0)e^{t\mathcal{L}_{FP}^\ddagger}1 \\ &= \int_0^\infty dt\ e^{t\mathcal{L}_{FP}^\ddagger}1 \end{aligned} \tag{C.12}$$

Here, $\mathcal{L}_{FP}^\ddagger$ is the adjoint operator, that satisfies $f(x)\mathcal{L}_{FP}[g(x)] = g(x)\mathcal{L}_{FP}^\ddagger[f(x)]$. If $\mathcal{L}_{FP}^\ddagger$ is applied at both sides of Eq. (C.12) it is obtained:

$$\begin{aligned}\mathcal{L}_{FP}^\ddagger\tau &= \int_0^\infty dt\ \mathcal{L}_{FP}^\ddagger e^{t\mathcal{L}_{FP}^\ddagger}1 \\ &= \int_0^\infty dt\ \frac{de^{t\mathcal{L}_{FP}^\ddagger}}{dt}1 \\ &= -1 \end{aligned} \tag{C.13}$$

Which gives a differential equation for the survival time τ that depends on the adjoint Fokker-Plank operator. It can be shown that this differential equation is given by:

$$\frac{\partial^2\tau}{\partial x^2} - \frac{1}{k_\mathrm{B}T}\frac{dV}{dx}\frac{\partial\tau}{\partial x} = -1. \tag{C.14}$$

In order to solve Eq. (C.14), absorbing and reflecting boundary conditions are taken, $\tau(b) = 0$ and $\frac{\partial\tau}{\partial x}|_{x=a} = 0$ respectively.

$$\tau(x) = \frac{1}{D}\int_x^b dy\ e^{\frac{V(y)}{k_\mathrm{B}T}}\int_a^y dz e^{-\frac{V(z)}{k_\mathrm{B}T}}. \tag{C.15}$$

Finally, the kinetic rate $k_\rightarrow$ is defined as the inverse of the mean first passage time evaluated at $x = a$, $k_\rightarrow = \tau(a)^{-1}$. Therefore:

$$k_\rightarrow = \frac{D}{\int_a^b dy\ e^{\frac{V(y)}{k_\mathrm{B}T}}\int_a^y dz e^{-\frac{V(z)}{k_\mathrm{B}T}}}. \tag{C.16}$$

Using the saddle point approximation in (C.16) it can be derived:

$$k_\rightarrow = \frac{\sqrt{-V''_\mathrm{max}V''_\mathrm{min}}}{2\pi k_\mathrm{B}T}D\exp\left[-\frac{(V_\mathrm{max}-V_\mathrm{min})}{k_\mathrm{B}T}\right], \tag{C.17}$$

where $V_\mathrm{max} = V(x^\ddagger)$ and $V_\mathrm{min} = V(x_\mathrm{min})$, and V''_max and V''_min are the values of the second derivatives of the one-dimensional potential evaluated at $x^\ddagger$ and x_min, respectively (that is, V''_max and V''_min are the stiffnesses of the TS and the minimum of the potential).

C.3 The kinetic barrier

By equating Eq. (C.16) and the Arrhenius Eq. (4.13) we get:

$$k_0 e^{-\frac{\Delta E^{\ddagger}}{k_{\mathrm{B}}T}} = \frac{D}{\int_a^b dy\ e^{\frac{V(y)}{k_{\mathrm{B}}T}} \int_a^y dz e^{-\frac{V(z)}{k_{\mathrm{B}}T}}} \tag{C.18}$$

Therefore,

$$\frac{\Delta E^{\ddagger}}{k_{\mathrm{B}}T} = \log\left(\frac{k_0}{D}\right) + \log\left(\int_a^b dy\ e^{\frac{V(y)}{k_{\mathrm{B}}T}} \int_a^y dz e^{-\frac{V(z)}{k_{\mathrm{B}}T}}\right) \tag{C.19}$$

$\Delta E^{\ddagger}$ is defined as the change in energy between the initial state and the transition state (TS) of a chemical reaction. As it is usually referred to as the kinetic barrier, is is rewritten as B_{eff}. By taking $k_0/D \sim 1$ we get:

$$B_{\mathrm{eff}} = k_{\mathrm{B}}T \log\left(\int_a^b dy\ e^{\frac{V(y)}{k_{\mathrm{B}}T}} \int_a^y dz e^{-\frac{V(z)}{k_{\mathrm{B}}T}}\right) \tag{C.20}$$

Therefore, Kramers' theory provides an analytical expression to evaluate the size of the kinetic barrier along a reaction pathway for any potential $V(x)$.

The force f is defined as:

$$f = -\frac{dV}{dx}, \tag{C.21}$$

and it can be shown that the derivative of B_{eff} as a function of f equals:

$$\begin{aligned}\frac{\partial B_{\mathrm{eff}}}{\partial f} =& e^{-\frac{B_{\mathrm{eff}}}{k_{\mathrm{B}}T}} \Big[\int_a^b dy\ (-y) e^{\frac{V(y)}{k_{\mathrm{B}}T}} \int_a^y dz e^{-\frac{V(z)}{k_{\mathrm{B}}T}} + \\ &+ \int_a^b dy\ e^{\frac{V(y)}{k_{\mathrm{B}}T}} \int_a^y dz (z) e^{-\frac{V(z)}{k_{\mathrm{B}}T}} \Big]\end{aligned} \tag{C.22}$$

Finally, applying the saddle point approximation in Eqs. (C.20) and (C.22), we get:

$$\frac{\partial B_{\mathrm{eff}}}{\partial f} \simeq e^{-\frac{V(x^{\ddagger})-V(x_{\min})}{k_{\mathrm{B}}T}} \left[-x^{\ddagger} e^{\frac{V(x^{\ddagger})-V(x_{\min})}{k_{\mathrm{B}}T}} + x_{\min} e^{\frac{V(x^{\ddagger})-V(x_{\min})}{k_{\mathrm{B}}T}} \right] \tag{C.23}$$

$$= -\left(x^{\ddagger} - x_{\min}\right) = -\Delta x_{\mathrm{eff}} \tag{C.24}$$

Therefore, form the knowledge of the profile $V(x)$ it is possible to compute the size of the effective barrier B_{eff} and the effective distance between the TS and the stable state Δx_{eff}.

C.4 Bell-Evans model

The Bell-Evans (BE) model is one of the first approaches proposed to analyze rupture force events at the level of single molecules [Bell 78]. In the context of DFS experiments,

it considers that force lowers the kinetic barrier $\Delta G^{\ddagger}$ of a system by the quantity $-fx^{\ddagger}$, being $x^{\ddagger}$ the distance between the initial state (which in the case of biomolecules is identified with the native state, and in the case of bimolecular interactions is identified with the bonded state) to the TS. Therefore:

$$B(f) = B_{BE} - fx^{\ddagger} \tag{C.25}$$

The expression for the force-dependent kinetic barrier given by Eq. (C.25) is introduced in the phenomenological Arrhenius Eq. (4.13) and the following expression for the kinetic rate is obtained:

$$k(f) = k_m \exp\left(\frac{fx^{\ddagger}}{k_{\mathrm{B}}T}\right), \qquad k_m = k_0 \exp\left(-\frac{\Delta G^{\ddagger}}{k_{\mathrm{B}}T}\right). \tag{C.26}$$

k_0 is the attempt rate at zero force and k_m is the kinetic rate of dissociation or unfolding at zero force; $\Delta G^{\ddagger}$ the height of the kinetic barrier; $x^{\ddagger}$ is the distance between the bonded and the TS; k_{B} is the Boltzmann constant; and T is the temperature (taken equal to 298 K) [Stri 99, Evan 09, Merk 99, Evan 10].

Unfolding or dissociation events

Suppose of a molecular system made of a single folded molecule, or made of two bonded molecules, to which a force that increases linearly with time ($f = rt$) is applied. The probability of the system to remain in the folded or the bounded state satisfies the following master equation:

$$\frac{dP(f)}{df} = -\frac{k_{\rightarrow}(f)}{r}P(f), \tag{C.27}$$

which can be analytically solved by introducing Eq. (C.26) to model the kinetic rate. The solution is given by:

$$\log\left(-r\log P(f)\right) = \log\left(\frac{k_m k_{\mathrm{B}}T}{rx^{\ddagger}}\right) + \log\left[\exp\left(\frac{fx^{\ddagger}}{k_{\mathrm{B}}T}\right) - 1\right]. \tag{C.28}$$

Next, an analytical expression for the probability density function of the rupture force, $p(f_{\mathrm{rupt}})$, can be obtained from the derivative of $P(f)$ respect to force f:

$$p(f_{\mathrm{rupt}}) = \frac{dP(f)}{df} = \frac{k_m}{v}\exp\left(\frac{f_{\mathrm{rupt}}x^{\ddagger}}{k_{\mathrm{B}}T}\right)\exp\left\{-\frac{k_m k_{\mathrm{B}}T}{vx^{\ddagger}}\left[\exp\left(\frac{f_{\mathrm{rupt}}x^{\ddagger}}{k_{\mathrm{B}}T}\right) - 1\right]\right\} \tag{C.29}$$

Finally, from $p(f_{\mathrm{rupt}})$ the most probable rupture force $\langle f_{\mathrm{rupt}}\rangle$ and the standard deviation σ_f can be calculated:

$$\langle f_{\mathrm{rupt}}\rangle = \frac{k_{\mathrm{B}}T}{x^{\ddagger}}\log\left(\frac{x^{\ddagger}v}{k_m k_{\mathrm{B}}T}\right) \tag{C.30}$$

$$\sigma_f = \log\left(\frac{3+\sqrt{5}}{2}\right)\frac{k_{\mathrm{B}}T}{x^{\ddagger}}. \tag{C.31}$$

Folding or association events

The kinetic rates must satisfy detailed balance, and consequently the folding or association kinetic rate can be written directly as:

$$k_{\leftarrow} = k_m \exp\left(\frac{fx^{\ddagger} - \Delta G}{k_{\mathrm{B}}T}\right). \tag{C.32}$$

ΔG is the difference in free energy between the native or bonded state to the unfolded or dissociated state. As before, the probability to remain in the unfolded or unbounded state satisfies the following master equation,

$$\frac{dP(f)}{df} = -\frac{k_{\leftarrow}(f)}{r}P(f), \tag{C.33}$$

which is formally identical to Eq. (C.27). Hence, the folding force distribution, or binding force distribution, is identical to the one given by Eq. (C.29) but replacing k_m by $k_m \exp\left(-\Delta G/k_{\mathrm{B}}T\right)$.

C.5 Dudko-Hummer-Szabo model

The Dudko-Hummer-Szabo (DHS) model considers two different analytical formulas for the potential of mean force $V(x)$. The first assumes a parabolic potential with a barrier of height $\Delta G^{\ddagger}$ located at $x^{\ddagger}$:

$$V(x) = \begin{cases} \Delta G^{\ddagger}\left(\frac{x}{x^{\ddagger}}\right)^2 & x < x^{\ddagger} \\ 0 & x \geq x^{\ddagger}. \end{cases} \tag{C.34}$$

The second potential, that also contains a barrier of height $\Delta G^{\ddagger}$ located at $x^{\ddagger}$, has a cubic shape:

$$V(x) = \frac{3}{2}\Delta G^{\ddagger}\left(\frac{x}{x^{\ddagger}}\right) - 2\Delta G^{\ddagger}\left(\frac{x}{x^{\ddagger}}\right)^3 \tag{C.35}$$

The DHS model includes the action of the force f on the potential of mean force as $V(x,f) = V(x) - fx$, where $V(x)$ is given by any of the two potentials given in Eqs. (C.34) and (C.35). Using $V(x,f)$, Kramer's equation (C.16) can be analytically solved for the two cases in order to get an expression for $k(f)$, and results can be merged together as:

$$k(f) = k_m\left(1 - \gamma\frac{fx^{\ddagger}}{\Delta G^{\ddagger}}\right)^{1/\gamma - 1} \exp\left\{\frac{\Delta G^{\ddagger}}{k_{\mathrm{B}}T}\left[1 - \left(1 - \gamma\frac{fx^{\ddagger}}{\Delta G^{\ddagger}}\right)^{1/\gamma}\right]\right\}, \tag{C.36}$$

where $\gamma = 1/2$ gives the solution related to the parabolic potential (Eq. C.34), and $\gamma = 2/3$ gives the solution for the cubic potential (Eq. C.35). Interestingly, for $\gamma = 1$ the BE model is recovered.

Using the kinetic rate given by Eq. (C.36), the differential master equation (C.27) can be solved and the following expression for the rupture force distribution is obtained:

$$p(f_{\text{rupt}}) = \frac{k_m}{v}\left(1-\gamma\frac{f_{\text{rupt}}x^{\ddagger}}{\Delta G^{\ddagger}}\right)\exp\left\{\frac{\Delta G^{\ddagger}}{k_{\text{B}}T}\left[1-\left(1-\gamma\frac{f_{\text{rupt}}x^{\ddagger}}{\Delta G^{\ddagger}}\right)^{1/\gamma}\right]-\right.$$
$$\left.-\frac{k_m k_{\text{B}}T}{vx^{\ddagger}}\left(e^{\frac{\Delta G^{\ddagger}}{k_{\text{B}}T}\left[1-\left(1-\gamma\frac{f_{\text{rupt}}x^{\ddagger}}{\Delta G^{\ddagger}}\right)^{1/\gamma}\right]}-1\right)\right\}. \tag{C.37}$$

Again, from Eq. (C.37) it is straightforward to derive the average rupture force and its standard deviation:

$$\langle f_{\text{rupt}}\rangle = \frac{\Delta G^{\ddagger}}{\gamma x^{\ddagger}}\left\{1-\left[\frac{k_{\text{B}}T}{\Delta G^{\ddagger}}\log\left(\frac{k_m k_{\text{B}}T e^{\frac{\Delta G^{\ddagger}}{k_{\text{B}}T}+0.577}}{x^{\ddagger}v}\right)\right]^{\gamma}\right\} \tag{C.38}$$

$$\sigma_f = \frac{k_{\text{B}}T\pi}{\sqrt{6}x^{\ddagger}}\left[\frac{k_{\text{B}}T}{\Delta G^{\ddagger}}\log\left(\frac{k_m k_{\text{B}}T e^{\frac{\Delta G}{k_{\text{B}}T}+1.064}}{x^{\ddagger}v}\right)\right]^{\gamma-1}. \tag{C.39}$$

The DHS model has been recently extended to address the unbinding or folding, or to characterize the presence of intermediate states. More information can be found in references [Pier 13, Gara 14].

Appendix D

Bell-Evans model and nucleic acid hairpins

As explained in Section 4.1.1, the configurations accessible by a nucleic acid hairpin are labeled with the number of open basepairs n. The free energy of each configuration at a given force f consists of three contributions,

$$\Delta G_n(f) = \Delta G_n^0 + \Delta G_n^{\mathrm{st}}(f) + \Delta G_n^{\mathrm{d}}(f). \tag{D.1}$$

ΔG_n^0 is the free-energy of formation of configuration n, $\Delta G_n^{\mathrm{st}}(f)$ is an elastic contribution due to the stretching of the released single-stranded nucleic acid at configuration n, and $\Delta G_n^{\mathrm{d}}(f)$ is the energetic contribution due to orientation of the double helix.

In this appendix, the force-dependent contributions $\Delta G_n^i(f)$ ($i = \mathrm{d},\ \mathrm{st}$) are written as:

$$\Delta G_n^i(f) = -\int_0^f x_n^i(f')df' = \int_0^{x_n^i(f)} F_n^i(x')dx' - f x_n^i(f), \tag{D.2}$$

being $x_n^i(f)$ the equilibrium end-to-end distance of the element i projected along the force axis, and $F_n^i(f)$ its inverse function.

According to the Arrhenius theory, a simplified version of the kinetic rates of unfolding and folding under tension in a two-state folder are given by [Moss 09b]:

$$k_{N\to U}(f) = k_0 \exp\left(-\frac{\Delta G_{n^\ddagger}(f) - \Delta G_0(f)}{k_{\mathrm{B}}T}\right) \tag{D.3a}$$

$$k_{N\leftarrow U}(f) = k_0 \exp\left(-\frac{\Delta G_{n^\ddagger}(f) - \Delta G_{\mathrm{N}}(f)}{k_{\mathrm{B}}T}\right), \tag{D.3b}$$

where k_0 is the attempt frequency at zero force, $\Delta G_{n^\ddagger}(f)$ is the free energy of the TS, $\Delta G_0(f)$ is the free energy of state N (hairpin folded, $n = 0$), and $\Delta G_{\mathrm{N}}(f)$ is the free energy of state U (unfolded haripin, $n = \mathrm{N}$, being N the number of basepairs in the hairpin stem).

The two expressions in Eq. (D.3) can be arranged in order to obtain:

$$k_{N\to U}(f) = k_0 \exp\left(-\frac{B(f) - f x^{\mathrm{st}}_{n^\ddagger}(f)}{k_{\mathrm{B}}T}\right), \tag{D.4a}$$

$$k_{N\leftarrow U}(f) = k_0 \exp\left(-\frac{B(f) - \Delta G(f) - f\left(x^{\mathrm{st}}_{n^\ddagger}(f) + x^{\mathrm{d}}_{n^\ddagger}(f) - x^{\mathrm{st}}_{\mathrm{N}}(f)\right)}{k_{\mathrm{B}}T}\right) \tag{D.4b}$$

where

$$B(f) = \Delta G^0_{n^\ddagger} - \Delta G^0_0 + \int_0^{x^{\mathrm{st}}_{n^\ddagger}(f)} F^{\mathrm{st}}_{n^\ddagger}(x)dx \tag{D.5}$$

$$\Delta G(f) = \Delta G^0_{\mathrm{N}} - \Delta G^0_0 + \int_0^{x^{\mathrm{st}}_{\mathrm{N}}(f)} F^{\mathrm{st}}_{\mathrm{N}}(x)dx - \int_0^{x^{\mathrm{d}}_{n^\ddagger}(f)} F^{\mathrm{d}}_{n^\ddagger}(x)dx. \tag{D.6}$$

At this stage, it is useful to define the molecular end-to-end distances $x_N(f)$, $x_{\mathrm{TS}}(f)$ and $x_U(f)$ of the native, the transition, and the unfolded states, respectively, as:

$$x_N(f) = x^{\mathrm{d}}_0(f) \tag{D.7a}$$

$$x_{\mathrm{TS}}(f) = x^{\mathrm{st}}_{n^\ddagger}(f) + x^{\mathrm{d}}_{n^\ddagger}(f) \tag{D.7b}$$

$$x_U(f) = x^{\mathrm{st}}_{\mathrm{N}}(f), \tag{D.7c}$$

and, since $x^{\mathrm{d}}_0(f) = x^{\mathrm{d}}_{n^\ddagger}(f)$, the kinetic rates in Eq. (D.4) can be further simplified:

$$k_{N\to U}(f) = k_0 \exp\left(-\frac{B(f) - f x_{N-\mathrm{TS}}(f)}{k_{\mathrm{B}}T}\right) \tag{D.8a}$$

$$k_{N\leftarrow U}(f) = k_0 \exp\left(-\frac{B(f) - \Delta G(f) + f x_{\mathrm{TS}-U}(f)}{k_{\mathrm{B}}T}\right), \tag{D.8b}$$

where $x_{A-B} = x_B - x_A$.

In DFS experiments, kinetic rates are usually measured close to the coexistence force f_c. Hence, a Taylor expansion close to the coexistence force is performed for the kinetic rates in Eq. (D.8). By defining $k_m = k_0 \exp\left(-B(f_c)/k_{\mathrm{B}}T\right)$, it can be demonstrated that to a first order the result is:

$$k_{N\to U}(f) \simeq k_m \exp\left(\frac{f x_{N-\mathrm{TS}}(f_c)}{k_{\mathrm{B}}T}\right) \tag{D.9a}$$

$$k_{N\leftarrow U}(f) \simeq k_m \exp\left(\frac{\Delta G(f_c) - f x_{\mathrm{TS}-U}(f_c)}{k_{\mathrm{B}}T}\right), \tag{D.9b}$$

Finally, at f_c the unfolding and folding kinetic rates must be identical ($k_{N\to U}(f_c) = k_{N\leftarrow U}(f_c)$). Therefore, it is satisfied that:

$$\Delta G(f_c) = f_c x_{N-U}(f_c), \tag{D.10}$$

where $x_{N-U} = x_{N-\mathrm{TS}} + x_{\mathrm{TS}-U}$.

Appendix E

Matching experimental and theoretical kinetic barriers

In Section 4.3.2 it is shown how to theoretically compute the force-dependent kinetic barrier of nucleic acid hairpins using the NN model (hereafter referred to as $B^{\mathrm{th}}_{NU}(f)$) using Eq. (4.32). Furthermore, it is explained how to extract its profile from unfolding and folding kinetic rates measured in pulling experiments using Eq. (4.23). By matching both estimations it is possible to obtain the free energy of formation of the hairpin ΔG^0_{NU}, and the attempt frequency at zero force k_0^{NU}.

Based on Eq. (4.23a), the following χ^2_{unf} function can be defined:

$$\chi^2_{\mathrm{unf}} = \sum_{i=1}^{N_u} \left[\frac{B^{\mathrm{th}}_{NU}(f_i)}{k_{\mathrm{B}}T} - \left(\log k_0^{NU} - \log k_{N\to U}(f_i) \right) \right]^2, \tag{E.1}$$

where N_u is the total number of points for the experimental unfolding kinetic rate, $\log k_{N\to U}(f_i)$ is the experimentally measured logarithm of the unfolding kinetic rate at force f_i ($i = 1, \ldots, N_u$), and $B^{\mathrm{th}}_{NU}(f_i)$ is the theoretical value of the height of the kinetic barrier obtained at f_i.

Similarly, based on Eq. (4.23b), the following χ^2_{fold} function can be defined:

$$\chi^2_{\mathrm{fold}} = \sum_{i=1}^{N_f} \left[\frac{B^{\mathrm{th}}(f_i)}{k_{\mathrm{B}}T} - \left(\log k_0^{NU} + \frac{\Delta G_{NU}(f_i)}{k_{\mathrm{B}}T} - \log k_{N\leftarrow U}(f_i) \right) \right]^2, \tag{E.2}$$

$$\tag{E.3}$$

where N_f is the total number of points for the experimental folding kinetic rate, $\log k_{N\leftarrow U}(f_i)$ is the experimentally measured logarithm of the folding kinetic rate at force f_i ($i = 1, \ldots, N_f$), $B^{\mathrm{th}}_{NU}(f_i)$ is the theoretical value of the height of the kinetic barrier obtained at f_i, and $\Delta G_{NU}(f_i) = \Delta G_U(f_i) - \Delta G_N(f_i)$ is the free-energy difference between states U and N at f_i (Eq. 4.1), which satisfies:

$$\Delta G_{NU}(f_i) = \Delta G^0_{NU} - \int_0^{f_i} x_{n=\mathrm{N}}(f')df' + \int_0^{f_i} x_{\mathrm{d}}(f')df'. \tag{E.4}$$

Finally, the total χ^2 function is defined:

$$\chi^2 = \chi^2_{\rm unf} + \chi^2_{\rm fold}. \tag{E.5}$$

The best match between the experimental and the theoretical profile of the force-dependent kinetic barrier, $B(f)$, is given by the values of $\log k_0^{NU}$ and ΔG^0_{NU} that minimize this χ^2 function. From the derivatives of χ^2 as a function of $\log k_0^{NU}$ and ΔG^0_{NU} is can be shown that best estimators for both magnitudes are:

$$\log k_0^{NU} = \frac{1}{N_u}\sum_{i=1}^{N_u}\left[\frac{B^{\rm th}(f_i)}{k_{\rm B}T} + \log k_{N\to U}(f_i)\right] \tag{E.6}$$

$$\Delta G^0_{NU} = \frac{1}{N_f}\sum_{i=1}^{N_f}\left[B^{\rm th}(f_i) + \int_0^{f_i} x_{n={\rm N}}(f')df' - \int_0^{f_i} x_d(f')df' + k_{\rm B}T\log k_{N\leftarrow U}(f_i)\right] - k_{\rm B}T\log k_0. \tag{E.7}$$

Unknown elastic properties

The χ^2 function defined in Eq. (E.5) depends on the elastic properties of the molecule under study. This is because both terms $B^{\rm th}_{NU}(f)$ and $\Delta G_{NU}(f)$ contain an energetic contribution due to the release of single-stranded nucleic acid upon unfolding. In case the elastic properties are unknown, the χ^2 function can also be minimized for the persistence length P. Hence, the trio of values of P, $\log k_0^{NU}$ and ΔG^0_{NU} that satisfy:

$$\frac{\partial\chi^2}{\partial P} = 0, \qquad \frac{\partial\chi^2}{\partial \log k_0^{NU}} = 0, \qquad \frac{\partial\chi^2}{\partial \Delta G^0_{NU}} = 0, \tag{E.8a,b,c}$$

give the best match between the theoretical and the experimental profile of the kinetic barrier when elastic properties are unknown.

The solution to the set of equations (E.8a,b,c) can be easily found numerically. The methodology followed is:

i. Select a range of reasonable values for the persistence length P.

ii. For each value of P, compute ΔG^0_{NU} and $\log k_0^{NU}$ according to Eq. (E.6).

iii. For each trio of values of P, ΔG^0_{NU} and $\log k_0^{NU}$, calculate the numerical value of the χ^2 function (Eq. E.5).

iv. Find the trio of values P, ΔG^0_{NU} and $\log k_0^{NU}$ such that χ^2 has an absolute minimum.

In Fig. E.1, the numerical evaluation of χ^2 as a function P for hairpins M15, U20 and N08 is shown. The result depends on the set of basepair free energies used to calculate $B^{\rm th}_{NU}(f)$ (UO versus UNZ set, Table 4.1). Table E.1 contains results obtained for the three hairpins. Numerical values are the average over results obtained using the two different sets of basepair free energies. Note that the estimations for the persistence length are in good agreement with the values determined in Chapter 3.

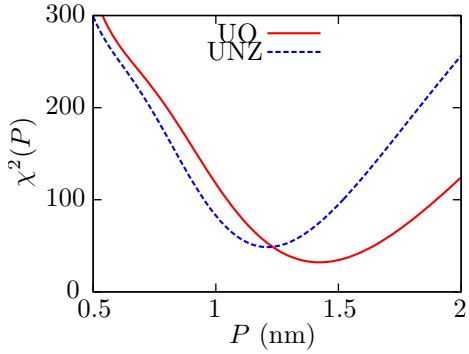

Figure E.1: χ^2 versus persistence length P. Red-solid line was obtained by using the UO set of basepair free energies from [Sant 98] for the evaluation of $B^{\text{th}}(f)$ and blue-dashed line was obtained by using the UNZ values from [Hugu 10b]. Result for hairpins M15 (**a**), U20 (**b**) and N08 (**c**).

Hairpin	P (nm)	$\log k_0$	ΔG^0_{NU} ($k_{\text{B}}T$)	χ^2
M15	1.3±0.1	8.0±0.5	80±2	40±8
U20	1.3±0.2	9.1±0.1	85±3	93±8
N08	1.6±0.1	7.8±0.5	68±2	80±8

Table E.1: Results for $\log k_0$, ΔG^0_{NU} and P obtained by minimization of χ^2. The value is the average over the two estimations obtained using the UO set of parameters [Sant 98] or the set of parameters derived from unzipping experiments [Hugu 10b]. Error-bars are standard errors of the such average.

Other unknown parameters

In the determination of the effect of monovalent ions in the thermodynamic properties of an RNA hairpin (Section 5.2) the theoretical expression of the kinetic barrier $B_{NU}(f)$ depends on the constant m, introduced in Eq. (5.4). In this case, the trio of values m, $\log k_0^{NU}$ and ΔG^0_{NU} can be determined by matching the theoretical and the experimental estimations of the kinetic barrier. Again the χ^2 function can be minimized to find the best match according to:

$$\frac{\partial \chi^2}{\partial m} = 0, \qquad \frac{\partial \chi^2}{\partial \log k_0^{NU}} = 0, \qquad \frac{\partial \chi^2}{\partial \Delta G^0_{NU}} = 0. \tag{E.9a,b,c}$$

The same procedure can be performed when the kinetic barrier $B_{NU}(f)$, or the free energy $\Delta G_{NU}(f)$, or both parameters, depend on an unknown constant $\mathcal{C}$ (such as the binding free energy g_{bind} in Section 5.4). The key idea is to minimize the χ^2 function for $\mathcal{C}$, together with the attempt frequency and the free energy of formation at zero force:

$$\frac{\partial \chi^2}{\partial \mathcal{O}} = 0, \qquad \frac{\partial \chi^2}{\partial \log k_0^{NU}} = 0, \qquad \frac{\partial \chi^2}{\partial \Delta G^0_{NU}} = 0. \tag{E.10a,b,c}$$

Appendix F

UV absorbance experiments in a short RNA hairpin

The melting profile of CD4-RNA hairpin is determined through UV absorbance at 260 nm in a buffer containing 100 mM Tris.HCl, 1 mM EDTA, without NaCl or $MgCl_2$. The melting temperature is measured at 70±1℃(dashed line in Fig. F.1a-b). The first derivative of the absorbance as a function of temperature reveals several maxima along the resulting profile (arrows in Fig. F.1b), which denote the presence of pre-melted states. For instance, regions with a richer A-U content in the middle of the stem may dissociate before the whole hairpin is unfolded (Fig. F.1c). This result invalidates the two-state assumption used to extract thermodynamic parameters from the melting curve.

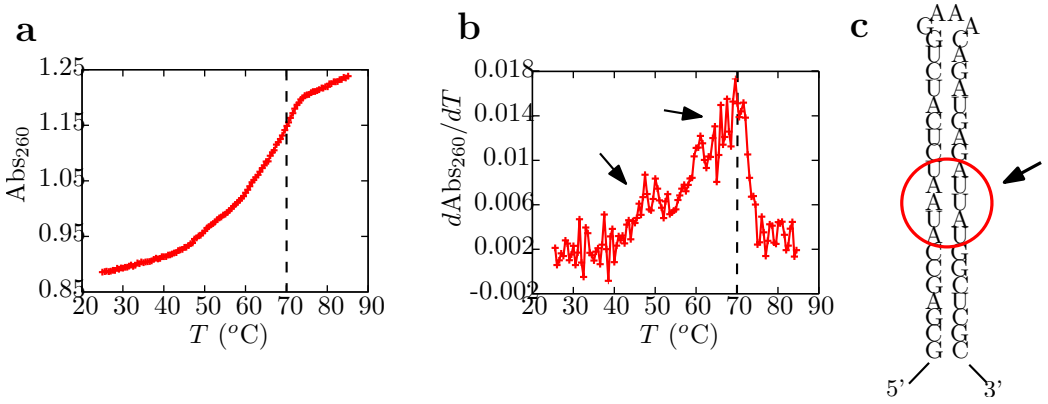

Figure F.1: UV Absorbance of the RNA hairpin. **a.** Experimental results of the absorbance of our RNA hairpin as a function of temperature. **b.** First derivative of the absorbance as a function of temperature. Its maximum (black dashed line) defines the melting temperature. **c.** CD4-RNA hairpin has a region with a high A-U content in the stem which could lead to premelted states.

By adding 100 mM NaCl to the buffer, the melting temperature is too high and the sample starts boiling and evaporating before any relevant signal can be obtained. Therefore, for CD4-RNA melting curves cannot be measured at different ionic conditions. Here, DFS provide an alternative and powerful methodology.

Appendix G

Ions and counterions

A nucleic acid molecule (dsDNA, ssDNA and RNA) can be sketched as a negatively charged linear filament. Positively charged ions in solution tend to surround the molecule and screen its charge. Such interactions may be weak but are long-ranged, and play a crucial role in the spatial distribution of ions that reflects the balance between thermal fluctuations and electrostatic interactions.

In what follows, different theories to address the interaction between DNA and ions in solution are presented.

G.1 Poisson–Boltzmann theory and Debye length

The Poisson–Boltzmann equation combines the Poisson equation of electrostatics with the equilibrium Boltzmann distribution [Jack 06].

Suppose a solution made of a collection N different types of ions with charge q_i and local concentration $c_i(\vec{r})$ expressed in molar units ($i = 1, \ldots, N$). The charged density can be expressed as:

$$\rho(\vec{r}) = \sum_{i=1}^{N} q_i N_{\mathrm{A}} c_i(\vec{r}), \tag{G.1}$$

where N_{A} is the Avogadro's number and $\vec{r}$ is the distance to some charged object of interest, such as an ion, a biopolymer, a membrane, etc. Under equilibrium conditions it can be assumed that the distribution of ions satisfies the Boltzmann distribution. As a result:

$$c_i(\vec{r}) = c_i^{\infty} \exp\left(-\frac{q_i V(\vec{r})}{k_{\mathrm{B}} T}\right), \tag{G.2}$$

being c_i^{∞} is the total concentration of ions in the solution and where the potential energy $V(\vec{r})$ satifies $V(\vec{r} \to \infty) \to 0$.

Given a charge distribution $\rho(\vec{r})$, the Poisson equation states that the electrical potential $V(\vec{r})$ satisfies:

$$\nabla^2 V(\vec{r}) = -\frac{\rho(\vec{r})}{\varepsilon}, \tag{G.3}$$

where ε is a dielectric constant. Inserting the charged distribution given in Eq. (G.1) into the Poisson equation (Eq. G.3) one gets the Poisson–Boltzmann equation:

$$\nabla^2 V(\vec{r}) = -\frac{1}{\varepsilon}\sum_{i=1}^{N} q_i N_\mathrm{A} c_i^\infty \exp\left(-\frac{q_i V(\vec{r})}{k_\mathrm{B}T}\right). \tag{G.4}$$

A first step to solve Eq. (G.4) is to consider that $q_i V(\vec{r})/k_\mathrm{B}T \ll 1$. Hence:

$$\nabla^2 V(\vec{r}) = -\frac{1}{\varepsilon}\sum_{i=1}^{N} q_i N_\mathrm{A} c_i^\infty \left(1 - \frac{q_i V(\vec{r})}{k_\mathrm{B}T}\right) \tag{G.5a}$$

$$= \frac{1}{\varepsilon}\sum_{i=1}^{N} \frac{N_\mathrm{A} c_i^\infty q_i^2}{k_\mathrm{B}T} V(\vec{r}), \tag{G.5b}$$

where the solution is assumed to be neutral ($\sum_{i=1}^{N} q_i c_i^\infty = 0$). The charge of each type of ion can be rewritten as $q_i = z_i e$, being z_i the valence and e the charge unit. In addition, the ionic strength is defined as $I = \frac{1}{2}\sum_i z_i^2 c_i^\infty$. Hence, Eq. (G.5a) becomes:

$$\nabla^2 V(\vec{r}) = \frac{1}{\lambda_D^2} V(\vec{r}), \tag{G.6}$$

which is known as the linearized version of the Poisson-Boltzmann equation, and where λ_D is known as the Debye screening length,

$$\lambda_D = \sqrt{\frac{\varepsilon k_\mathrm{B}T}{2N_\mathrm{A}e^2 I}}. \tag{G.7}$$

λ_D is a fundamental length in ionic solutions.

G.1.1 Poisson–Boltzmann equation for a single charge: Debye–Hückel approximation

Suppose a negative ion with charge $-q$ and radius r_0 surrounded by a cloud of positive ions (Fig. G.1a). Equation (G.6) can be solved in the spherical coordinates and the following expression for the potential energy is obtained:

$$V(r) = -\frac{q}{4\pi\varepsilon r} - \frac{q}{4\pi\varepsilon r}\left(\frac{e^{-(r-r_0)/\lambda_D}}{1 + r_0/\lambda_D} - 1\right) \tag{G.8}$$

$$= -\frac{q}{4\pi\varepsilon r}\frac{e^{-(r-r_0)/\lambda_D}}{1 + r_0/\lambda_D}, \tag{G.9}$$

assuming that:

$$V(r \to \infty) \to 0 \qquad\qquad E(r_0) = -\frac{q}{4\pi\varepsilon r_0^2}, \tag{G.10}$$

where $E(r)$ is the electric field, which must be continuous. This result in known as the Debye–Hückel approximation. The first term in the solution given by Eq. (G.8) is the Coulombic part due to the charge in the origin, and the second term is the ionic screening contribution. The result provided in Eq. (G.8) allows us to calculate the reversible work, or free energy, required to place a charge $-q$ to the surface of a sphere of radius r_0 as:

$$W = \int_0^{-q} V(r_0) dq' \tag{G.11a}$$

$$= \frac{1}{4\pi\varepsilon r_0} \int_0^{-q} q' dq' + \frac{1}{4\pi\varepsilon r_0} \left(\frac{1}{1 + r_0/\lambda_D} - 1 \right) \int_0^{-q} q' dq' \tag{G.11b}$$

$$= \frac{1}{4\pi\varepsilon r_0} \frac{q^2}{2} + \frac{1}{4\pi\varepsilon r_0} \left(\frac{1}{1 + r_0/\lambda_D} - 1 \right) \frac{q^2}{2} \tag{G.11c}$$

$$= \frac{1}{4\pi\varepsilon r_0} \frac{q^2}{2} - \frac{1}{4\pi\varepsilon} \frac{1}{\lambda_D + r_0} \frac{q^2}{2}. \tag{G.11d}$$

The first term in Eq. (G.11d) is independent of the concentration of ions in the solution. The second provides the dependence of the reversible work W with the concentration of ions. By assuming $\lambda_D \gg r_0$:

$$W \propto \frac{1}{\lambda_D} \propto I^{-1/2}. \tag{G.12}$$

It is important to stress that the Debye-Hückel theory is only exact in the limit of very low concentration of ions (*i. e.*, in the dilute limit). Otherwise, the approximation used to linearize the Poisson-Boltzmann equation might not be valid. In addition, the theory is a mean field approach that does not account for correlations between the ions in solution.

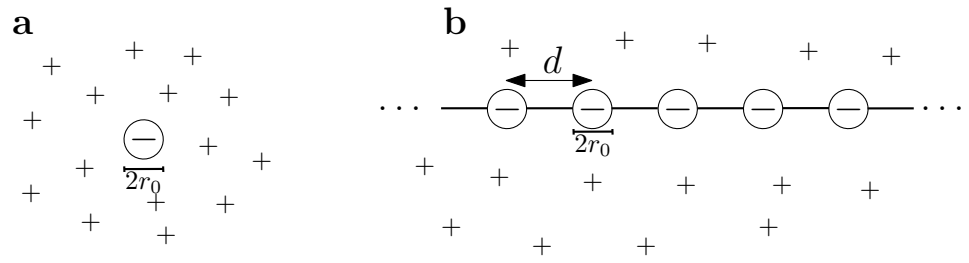

Figure G.1: Charge distribution in ionic solutions. a. Representation of a negative ion with charge $-q$ and radius r_0 surrounded by a cloud of positive ions. **b.** Linear charges of negative charges of radius r_0 separated by a constant distance d and surrounded by a cloud of positive ions.

G.1.2 Poisson–Boltzmann equation for a linear charge distribution: Debye–Hückel screening

Nucleic acid molecules can be treated as a linear charge density. The own repulsion endows dsDNA and dsRNA molecules with its large persistence length. For simplicity,

the nucleic acid molecules will be treated as a linear chain of charges of radius r_0 separated by a distance d (Fig. G.1). Suppose now we work in the dilute limit (that is, $q_i V(\vec{r})/k_{\mathrm{B}}T \ll 1$) and take $\lambda_d \gg r_0$. First, the potential energy at the surface of one of the charges of the molecule is:

$$V(r) = -2\sum_{n>0} \frac{q}{4\pi\varepsilon r_n} \frac{e^{-(r_n - r_0)/\lambda_D}}{1 + r_0/\lambda_D} \tag{G.13}$$

$$\simeq -2\sum_{n>0} \frac{q}{4\pi\varepsilon nd} e^{-nd/\lambda_D}, \tag{G.14}$$

where has been assumed that $d \gg r_0$, n is a integer that enumerates the charges on the right of the charge that has been taking as the origin, and the factor 2 in front of the sum accounts for the charges on the left. The molecule can be assumed to be infinitely large and, as a result:

$$V(r) = -2\sum_{n=1}^{\infty} \frac{q}{4\pi\varepsilon nd} e^{-nd/\lambda_D} \tag{G.15a}$$

$$= -2\frac{q}{4\pi\varepsilon} \int_d^{\infty} \frac{e^{-x/\lambda_D}}{x} \frac{dx}{d} \tag{G.15b}$$

$$= -2\frac{q}{4\pi\varepsilon d} \int_{d/\lambda_D}^{\infty} \frac{e^{-x}}{x} dx. \tag{G.15c}$$

For $d \ll \lambda_D$ this can be approximated to:

$$V(r) = 2\frac{q}{4\pi\varepsilon d} \log\left(\frac{d}{\lambda_D}\right). \tag{G.16}$$

Thus, the free energy required to place a charge q on the surface of infinite linear distribution of charges (or, equivalently, a charged polymer) is equal to:

$$W = 2\frac{1}{4\pi\varepsilon d} \log\left(\frac{d}{\lambda_D}\right) \int_0^q q' dq' \tag{G.17}$$

$$= -\frac{q^2}{4\pi\varepsilon d} \log\left(\frac{d}{\lambda_D}\right). \tag{G.18}$$

Hence [Mann 72]:

$$W \propto \log \lambda_D \propto -\log I. \tag{G.19}$$

Therefore, in the limit of low dilutions, the free energy to place a charge close to a DNA molecule surrounded by Na^+ ions, satisfies:

$$W \propto -\log[\mathrm{Na}^+], \tag{G.20}$$

since $I = [\mathrm{Na}^+]/2$. Therefore, it has been here shown that the presence of monovalent ions $[\mathrm{Na}^+]$ in a solution has a logarithmic and stabilizing effect on the thermodynamic

stability of an RNA double helix present in the solution. This justifies the choice of $m \log[\mathrm{Na}^+]$ to take into account the effect of the monovalent ions in the free energy of formation of a single base pair, being m a constant. It is important to stress that this result is only valid at low ionic concentrations.

G.2 Counterion condensation theory

When a polyelectrolyte has a high charge density, it attracts ions very strongly. Under this circumstance, a new form of association called "counterion-condensation" (CC) takes place [Mann 78]. The key idea is to treat DNA as a continuous uniformly and negatively charged rigid rod and find the conditions at which the ion-polyelectrolyte contribution to the partition function converges. A key parameter is given by:

$$\xi = \left| \frac{qe}{\varepsilon d k_{\mathrm{B}} T} \right|, \tag{G.21}$$

where q/d is the linear charge density of the polyelectrolyte, and e is the unit charge. When $\xi \geq 1/Z$ (being Z the valence of the counterion, $Z = 1$ for Na^{+1} and $Z = 2$ for Mg^{+2}) the counterion-polyelectrolyte system has a very high free energy and is unstable. In contrast, for $\xi < 1/Z$ the polyelectrolyte is screened by the counterions and the Debye-Hückel screening effect properly describes the system. Hence, when $\xi \geq 1/Z$, ions in the solution bind to the molecule and increase the effective mean spacing between the charges of the polyelectrolyte from d to d_{eff} until its charge density is screened (and $\xi_{\mathrm{eff}} < 1$) and the Poisson-Boltzmann theory can be applied. Interestingly, the fraction of bound ions in this condensation process is independent of the bulk concentration [Jack 06].

In reference [Mann 02], S. G. Manning presents analytical formulas for the free energy of formation of nucleic acid molecules based on the CC theory that extend the standard line charge model to helical and double helical charge arrays. The polyelectrolyte free energy, ΔG_{dh}, can be written in the form:

$$\Delta G_{\mathrm{dh}} = \Delta G_l - \frac{2N k_{\mathrm{B}} T}{Z} \left(2 - \frac{1}{Z\xi}\right) \left[\log \left(\frac{1 - e^{-2d/\lambda_D}}{1 - e^{-d/\lambda_D}} \right) - \Omega_{\mathrm{dh}} \right], \tag{G.22}$$

where N is the number of base pairs along the double helix; λ_D is the Debye length (Eq. G.7); d is the uniform spacing of a line of $2N$ charges, equal to one half of the interphosphate distance for B-dsDNA ($d = 1.688$ Å); and $\xi = 4.23$, also for B-dsDNA.

The term Ω_{dh} contains information about the polyelectrolyte being a double helix, and is calculated as:

$$\Omega_{\mathrm{dh}} = S_0 + S_1 + S_2 + S_3, \tag{G.23}$$

where

$$S_0 = \frac{1}{2}\frac{d}{r_{00'}} e^{-r_{00'}/\lambda_D} \tag{G.24a}$$

$$S_1 = \frac{1}{2}\sum_{n=1}^{\infty}\left[2d\frac{e^{-r_n/\lambda_D}}{r_n} - \frac{e^{-2nd/\lambda_D}}{n}\right] \tag{G.24b}$$

$$S_2 = \frac{1}{4}\sum_{n=1}^{\infty}\left[2d\frac{e^{-r_{0n'}/\lambda_D}}{r_{0n'}} - \frac{e^{-2nd/\lambda_D}}{n}\right] \tag{G.24c}$$

$$S_3 = \frac{1}{4}\sum_{n=1}^{\infty}\left[2d\frac{e^{-r_{0-n'}/\lambda_D}}{r_{0-n'}} - \frac{e^{-2nd/\lambda_D}}{n}\right], \tag{G.24d}$$

and

$$r_n = 2nd\sqrt{1 + \frac{a^2}{2n^2d^2}\left[1 - \cos\left(\frac{2nd}{h}\right)\right]} \tag{G.25a}$$

$$r_{0n'} = 2nd\sqrt{\left(1 + \frac{\Delta\delta}{2nd}\right)^2 + \frac{a^2}{2n^2d^2}\left[1 - \cos\left(\frac{2nd}{h} + \Delta\alpha\right)\right]} \tag{G.25b}$$

$$r_{0-n'} = 2nd\sqrt{\left(1 - \frac{\Delta\delta}{2nd}\right)^2 + \frac{a^2}{2n^2d^2}\left[1 - \cos\left(\frac{2nd}{h} - \Delta\alpha\right)\right]}, \tag{G.25c}$$

where a is the radius of the helix, h is the increment of the axial coordinate per radian change of rotation angle, and $\Delta\delta$ and $\Delta\alpha$ are a vertical displacement and a phase angle, respectively, that have definite numerical values given in crystallographic tables. For B-dsDNA, $a = 9.225$ Å, $h = 5.364$ Å, $\Delta\delta = 3.696$ Å, $\Delta\alpha = -3.218$ rad.

The term ΔG_l equals the free energy of formation of a linear distribution of charges, *i. e.* it is the reversible work required to assemble the N charges on the line from their isolated reference state at infinity. It satisfies:

$$\Delta G_l = 2Nk_{\mathrm{B}}T\left[-\frac{1}{Z}\left(2 - \frac{1}{Z\xi}\right)\log\left(1 - e^{-d/\lambda_D}\right) - \frac{1}{Z} + \frac{1}{Z^2\xi}\right]. \tag{G.26}$$

Equation (G.26) is an original result of the CC theory [Mann 78, Mann 96].

G.3 Tightly Bound Ion model

The Tightly Bound Ion (TBI) model for mixed ion solutions treats monovalent ions as ionic background and multivalent ions as responsible from ion–ion correlation effects, and takes into account only non-sequence-specific electrostatic effects of ions on nucleic acid hairpins [Tan 05, Tan 07, Tan 08].

In this section the empirical equations of the TBI model [Tan 05, Tan 06] used to predict the hairpin free energies at different ionic conditions as presented in Section 5.2 are summarized [Tan 07, Tan 08].

Bases in the loop	N_l=4
Bases in the helix stem	N=20
Hairpin diameter	d=1.7 nm
Interphosphate distance	a=0.6 nm
Enthalpy at 1 M [Mon^+], 0 M [Mg^{2+}]	ΔH_0=199 kcal/mol
Entropy at 1 M [Mon^+], 0 M [Mg^{2+}]	ΔS_0=527.16 mkcal/Kmol

The external parameters, obtained from the Mfold server [Walt 94, Math 99, Zuke 03], are:

In what follows x is the concentration of monovalent salt and y is the concentration of magnesium ions. Both parameters are given in units of M. Temperature T is given in Kelvin.

The empirical set of equations are given in what follows.

In presence of monovalent ions:

In presence of monovalent ions, the free energy of formation of a hairpin reads as:

$$G_{\rm Mon}(x,T) = G^l_{\rm Mon}(x) + G^h_{\rm Mon}(x,T), \tag{G.27}$$

where $G^l_{\rm Mon}(x)$ is the free energy of formation of the loop and $G^h_{\rm Mon}(x,T)$ is the free energy of formation of the double helix. The contribution of the loop satisfies:

$$G^l_{\rm Mon}(x) = -(\ln Z^l_{\rm Mon}(x) - \ln Z^c_{\rm Mon}(x)), \tag{G.28}$$

where

$$\ln Z^l_{\rm Mon}(x) = a^l_1(x)\log(N_l - a/d + 1) + b^l_1(x)(N_l - a/d + 1)^2 - b^l_1(x) \tag{G.29a}$$
$$\ln Z^c_{\rm Mon}(x) = c^l_1(x)N_l - d^l_1(x) \tag{G.29b}$$
$$a^l_1(x) = (0.02N_l - 0.026)\log(x) + 0.54N_l + 0.78 \tag{G.29c}$$
$$b^l_1(x) = \left(-\frac{0.01}{(N_l+1)} + 0.006\right)\log(x) - \frac{7}{(N_l+1)^2} - 0.01 \tag{G.29d}$$
$$c^l_1(x) = 0.07\log(x) + 1.8 \tag{G.29e}$$
$$d^l_1(x) = 0.21\log(x) + 1.5. \tag{G.29f}$$

The contribution of the hairpin satisfies:

$$G^h_{\rm Mon}(x,T) = H_0 - TS_{\rm Mon}(x)0.001, \tag{G.30}$$

where the factor 0.001 in the last term account for the different units given for the enthalpy H_0 and the entropy $S_{\mathrm{Mon}}(x)$, and:

$$S_{\mathrm{Mon}}(x) = S_0 - 3.22(N-1)g_1(x) \tag{G.31a}$$
$$g_1(x) = a_1^h(x) + b_1^h(x)/N \tag{G.31b}$$
$$a_1^h(x) = -0.075\log(x) + 0.012\log^2(x) \tag{G.31c}$$
$$b_1^h(x) = 0.018\log^2(x). \tag{G.31d}$$

In presence of only magnesium ions:

When only magnesium ions are contained in the solution surrounding the nucleic acid molecule, the free energy of formation is given by:

$$G_{\mathrm{Mg}}(y,T) = G_{\mathrm{Mg}}^l(y) + G_{\mathrm{Mg}}^h(y,T), \tag{G.32}$$

where $G_{\mathrm{Mg}}^l(y)$ is the energetic contribution of the loop and $G_{\mathrm{Mg}}^h(y,T)$ is the one coming from the hairpin double helix. The first satisfies:

$$G_{\mathrm{Mg}}^l(y) = -(\ln Z_{\mathrm{Mg}}^l(y) - \ln Z_{\mathrm{Mg}}^c(y)), \tag{G.33}$$

where

$$\ln Z_{\mathrm{Mg}}^l(y) = a_2^l(y)\log(N_l - a/d + 1) + b_2^l(y)(N_l - a/d + 1)^2 - b_2^l(y) \tag{G.34a}$$
$$\ln Z_{\mathrm{Mg}}^c(y) = c_2^l(y)N_l - d_2^l(y) \tag{G.34b}$$
$$a_2^l(y) = \left(-\frac{1}{N_l+1} + 0.32\right)\log(y) + 0.7N_l + 0.43 \tag{G.34c}$$
$$b_2^l(y) = 0.0002(N_l+1)\log(y) - 5.9/(N_l+1)^2 - 0.003 \tag{G.34d}$$
$$c_2^l(y) = 0.067\log(y) + 2.2 \tag{G.34e}$$
$$d_2^l(y) = 0.163\log(y) + 2.53 \tag{G.34f}$$
$$\tag{G.34g}$$

On the other hand, the energetic contribution due to double helix formation is:

$$G_{\mathrm{Mg}}^h(y,T) = H_0 - TS_{\mathrm{Mg}}(y)0.001, \tag{G.35}$$

where

$$S_{\mathrm{Mg}}(y) = S_0 - 3.22(N-1)g_2(y) \tag{G.36a}$$
$$g_2(y) = a_2^h(y) + b_2^h(y)/N^2 \tag{G.36b}$$
$$a_2^h(y) = -0.6/N + 0.025\log(y) + 0.0068\log^2(y) \tag{G.36c}$$
$$b_2^h(y) = \log(y) + 0.38\log^2(y) \tag{G.36d}$$

Mixed monovalent/magnesium ionic conditions:

In presence of both monovalent and magnesium ions in the solution, the free energy of formation of the nucleic acid hairpin is given by:

$$G_{\mathrm{Mon,Mg}}(x,y,T) = G^{l}_{\mathrm{Mon,Mg}}(x,y) + G^{h}_{\mathrm{Mon,Mg}}(x,y,T), \tag{G.37}$$

where $G^{l}_{\mathrm{Mon,Mg}}(x,y)$ and $G^{h}_{\mathrm{Mon,Mg}}(x,y,T)$ are the energetic contributions due to loop and helix formation respectively. In the case of the loop formation, it is satisfied:

$$G^{l}_{\mathrm{Mon,Mg}}(x,y) = x^{l}_{1}(x,y)G^{l}_{\mathrm{Mon}}(x) + (1 - x^{l}_{1}(x,y))G^{l}_{\mathrm{Mg}}(y), \tag{G.38}$$

where

$$x^{l}_{1}(x,y) = x/(x + (7.2 - 20/N_l)(40 - \log(x))y). \tag{G.39}$$

For the case of the double helix:

$$G^{h}_{\mathrm{Mon,Mg}}(x,y,T) = H_0 - TS_{\mathrm{Mon,Mg}}(x,y)0.001, \tag{G.40}$$

where

$$\begin{aligned} S_{\mathrm{Mon,Mg}}(x,y) = S_0 - 3.22\big[&(N-1)\left(x^{h}_{1}(x,y)g_1(x) + (1 - x^{h}_{1}(x,y))g_2(y)\right) \\ &+ g_{1,2}(x,y)\big] \end{aligned} \tag{G.41a}$$

$$x^{h}_{1}(x,y) = \frac{x}{x + (8.1 - 32.4/N)(5.2 - \log(x))y} \tag{G.41b}$$

$$g_{1,2}(x,y) = -0.6x^{h}_{1}(x,y)(1 - x^{h}_{1}(x,y))\log(x)\log((1/x^{h}_{1}(x,y) - 1)x)/N \tag{G.41c}$$

Appendix H

The chemical reaction of binding processes

From the point of view of chemical reactions, the binding of a ligand (L) to a DNA hairpin (H) can be written as:

$$\mathrm{H} + \mathrm{L} \rightleftarrows \mathrm{B}, \tag{H.1}$$

where B denotes the bound state.

The concentrations of each chemical specie satisfy the following differential equations:

$$\frac{d[\mathrm{H}]}{dt} = -k_{\rightarrow}[\mathrm{H}][\mathrm{L}] + k_{\leftarrow}[\mathrm{B}] \tag{H.2a}$$

$$\frac{d[\mathrm{L}]}{dt} = -k_{\rightarrow}[\mathrm{H}][\mathrm{L}] + k_{\leftarrow}[\mathrm{B}] \tag{H.2b}$$

$$\frac{d[\mathrm{B}]}{dt} = +k_{\rightarrow}[\mathrm{H}][\mathrm{L}] - k_{\leftarrow}[\mathrm{B}] \tag{H.2c}$$

For convenience, the probability to be in the bounded state of the system described by the chemical reaction in Eq. (H.1) is defined:

$$p_{\mathrm{B}} = \frac{[\mathrm{B}]}{[\mathrm{B}] + [\mathrm{H}]}, \tag{H.3}$$

and according to the set of chemical reactions written in Eq. (H.2), its time evolution satisfies:

$$\frac{dp_{\mathrm{B}}}{dt} = \frac{1}{[\mathrm{H}] + [\mathrm{B}]} \left(+k_{\rightarrow}[\mathrm{H}][\mathrm{L}] - k_{\leftarrow}[\mathrm{B}]\right) \tag{H.4a}$$

$$= +k_{\rightarrow}[\mathrm{L}]\left(1 - p_{\mathrm{B}}\right) - k_{\leftarrow}p_{\mathrm{B}}. \tag{H.4b}$$

When the system is in equilibrium, $dp_{\mathrm{B}}/dt = 0$ and the probability p_{B} can be written as:

$$p_{\mathrm{B}} = \frac{1}{1 + \dfrac{k_{\leftarrow}}{k_{\rightarrow}[\mathrm{L}]}} \tag{H.5a}$$

$$= \frac{1}{1 + \dfrac{K_{\mathrm{eq}}}{[\mathrm{L}]}}, \tag{H.5b}$$

where K_{eq} is the equilibrium constant of the binding chemical reaction (Eq. H.1).

The equilibrium binding probability can also be obtained from the point of view of statistical mechanics. The binding reaction of a single ligand molecule to a DNA hairpin has the following grand-partition function:

$$Z_{GC} = \sum_{N=0}^{\infty} z^N \sum_{\Omega_N} e^{-\frac{E(\Omega_N)}{k_{\mathrm{B}}T}} \tag{H.6a}$$

$$= 1 + ze^{-\beta\varepsilon}, \tag{H.6b}$$

where $\beta = 1/k_{\mathrm{B}}T$. The first term in Eq. (H.6b) corresponds to the situation where no ligand is bound to the hairpin (left side of the chemical reaction given by Eq. H.1); the second term accounts for the bounded state, where a ligand molecule binds to the hairpin with an energy ε. There are no extra terms because it is considered that only one ligand molecule can bind the hairpin at a time.

The binding probability in equilibrium can be computed as the ratio of the grand-canonical partition function restricted to the bound configuration over the total grand-partition function (Eq. H.6b):

$$p_{\mathrm{B}} = \frac{ze^{-\beta\varepsilon}}{Z_{GC}} = \frac{ze^{-\beta\varepsilon}}{1 + ze^{-\beta\varepsilon}} \tag{H.7}$$

Hence, by equaling Eqs. (H.5b) and (H.7) the following identity is obtained:

$$z^{-1}e^{\beta\varepsilon} = \frac{K_{\mathrm{eq}}}{[\mathrm{L}]} \tag{H.8}$$

or, equivalently:

$$\log\left(\frac{K_{\mathrm{eq}}}{[\mathrm{L}]}\right) = \beta(\varepsilon - \mu), \tag{H.9}$$

which gives the definition for the chemical potential of the ligand:

$$\mu = \varepsilon + k_{\mathrm{B}}T \log\left(\frac{K_{\mathrm{eq}}}{[\mathrm{L}]}\right). \tag{H.10}$$

Appendix I

Fluctuation relations: mathematical demonstrations

I.1 The Crooks fluctuation relation

Consider a thermodynamic system, initially set to equilibrium conditions, to which an experimental protocol $\lambda(t)$ is applied. λ denotes the control parameter, and its time evolution is discretized for the sake of simplicity as:

$$\lambda(t) = \{\lambda_0, \lambda_1, \ldots, \lambda_i, \ldots, \lambda_N\}, \tag{I.1}$$

being the subindex i ($i = 1, \ldots, N$) a time label, such that $i\Delta t$ indicates the time and $\tau = N\Delta t$ is equal to the time-duration of the protocol. The time reversed protocol of $\lambda(t)$, defined as $\tilde{\lambda}(t) = \lambda(\tau - t)$, satisfies:

$$\tilde{\lambda}(t) = \{\tilde{\lambda}_0, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_i, \ldots, \tilde{\lambda}_N\} \tag{I.2a}$$

$$= \{\lambda_N, \lambda_{N-1}, \ldots, \lambda_{N-i}, \ldots, \lambda_0\}. \tag{I.2b}$$

During the protocol $\lambda(t)$ –hereafter referred to as forward protocol– the system explores different configurations $\mathcal{C}_i$ accessible in its phase space $\mathcal{S}$. Thus, a trajectory Γ of the system evolving under such protocol can be characterized by the sequence of sampled configurations:

$$\Gamma = \{\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_i, \ldots, \mathcal{C}_N\}. \tag{I.3}$$

Similarly, the time-reversed trajectory $\tilde{\Gamma}$ can be defined as:

$$\tilde{\Gamma} = \{\tilde{\mathcal{C}}_0, \tilde{\mathcal{C}}_1, \ldots, \tilde{\mathcal{C}}_i, \ldots, \tilde{\mathcal{C}}_N\} \tag{I.4a}$$

$$= \{\mathcal{C}_N, \mathcal{C}_{N-1}, \ldots, \mathcal{C}_{N-i}, \ldots, \mathcal{C}_0\}. \tag{I.4b}$$

For a given value of λ_i, the equilibrium probability to sample a given configuration $\mathcal{C}_i$ is given by the Gibbsian equilibrium distribution function,

$$P^{\mathrm{eq}}_{\lambda_i}(\mathcal{C}_i) = \frac{e^{-\beta E_{\lambda_i}(\mathcal{C}_i)}}{Z_{\lambda_i}}, \tag{I.5}$$

where $\beta = 1/k_{\mathrm{B}}T$, k_{B} is the Boltzmann constant, T is the temperature in Kelvin, $E_{\lambda_i}(\mathcal{C}_i)$ is the energy of the configuration $\mathcal{C}_i$ when $\lambda = \lambda_i$, and Z_{λ_i} is the partition function of the system at $\lambda = \lambda_i$:

$$Z_{\lambda_i} = \sum_{\mathcal{C}_i \in \mathcal{S}} \exp\left(-\frac{E_{\lambda_i}(\mathcal{C}_i)}{k_{\mathrm{B}}T}\right). \tag{I.6}$$

In addition, it is assumed that the system satisfies the following detailed balance condition:

$$\frac{\mathcal{W}_i(\mathcal{C}_i \to \mathcal{C}_{i+1})}{\mathcal{W}_i(\mathcal{C}_{i+1} \to \mathcal{C}_i)} = \exp\left[-\beta\left(E_{\lambda_{i+1}}(\mathcal{C}_{i+1}) - E_{\lambda_{i+1}}(\mathcal{C}_i)\right)\right], \tag{I.7}$$

where $\mathcal{W}_i(\mathcal{C}_i \to \mathcal{C}_{i+1})$ is the transition probability at time i to go to configuration $\mathcal{C}_{i+1}$ from configuration $\mathcal{C}_i$.

Under these assumptions (initially equilibrated system that fulfills detailed balance along the whole trajectory), the probability of the system to follow a given trajectory Γ is equal to:

$$P(\Gamma) = P^{\mathrm{eq}}_{\lambda_0}(\mathcal{C}_0) \prod_{i=0}^{N-1} \mathcal{W}_i(\mathcal{C}_i \to \mathcal{C}_{i+1}), \tag{I.8}$$

whereas the probability of the system to follow the time-reversed trajectory $\tilde{\Gamma}$ is:

$$P(\tilde{\Gamma}) = P^{\mathrm{eq}}_{\tilde{\lambda}_0}(\tilde{\mathcal{C}}_0) \prod_{i=0}^{N-1} \mathcal{W}_i(\tilde{\mathcal{C}}_i \to \tilde{\mathcal{C}}_{i+1}) \tag{I.9a}$$

$$= P^{\mathrm{eq}}_{\lambda_N}(\mathcal{C}_N) \prod_{i=0}^{N-1} \mathcal{W}_i(\mathcal{C}_{N-i} \to \mathcal{C}_{N-i-1}) \tag{I.9b}$$

$$= P^{\mathrm{eq}}_{\lambda_N}(\mathcal{C}_N) \prod_{i=0}^{N-1} \mathcal{W}_i(\mathcal{C}_{i+1} \to \mathcal{C}_i). \tag{I.9c}$$

The ratio between both probabilities is equal to:

$$\frac{P(\Gamma)}{P(\tilde{\Gamma})} = \frac{P^{\mathrm{eq}}_{\lambda_0}(\mathcal{C}_0) \prod_{i=0}^{N-1} \mathcal{W}_i(\mathcal{C}_i \to \mathcal{C}_{i+1})}{P^{\mathrm{eq}}_{\lambda_N}(\mathcal{C}_N) \prod_{i=0}^{N-1} \mathcal{W}_i(\mathcal{C}_{i+1} \to \mathcal{C}_i)} \tag{I.10a}$$

$$= \frac{e^{-\beta E_{\lambda_0}(\mathcal{C}_0)}}{Z_{\lambda_0}} \frac{Z_{\lambda_N}}{e^{-\beta E_{\lambda_N}(\mathcal{C}_N)}} \prod_{i=0}^{N-1} e^{-\beta\left(E_{\lambda_{i+1}}(\mathcal{C}_{i+1}) - E_{\lambda_{i+1}}(\mathcal{C}_i)\right)} \tag{I.10b}$$

$$= \frac{Z_{\lambda_N}}{Z_{\lambda_0}} e^{-\beta\left[E_{\lambda_0}(\mathcal{C}_0) + \sum_{i=0}^{N-1}\left(E_{\lambda_{i+1}}(\mathcal{C}_{i+1}) - E_{\lambda_{i+1}}(\mathcal{C}_i)\right) - E_{\lambda_N}(\mathcal{C}_N)\right]} \tag{I.10c}$$

$$= \frac{Z_{\lambda_N}}{Z_{\lambda_0}} e^{-\beta \sum_{i=0}^{N-1}\left(E_{\lambda_i}(\mathcal{C}_i) - E_{\lambda_{i+1}}(\mathcal{C}_i)\right)}. \tag{I.10d}$$

By identifying the work for a trajectory exerted upon the system along the protocol $\lambda(t)$ as:

$$W(\Gamma) = \sum_{i=0}^{N-1} \left(E_{\lambda_{i+1}}(\mathcal{C}_i) - E_{\lambda_i}(\mathcal{C}_i)\right), \tag{I.11}$$

and the logarithm of the partition function with the free energy of the system:

$$F_{\lambda_i} = -k_{\mathrm{B}}T \log Z_{\lambda_i}, \tag{I.12}$$

it is finally obtained that:

$$\frac{P(\Gamma)}{P(\tilde{\Gamma})} = e^{-\beta\Delta F} e^{\beta W(\Gamma)}, \tag{I.13}$$

or equivalently,

$$P(\Gamma)e^{-\beta W(\Gamma)} = P(\tilde{\Gamma})e^{-\beta\Delta F}, \tag{I.14}$$

where $\Delta F = F_{\lambda_N} - F_{\lambda_0}$.

The average over the protocol $\lambda(t)$ of any observable of the trajectory $\mathcal{F}(\Gamma)$ is calculated according to:

$$\langle \mathcal{F} \rangle_F = \sum_{\Gamma} P(\Gamma)\mathcal{F}(\Gamma), \tag{I.15}$$

where $\langle \ldots \rangle_F$ is the average taken over all the possible trajectories followed by the system along the forward protocol. Thus, by multiplying both sides of Eq. (I.14) by $\mathcal{F}(\Gamma)$ and averaging over all possible trajectories, on gets:

$$\left\langle \mathcal{F}e^{-\beta W} \right\rangle_F = \left\langle \tilde{\mathcal{F}} \right\rangle_R e^{-\beta\Delta F}, \tag{I.16}$$

where $\left\langle \tilde{\mathcal{F}} \right\rangle_R$ denotes the average over reversed trajectories of the time-reversed functional $\tilde{\mathcal{F}}$.

In case $\mathcal{F}(\Gamma)$ is taken as $\delta(W - W(\Gamma))$, Eq. (I.16) becomes:

$$P_F(W)e^{-\beta W} = P_R(-W)e^{-\beta\Delta F}, \tag{I.17}$$

where $P_F(W)$ is the probability density function of work values along the forward protocol and $P_R(-W)$ is the probability density function of work values with a minus sign along the reversed protocol. Finally, Eq. (I.17) can be rewritten in order to get the Crooks Fluctuation Relation (Eq. 8.2):

$$\frac{P_F(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta F}{k_{\mathrm{B}}T}\right). \tag{I.18}$$

Finally, because in single-molecule experiments neither the volume or pressure are manipulated, the variations of the Helmholtz free energy and of the Gibbs free energy are identical, and therefore:

$$\frac{P_F(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta G}{k_{\mathrm{B}}T}\right). \tag{I.19}$$

I.2 The Jarzynski fluctuation relation

In order to obtain the Jarzynski fluctuation relation the functional of the trajectory can be taken equal to 1, *i. e.* $\mathcal{F}(\Gamma) = 1$, and from Eq. (I.16) it is obtained:

$$\left\langle e^{-\beta W} \right\rangle_F = e^{-\beta \Delta F}. \tag{I.20}$$

Again, if ΔF is identified with ΔG the Jarzynski equality (Eq. 8.3) is recovered:

$$\left\langle e^{-\beta W} \right\rangle_F = e^{-\beta \Delta G}. \tag{I.21}$$

However, the Jarzynski fluctuation relation was derived earlier that the Crooks relation. An alternative direct demonstration that does not use the CFR is given in what follows.

Imagine a system with a Hamiltonian $\mathcal{H}(\lambda, \mathcal{C}) = E_\lambda(\mathcal{C})$, being λ the control parameter and $\mathcal{C}$ the configuration of the system. Initially the system is in equilibrium. Under a determined experimental protocol $\lambda(t)$ the system follows a set of configurations $\mathcal{C}(t)$ that may change under different independent repetitions of the same protocol. The work along a given trajectory Γ is (Eq. 8.1):

$$W(\Gamma) = \int_\Gamma dt \frac{d\lambda}{dt} \frac{\partial \mathcal{H}}{\partial \lambda}. \tag{I.22}$$

Despite being the system under under non-equilibrium conditions, it is assumed that detailed balance is satisfied.

For the sake of simplicity, time is discretized. Hence, the protocol and the trajectory may be written as in Eqs. (I.1) and (I.3), respectively. The work is written as in Eq. (I.11) and the system satisfies the dynamics given by Eq. I.7.

The average of the exponential of the work over all possible trajectories is then:

$$\langle \exp(-\beta W) \rangle = \sum_\Gamma \exp(-\beta W(\Gamma))\, P(\Gamma), \tag{I.23}$$

where $P(\Gamma)$ is the probability for a given trajectory and can be expressed as in Eq. (I.8). Therefore:

$$\langle \exp(-\beta W) \rangle = \sum_\Gamma e^{-\beta \sum_{i=0}^{N-1} \left(E_{\lambda_{i+1}}(\mathcal{C}_i) - E_{\lambda_i}(\mathcal{C}_i) \right)} P^{\text{eq}}_{\lambda_0}(\mathcal{C}_0) \prod_{i=0}^{N-1} \mathcal{W}_i(\mathcal{C}_i \to \mathcal{C}_{i+1}) \tag{I.24a}$$

$$= \sum_\Gamma P^{\text{eq}}_{\lambda_0}(\mathcal{C}_0) \prod_{i=0}^{N-1} \frac{e^{-\beta E_{\lambda_{i+1}}(\mathcal{C}_i)}}{e^{-\beta E_{\lambda_i}(\mathcal{C}_i)}} \mathcal{W}_i(\mathcal{C}_i \to \mathcal{C}_{i+1}) \tag{I.24b}$$

$$= \sum_\Gamma P^{\text{eq}}_{\lambda_0}(\mathcal{C}_0) \prod_{i=0}^{N-1} \frac{e^{-\beta E_{\lambda_{i+1}}(\mathcal{C}_i)}}{e^{-\beta E_{\lambda_i}(\mathcal{C}_i)}} \frac{e^{-\beta E_{\lambda_{i+1}}(\mathcal{C}_{i+1})}}{e^{-\beta E_{\lambda_{i+1}}(\mathcal{C}_i)}} \mathcal{W}_i(\mathcal{C}_{i+1} \to \mathcal{C}_i) \tag{I.24c}$$

$$= \sum_\Gamma P^{\text{eq}}_{\lambda_0}(\mathcal{C}_0) \frac{e^{-\beta E_{\lambda_N}(\mathcal{C}_N)}}{e^{-\beta E_{\lambda_0}(\mathcal{C}_0)}} \prod_{i=0}^{N-1} \mathcal{W}_i(\mathcal{C}_{i+1} \to \mathcal{C}_i). \tag{I.24d}$$

Now, using Eqs. (I.5), (I.9) and (I.12):

$$\langle \exp(-\beta W) \rangle = \frac{Z_{\lambda_N}}{Z_{\lambda_0}} \sum_{\Gamma} P^{\mathrm{eq}}_{\lambda_N}(\mathcal{C}_N) \prod_{i=0}^{N-1} \mathcal{W}_i(\mathcal{C}_{i+1} \to \mathcal{C}_i) \tag{I.25a}$$

$$= \frac{e^{-\beta F_{\lambda_N}}}{e^{-\beta F_{\lambda_0}}} \sum_{\Gamma} P(\tilde{\Gamma}) \tag{I.25b}$$

$$= \exp(-\beta \Delta F), \tag{I.25c}$$

the Jarzynski relation is recovered.

I.3 The extended fluctuation relation

In order to demonstrate the extended fluctuation relation (Eq. 9.3) the steps are identical to the ones followed in Section I.1. The only difference is that the the system starts the forward and reversed protocols in *partial* equilibrium conditions, rather than in total equilibrium. A system is in partial equilibrium when its phase space $\mathcal{S}$ can be divided into regions $\mathcal{S}'$ that are partially equilibrated. However, the statistical weights of the different regions $\mathcal{S}'$ do not necessarily follow an equilibrium distributions. Therefore, when a system is in partial equilibrium in λ_i, its configuration $\mathcal{C}_i$ inside $\mathcal{S}'$ is sampled according to the following distribution function:

$$P_{\lambda_i,\mathcal{S}'} = \chi_{\mathcal{S}'}(\mathcal{C}_i) \frac{e^{-\beta E_{\lambda_i}(\mathcal{C}_i)}}{Z_{\lambda_i,\mathcal{S}'}}, \tag{I.26}$$

where $Z_{\lambda_i,\mathcal{S}'}$ is the partition function restricted to the subset $\mathcal{S}'$:

$$Z_{\lambda_i,\mathcal{S}'} = \sum_{\mathcal{C}_i \in \mathcal{S}'} \exp\left(-\frac{E_{\lambda_i}(\mathcal{C}_i)}{k_{\mathrm{B}}T}\right), \tag{I.27}$$

and $\chi_{\mathcal{S}'}(\mathcal{C}_i) = 1$ is $\mathcal{C}_i \in \mathcal{S}'$ and 0 otherwise.

Hence, the probability to sample a given trajectory Γ and its reversed $\tilde{\Gamma}$ (Eqs. I.3 and I.4) are given by:

$$P(\Gamma) = \chi_{\mathcal{S}_0}(\mathcal{C}_0) \frac{e^{-\beta E_{\lambda_0}(\mathcal{C}_0)}}{Z_{\lambda_0,\mathcal{S}_0}} \prod_{i=0}^{N-1} \mathcal{W}_i(\mathcal{C}_i \to \mathcal{C}_{i+1}) \tag{I.28}$$

$$P(\tilde{\Gamma}) = \chi_{\mathcal{S}_N}(\mathcal{C}_N) \frac{e^{-\beta E_{\lambda_N}(\mathcal{C}_N)}}{Z_{\lambda_N,\mathcal{S}_N}} \prod_{i=0}^{N-1} \mathcal{W}_i(\mathcal{C}_{i+1} \to \mathcal{C}_i), \tag{I.29}$$

where $\mathcal{S}_0$ and $\mathcal{S}_N$ are the initial and final subsets of the phase space that contain the initial configuration of the forward and the reversed protocol.

Following the same steps as before, one gets:

$$\frac{P(\Gamma)}{P(\tilde{\Gamma})} = \frac{\chi_{\mathcal{S}_0}(\mathcal{C}_0)\frac{e^{-\beta E_{\lambda_0}(\mathcal{C}_0)}}{Z_{\lambda_0,\mathcal{S}_0}}\prod_{i=0}^{N-1}\mathcal{W}_i(\mathcal{C}_i\to\mathcal{C}_{i+1})}{\chi_{\mathcal{S}_N}(\mathcal{C}_N)\frac{e^{-\beta E_{\lambda_N}(\mathcal{C}_N)}}{Z_{\lambda_N,\mathcal{S}_N}}\prod_{i=0}^{N-1}\mathcal{W}_i(\mathcal{C}_{i+1}\to\mathcal{C}_i)} \tag{I.30a}$$

$$= \frac{\chi_{\mathcal{S}_0}(\mathcal{C}_0)e^{-\beta E_{\lambda_0}(\mathcal{C}_0)}Z_{\lambda_N,\mathcal{S}_N}}{\chi_{\mathcal{S}_N}(\mathcal{C}_N)e^{-\beta E_{\lambda_N}(\mathcal{C}_N)}Z_{\lambda_0,\mathcal{S}_0}}\prod_{i=0}^{N-1}\exp\left[-\beta\left(E_{\lambda_{i+1}}(\mathcal{C}_{i+1})-E_{\lambda_{i+1}}(\mathcal{C}_i)\right)\right] \tag{I.30b}$$

$$= \frac{\chi_{\mathcal{S}_0}(\mathcal{C}_0)Z_{\lambda_N,\mathcal{S}_N}}{\chi_{\mathcal{S}_N}(\mathcal{C}_N)Z_{\lambda_0,\mathcal{S}_0}}e^{\beta W(\Gamma)} \tag{I.30c}$$

Now, the average over the forward trajectories that start in partial equilibrium in a configuration $\mathcal{C}_0 \in \mathcal{S}_0$ and end in $\mathcal{C}_N \in \mathcal{S}_N$ for a functional $\mathcal{F}(\Gamma)$ of the trajectory is computed as:

$$\langle\mathcal{F}\rangle_F^{\mathcal{S}_0\to\mathcal{S}_N} = \frac{\sum_\Gamma P(\Gamma)\mathcal{F}(\Gamma)\chi_{\mathcal{S}_N}(\mathcal{C}_N)}{\sum_\Gamma P(\Gamma)\chi_{\mathcal{S}_N}(\mathcal{C}_N)}, \tag{I.31}$$

where $P(\Gamma)$ is given by Eq. (I.28). The quantity $\sum_\Gamma P(\Gamma)\chi_{\mathcal{S}_N}(\mathcal{C}_N)$ is equal to the probability $\phi_F^{\mathcal{S}_0\to\mathcal{S}_N}$ to have a forward trajectory starting at $\mathcal{S}_0$ and ending at $\mathcal{S}_N$. As a result:

$$\left\langle\mathcal{F}e^{-\beta W}\right\rangle_F^{\mathcal{S}_0\to\mathcal{S}_N} = \frac{1}{\phi_F^{\mathcal{S}_0\to\mathcal{S}_N}}\sum_\Gamma P(\Gamma)\mathcal{F}(\Gamma)e^{-\beta W(\Gamma)}\chi_{\mathcal{S}_N}(\mathcal{C}_N) \tag{I.32a}$$

$$= \frac{1}{\phi_F^{\mathcal{S}_0\to\mathcal{S}_N}}\sum_\Gamma \frac{\chi_{\mathcal{S}_0}(\mathcal{C}_0)Z_{\lambda_N,\mathcal{S}_N}}{\chi_{\mathcal{S}_N}(\mathcal{C}_N)Z_{\lambda_0,\mathcal{S}_0}}P(\tilde{\Gamma})\mathcal{F}(\Gamma)\chi_{\mathcal{S}_N}(\mathcal{C}_N) \tag{I.32b}$$

$$= \frac{1}{\phi_F^{\mathcal{S}_0\to\mathcal{S}_N}}\frac{Z_{\lambda_N,\mathcal{S}_N}}{Z_{\lambda_0,\mathcal{S}_0}}\sum_\Gamma \chi_{\mathcal{S}_0}(\mathcal{C}_0)P(\tilde{\Gamma})\mathcal{F}(\Gamma) \tag{I.32c}$$

$$= \frac{1}{\phi_F^{\mathcal{S}_0\to\mathcal{S}_N}}\frac{Z_{\lambda_N,\mathcal{S}_N}}{Z_{\lambda_0,\mathcal{S}_0}}\sum_\Gamma \chi_{\tilde{\mathcal{S}}_N}(\tilde{\mathcal{C}}_N)P(\tilde{\Gamma})\mathcal{F}(\Gamma) \tag{I.32d}$$

$$= \frac{1}{\phi_F^{\mathcal{S}_0\to\mathcal{S}_N}}\frac{Z_{\lambda_N,\mathcal{S}_N}}{Z_{\lambda_0,\mathcal{S}_0}}\frac{\sum_\Gamma \chi_{\tilde{\mathcal{S}}_N}(\tilde{\mathcal{C}}_N)P(\tilde{\Gamma})\mathcal{F}(\Gamma)}{\sum_\Gamma \chi_{\tilde{\mathcal{S}}_N}(\tilde{\mathcal{C}}_N)P(\tilde{\Gamma})}\sum_\Gamma \chi_{\tilde{\mathcal{S}}_N}(\tilde{\mathcal{C}}_N)P(\tilde{\Gamma}) \tag{I.32e}$$

$$= \frac{1}{\phi_F^{\mathcal{S}_0\to\mathcal{S}_N}}\frac{Z_{\lambda_N,\mathcal{S}_N}}{Z_{\lambda_0,\mathcal{S}_0}}\left\langle\tilde{\mathcal{F}}\right\rangle_R^{\tilde{\mathcal{S}}_0\to\tilde{\mathcal{S}}_N}\sum_\Gamma \chi_{\tilde{\mathcal{S}}_N}(\tilde{\mathcal{C}}_N)P(\tilde{\Gamma}), \tag{I.32f}$$

where $\langle\ldots\rangle_R^{\tilde{\mathcal{S}}_0\to\tilde{\mathcal{S}}_N}$ denotes the average taken over reversed trajectories that start in partial equilibrium at $\tilde{\mathcal{S}}_0 = \mathcal{S}_N$ and end in $\tilde{\mathcal{S}}_0 = \mathcal{S}_N$. Now, the quantity $\sum_\Gamma \chi_{\tilde{\mathcal{S}}_N}(\tilde{\mathcal{C}}_N)P(\tilde{\Gamma})$ is equal to $\phi_R^{\tilde{\mathcal{S}}_0\to\tilde{\mathcal{S}}_N} = \phi_R^{\mathcal{S}_N\to\mathcal{S}_0} = \phi_R^{\mathcal{S}_0\leftarrow\mathcal{S}_N}$, *i. e.* it is equal to the probability to have a reversed trajectory starting in partial equilibrium at $\mathcal{S}_N$ and ending at $\mathcal{S}_0$. Therefore:

$$\left\langle\mathcal{F}e^{-\beta W}\right\rangle_F^{\mathcal{S}_0\to\mathcal{S}_N} = \frac{\phi_R^{\mathcal{S}_0\leftarrow\mathcal{S}_N}}{\phi_F^{\mathcal{S}_0\to\mathcal{S}_N}}\frac{Z_{\lambda_N,\mathcal{S}_N}}{Z_{\lambda_0,\mathcal{S}_0}}\left\langle\tilde{\mathcal{F}}\right\rangle_R^{\mathcal{S}_0\leftarrow\mathcal{S}_N} \tag{I.33}$$

In order to obtain the extended fluctuation relation, the restricted partition function $Z_{\lambda_i,\mathcal{S}_i}$ is assumed to satisfy Eq. (I.12):

$$F_{\lambda_i,\mathcal{S}_i} = -k_{\mathrm{B}}T \log Z_{\lambda_i,\mathcal{S}_i}, \tag{I.34}$$

and the functional $\mathcal{F}(\Gamma)$ is taken equal to $\delta(W - W(\Gamma))$. Finally:

$$\frac{\phi_F^{\mathcal{S}_0\to\mathcal{S}_N}}{\phi_R^{\mathcal{S}_0\leftarrow\mathcal{S}_N}} \frac{P_F^{\mathcal{S}_0\to\mathcal{S}_N}(W)}{P_R^{\mathcal{S}_0\leftarrow\mathcal{S}_N}(-W)} = e^{\beta(W-\Delta F)}, \tag{I.35}$$

where $\Delta F = F_{\lambda_N,\mathcal{S}_N} - F_{\lambda_0,\mathcal{S}_0}$, and where again $\Delta F = \Delta G$.

Appendix J

Free-energy recovery from equilibrium experiments

In the equilibrium-based hopping experiments the trap-pipette distance λ is kept stationary and the molecule executes transitions between accessible states (Section 2.3.2). These experiments allow to measure free-energy differences between the sampled states. In this appendix it is shown how to extract the free energies of formation for molecules I1 and M1 from these experiments.

J.1 Molecule I1

For the case of I1, there are three different states, which are N, I and U. A typical trace is shown in Fig. J.1a.

The histogram of the measured force along an equilibrium trace can be fitted to a sum of three Gaussians:

$$p(f) = \frac{w_N}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{1}{2}\frac{(f-\langle f_N\rangle)^2}{\sigma_N^2}} + \frac{w_I}{\sqrt{2\pi\sigma_I^2}} e^{-\frac{1}{2}\frac{(f-\langle f_I\rangle)^2}{\sigma_I^2}} + \frac{w_U}{\sqrt{2\pi\sigma_U^2}} e^{-\frac{1}{2}\frac{(f-\langle f_U\rangle)^2}{\sigma_U^2}} \tag{J.1}$$

where w_N, w_I and w_U are the relative weights of each state N, I and U respectively ($w_N + w_I + w_U = 1$); $\langle f_N\rangle$, $\langle f_I\rangle$ and $\langle f_U\rangle$ are their average forces; and σ_N^2, σ_I^2 and σ_U^2 are the respective variances. The free energy of formation of I and U (relative to N) can be obtained using the detailed balance condition, by relating the logarithm of the relative weights of the Gaussians to the free-energy differences between states. To get the free energy at zero force the elastic contribution of each state must be subtracted:

$$\Delta G_{NI}^0 = -k_B T \log\left(\frac{w_I}{w_N}\right) + \int_0^{avf_I} x_I(f)df \tag{J.2}$$

$$\Delta G_{NU}^0 = -k_B T \log\left(\frac{w_U}{w_N}\right) - \int_0^{\langle f_N\rangle} x_d(f)df + \int_0^{\langle f_U\rangle} x_U(f)df \tag{J.3}$$

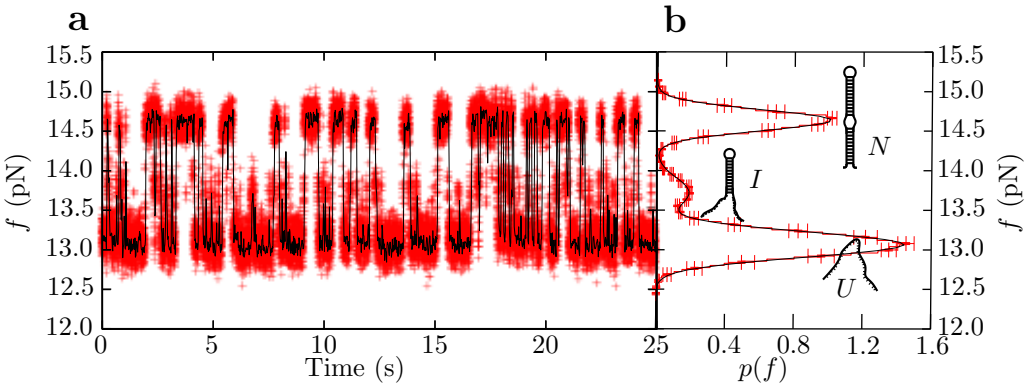

Figure J.1: Hopping experiments with hairpin I1. In these experiments the trap-pipette distance λ is kept stationary. **a.** Sample trace during 25 seconds. In red data is shown at full resolution and in black data is filtered. The molecule executes transitions between N (at ∼14.5 pN), I (at ∼13.8 pN) and U (at ∼13 pN). **b.** Histogram of the measured forces during the hopping experiment. The free energy of formation of each state can be obtained by fitting the histogram to a sum of three Gaussians and using the detailed balance condition, by relating the logarithm of the relative weights of the Gaussians to the free energy differences between the states.

Results are in agreement with free energy predictions using the NN model with the UO set of base-pair free energies and data from UNZ [Sant 98, Zuke 03, Hugu 10b] (Table 9.1).

J.2 Molecule M1

In the case of equilibrium experiments with M1, kinetics are very fast and some transitions are missed. Hopping traces in M1 reveal two clearly separated hopping regions (Fig. J.2): in one region states N and an intermediate I_N coexist; in the other region states M, U and another intermediate I_M coexist. A closer look of the traces at the interphase between both regions shows that there are no trajectories directly connecting N and M (Fig. J.2).

The histogram of the measured forces along an equilibrium trace can be fitted to a sum of five Gaussians:

$$p(f) = \frac{w_N}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{1}{2}\frac{(f-\langle f_N\rangle)^2}{\sigma_N^2}} + \frac{w_{I_N}}{\sqrt{2\pi\sigma_{I_N}^2}} e^{-\frac{1}{2}\frac{(f-\langle f_{I_N}\rangle)^2}{\sigma_{I_N}^2}} + \frac{w_U}{\sqrt{2\pi\sigma_U^2}} e^{-\frac{1}{2}\frac{(f-\langle f_U\rangle)^2}{\sigma_U^2}} + \frac{w_{I_M}}{\sqrt{2\pi\sigma_{I_M}^2}} e^{-\frac{1}{2}\frac{(f-\langle f_{I_M}\rangle)^2}{\sigma_{I_M}^2}} + \frac{w_N}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{1}{2}\frac{(f-\langle f_N\rangle)^2}{\sigma_N^2}} \quad \text{(J.4)}$$

where w_i, $i = N, I_N, U, I_M, M$ are the relative weights of each state ($\sum_i w_i = 1$); $\langle f_i\rangle$ are their average forces; and σ_i^2 are the variances. The free energy of formation of N and

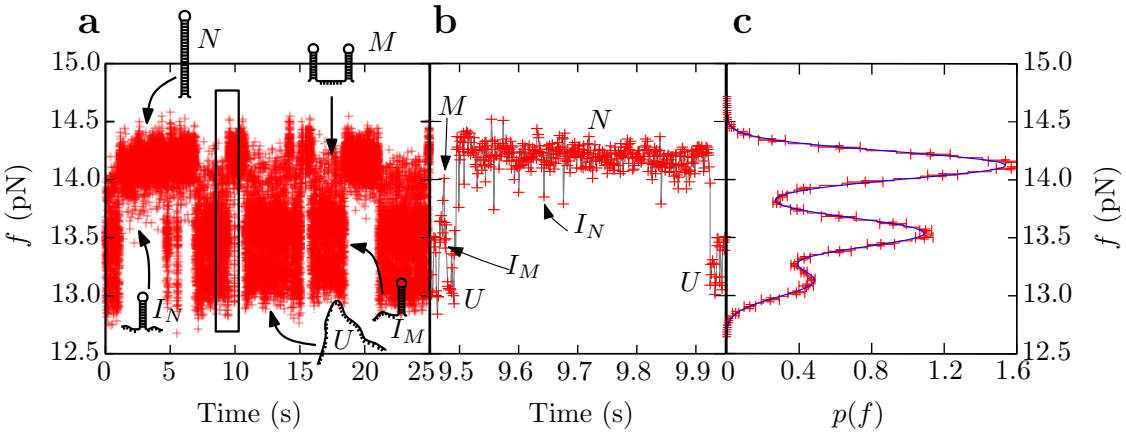

Figure J.2: Hopping experiments with hairpin M1. In these experiments the trap-pipette distance λ is kept stationary. **a.** Sample trace at full resolution. The molecule executes transitions between N (∼14.1 pN), M (∼13.9 pN), two intermediate states I_N and I_M on-pathway to the native and to the misfolded state respectively (∼13.5 pN in both cases since the molecular extension is identical) and U (∼13.0 pN). **b.** Detail of the boxed region in panel a. A closer look of the traces at the interphase between both regions shows that there are not trajectories directly connecting N and M. **c.** Histogram of the measured forces during the hopping experiment. The free energy of formation of each state can be obtained using a fit to a sum of five Gaussians.

M (relative to U) can be obtained using the detailed balance condition, by relating the logarithm of the relative weights of the Gaussians to the free energy differences between states. To get the free energy at zero force the elastic contribution of each state at the average force of the equilibrium trace $\langle f \rangle$ must be subtracted:

$$\Delta G^0_{NU} = -k_B T \log\left(\frac{w_U}{w_N}\right) - \int_0^{\langle f_N \rangle} x_d(f)df + \int_0^{\langle f_U \rangle} x_U(f)df \tag{J.5}$$

$$\Delta G^0_{MU} = -k_B T \log\left(\frac{w_U}{w_M}\right) - 2\int_0^{\langle f_N \rangle} x_d(f)df - \int_0^{\langle f_M \rangle} x_M(f)df + \int_0^{\langle f_U \rangle} x_U(f)df \tag{J.6}$$

Results are in agreement with free energy predictions using the NN model with the UO and the UNZ set of base-pair free energies [Sant 98, Zuke 03, Hugu 10b] (Table 9.4). Intermediate states are not characterized.

Bibliography

[Albe 00] B. Alberts. *Molecular biology of the cell*. Garland Science, 2000.

[Alem 11] A. Alemany, M. Ribezzi-Crivellari, and F. Ritort. "Recent progress in fluctuation theorems and free energy recovery". *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond*, pp. 155–179, 2011.

[Alem 12] A. Alemany, A. Mossa, I. Junier, and F. Ritort. "Experimental free-energy measurements of kinetic molecular states using fluctuation theorems". *Nature Physics*, Vol. 8, pp. 688–694, 2012.

[Alem 13] A. Alemany, N. Sanvicens, S. de Lorenzo, M.-P. Marco, and F. Ritort. "Bond Elasticity Controls Molecular Recognition Specificity in Antibody–Antigen Binding". *Nano letters*, Vol. 13, No. 11, pp. 5197–5202, 2013.

[Alem 14] A. Alemany and F. Ritort. "Determination of the elastic properties of short ssDNA molecules by mechanically folding and unfolding DNA hairpins". *Biopolymers*, Vol. accepted, 2014.

[Andr 94] S. Andrej, E. Shakhnovich, M. Karplus, *et al.* "How does a protein fold?". *Nature*, Vol. 369, No. 6477, pp. 248–251, 1994.

[Anta 91] V. P. Antao, S. Y. Lai, and I. Tinoco Jr. "A thermodynamic study of unusually stable RNA and DNA hairpins". *Nucleic acids research*, Vol. 19, No. 21, pp. 5901–5905, 1991.

[Anta 92] V. P. Antao and I. Tinoco Jr. "Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops". *Nucleic acids research*, Vol. 20, No. 4, pp. 819–824, 1992.

[Arrh 89] S. Arrhenius. "On the reaction rate of the inversion of non-refined sugar upon souring". *Zeitschrift für Physikalische Chemie*, Vol. 4, pp. 266–248, 1889.

[Ashk 70] A. Ashkin. "Acceleration and trapping of particles by radiation pressure". *Physical Review Letters*, Vol. 24, No. 4, pp. 156–159, 1970.

[Ashk 86] A. Ashkin, J. Dziedzic, J. Bjorkholm, and S. Chu. "Observation of a single-beam gradient force optical trap for dielectric particles". *Optics letters*, Vol. 11, No. 5, pp. 288–290, 1986.

[Ashk 97] A. Ashkin. "Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime". *Methods in cell biology*, Vol. 55, pp. 1–27, 1997.

[Aust 74] R. Austin, K. Beeson, L. Eisenstein, H. Frauenfelder, I. Gunsalus, and V. Marshall. "Activation energy spectrum of a biomolecule: Photodissociation of carbonmonoxy myoglobin at low temperatures". *Physical Review Letters*, Vol. 32, No. 8, p. 403, 1974.

[Baum 00] W. Baumgartner, P. Hinterdorfer, W. Ness, A. Raab, D. Vestweber, H. Schindler, and D. Drenckhahn. "Cadherin interaction probed by atomic force microscopy". *Proceedings of the National Academy of Sciences*, Vol. 97, No. 8, pp. 4005–4010, 2000.

[Baum 97] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante. "Ionic effects on the elasticity of single DNA molecules". *Proceedings of the National Academy of Sciences*, Vol. 94, No. 12, pp. 6185–6190, 1997.

[Bell 78] G. I. Bell. "Models for the specific adhesion of cells to cells". *Science*, Vol. 200, No. 4342, pp. 618–627, 1978.

[Benn 76] C. H. Bennett. "Efficient estimation of free energy differences from Monte Carlo data". *Journal of Computational Physics*, Vol. 22, No. 2, pp. 245–268, 1976.

[Berg 99] C. Berger, S. Weber-Bornhauser, J. Eggenberger, J. Hanes, A. Plückthun, and H. R. Bosshard. "Antigen recognition by conformational selection". *FEBS letters*, Vol. 450, No. 1, pp. 149–153, 1999.

[Best 08] 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.

[Biza 12] C. V. Bizarro, A. Alemany, and F. Ritort. "Non-specific binding of Na+ and Mg2+ to RNA determined by force spectroscopy methods". *Nucleic acids research*, Vol. 40, No. 14, pp. 6922–6935, 2012.

[Blom 89] M. Blommers, J. Walters, C. Haasnoot, J. Aelen, G. Van der Marel, J. Van Boom, and C. Hilbers. "Effects of base sequence on the loop folding in DNA hairpins". *Biochemistry*, Vol. 28, No. 18, pp. 7491–7498, 1989.

[Boeh 09] D. D. Boehr, R. Nussinov, and P. E. Wright. "The role of dynamic conformational ensembles in biomolecular recognition". *Nature chemical biology*, Vol. 5, No. 11, pp. 789–796, 2009.

[Borg 08] A. Borgia, P. M. Williams, and J. Clarke. "Single-molecule studies of protein folding". *Annual Review of Biochemistry*, Vol. 77, pp. 101–125, 2008.

[Bosc 13] 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*, p. gkt1089, 2013.

[Bouc 99] C. Bouchiat, M. D. Wang, J. F. Allemand, T. Strick, S. M. Block, and V. Croquette. "Estimating the persistence length of a worm-like chain molecule from force-extension measurements". *Biophysical journal*, Vol. 76, No. 1, pp. 409–413, 1999.

[Broo 98] C. L. Brooks III. "Simulations of protein folding and unfolding". *Current opinion in structural biology*, Vol. 8, No. 2, pp. 222–226, 1998.

[Bryn 87] J. D. Bryngelson and P. G. Wolynes. "Spin glasses and the statistical mechanics of protein folding". *Proceedings of the National Academy of Sciences*, Vol. 84, No. 21, pp. 7524–7528, 1987.

[Bust 04] C. Bustamante, Y. R. Chemla, N. R. Forde, and D. Izhaky. "Mechanical processes in biochemistry". *Annual Review of Biochemistry*, Vol. 73, pp. 705–748, 2004.

[Bust 05] C. Bustamante, J. Liphardt, and F. Ritort. "The nonequilibrium thermodynamics of small systems". *Physics Today*, Vol. 58, pp. 43–48, 2005.

[Bust 06] C. Bustamante and S. B. Smith. "Light-force sensor and method for measuring axial optical-trap forces from changes in light momentum along an optic axis". *US Patent*, Vol. 7, pp. 133,132,B2, 2006.

[Bust 94] C. Bustamante, J. Marko, E. D. Siggia, and S. B. Smith. "Entropic elasticity of λ-phage DNA". *Science*, Vol. 265, No. 5178, pp. 1599–1600, 1994.

[Camu 14] J. Camunas-Soler, A. Alemany, and F. Ritort. "Recovery of free energies of binding from fluctuation relations". *In preparation*, 2014.

[Carr 00] M. Carrión-Vázquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. Li, and J. M. Fernandez. "Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering". *Progress in biophysics and molecular biology*, Vol. 74, No. 1, pp. 63–91, 2000.

[Cecc 05] 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.

[Cecc 11] 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.

[Chen 12] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack. "Ionic strength-dependent persistence lengths of single-stranded RNA and DNA". *Proceedings of the National Academy of Sciences*, Vol. 109, No. 3, pp. 799–804, 2012.

[Chun 09] H. S. Chung, J. M. Louis, and W. A. Eaton. "Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories". *Proceedings of the National Academy of Sciences*, Vol. 106, No. 29, pp. 11837–11844, 2009.

[Chun 12] H. S. Chung, K. McHale, J. M. Louis, and W. A. Eaton. "Single-molecule fluorescence experiments determine protein folding transition path times". *Science*, Vol. 335, No. 6071, pp. 981–984, 2012.

[Cocc 03] S. Cocco, J. Marko, and R. Monasson. "Slow nucleic acid unzipping kinetics from sequence-defined barriers". *The European Physical Journal E*, Vol. 10, No. 2, pp. 153–161, 2003.

[Cole 72] P. Cole, S. Yang, and D. Crothers. "Conformational changes of transfer ribonucleic acid. Equilibrium phase diagrams". *Biochemistry*, Vol. 11, No. 23, pp. 4358–4368, 1972.

[Coll 05] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr, and C. Bustamante. "Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies". *Nature*, Vol. 437, pp. 231–234, 2005.

[Colu 03] I. Coluzza, H. Muller, and D. Frenkel. "Designing refoldable model molecules". *Physical Review E*, Vol. 68, No. 4, p. 046703, 2003.

[Croo 00] G. E. Crooks. "Path-ensemble averages in systems driven far from equilibrium". *Physical Review E*, Vol. 61, No. 3, p. 2361, 2000.

[Croo 99] G. E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". *Physical Review E*, Vol. 60, No. 3, p. 2721, 1999.

[Cui 00] Y. Cui and C. Bustamante. "Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure". *Proceedings of the National Academy of Sciences*, Vol. 97, No. 1, pp. 127–132, 2000.

[Dell 13] C. Dellago and G. Hummer. "Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics". *Entropy*, Vol. 16, No. 1, pp. 41–61, 2013.

[Derr 81] B. Derrida. "Random-energy model: An exactly solvable model of disordered systems". *Physical Review B*, Vol. 24, No. 5, p. 2613, 1981.

[Dess 02] M.-N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette. "Stretching single stranded DNA, a model polyelectrolyte". *Physical Review Letters*, Vol. 89, No. 24, p. 248102, 2002.

[Diet 04] H. Dietz and M. Rief. "Exploring the energy landscape of GFP by single-molecule mechanical experiments". *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, pp. 16192–16197, 2004.

[Dill 97] K. A. Dill and H. S. Chan. "From Levinthal to pathways to funnels". *Nature structural biology*, Vol. 4, No. 1, pp. 10–19, 1997.

[Dinn 00] A. R. Dinner, A. Šali, L. J. Smith, C. M. Dobson, and M. Karplus. "Understanding protein folding via free-energy surfaces from theory and experiment". *Trends in biochemical sciences*, Vol. 25, No. 7, pp. 331–339, 2000.

[Doua 06] F. Douarche, S. Joubaud, N. B. Garnier, A. Petrosyan, and S. Ciliberto. "Work fluctuation theorems for harmonic oscillators". *Physical Review Letters*, Vol. 97, No. 14, p. 140603, 2006.

[Dove 62] W. F. Dove and N. Davidson. "Cation effects on the denaturation of DNA". *Journal of Molecular Biology*, Vol. 5, No. 5, pp. 467–478, 1962.

[Dudk 06] O. K. Dudko, G. Hummer, and A. Szabo. "Intrinsic rates and activation free energies from single-molecule pulling experiments". *Physical Review Letters*, Vol. 96, No. 10, p. 108101, 2006.

[Dudk 08] O. K. Dudko, G. Hummer, and A. Szabo. "Theory, analysis, and interpretation of single-molecule force spectroscopy experiments". *Proceedings of the National Academy of Sciences*, Vol. 105, No. 41, pp. 15755–15760, 2008.

[Enge 11] S. Engel, A. Alemany, N. Forns, P. Maass, and F. Ritort. "Folding and unfolding of a triple-branch DNA molecule with four conformational states". *Philosophical Magazine*, Vol. 91, No. 13-15, pp. 2049–2065, 2011.

[Erba 99] E. Erba, D. Bergamaschi, S. Ronzoni, M. Faretta, S. Taverna, M. Bonfanti, C. Catapano, G. Faircloth, J. Jimeno, and M. D'incalci. "Mode of action of thiocoraline, a natural marine compound with anti-tumour activity". *British journal of cancer*, Vol. 80, No. 7, p. 971, 1999.

[Este 05] M.-C. Estévez, R. Galve, F. Sánchez-Baeza, and M.-P. Marco. "Direct competitive enzyme-linked immunosorbent assay for the determination of the highly polar short-chain sulfophenyl carboxylates". *Analytical chemistry*, Vol. 77, No. 16, pp. 5283–5293, 2005.

[Evan 01] 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.

[Evan 09] E. Evans, K. Halvorsen, K. Kinoshita, and W. P. Wong. "A New Approach to Analysis of Single-Molecule Force Measurements". In: *Handbook of Single-Molecule Biophysics*, pp. 571–589, Springer, 2009.

[Evan 10] E. Evans, K. Kinoshita, S. Simon, and A. Leung. "Long-Lived, High-Strength States of ICAM-1 Bonds to β_2 Integrin, I: Lifetimes of Bonds to Recombinant α_Lβ_2 Under Force". *Biophysical journal*, Vol. 98, No. 8, pp. 1458–1466, 2010.

[Evan 35] M. G. Evans and M. Polanyi. "Some applications of the transition state method to the calculation of reaction velocities, especially in solution". *Transactions of the Faraday Society*, Vol. 31, pp. 875–894, 1935.

[Evan 97] E. Evans and K. Ritchie. "Dynamic strength of molecular adhesion bonds". *Biophysical journal*, Vol. 72, No. 4, pp. 1541–1555, 1997.

[Eyri 35] H. Eyring. "The activated complex in chemical reactions". *The Journal of Chemical Physics*, Vol. 3, p. 107, 1935.

[Ferr 11] A. C. M. Ferreon and A. A. Deniz. "Protein folding at single-molecule resolution". *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, Vol. 1814, No. 8, pp. 1021–1029, 2011.

[Fers 00] A. R. Fersht. "A kinetically significant intermediate in the folding of barnase". *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 97, No. 26, pp. 14121–14126, 2000.

[Fers 93] A. R. Fersht. "Protein folding and stability: the pathway of folding of barnase". *FEBS letters*, Vol. 325, No. 1, pp. 5–16, 1993.

[Fiel 96] B. A. Fields, F. A. Goldbaum, W. Dall'Acqua, E. L. Malchiodi, A. Cauerhff, F. P. Schwarz, X. Ysern, R. J. Poljak, and R. A. Mariuzza. "Hydrogen Bonding and Solvent Structure in an Antigen-Antibody Interface. Crystal Structures and Thermodynamic Characterization of Three Fv Mutants Complexed with Lysozyme,#". *Biochemistry*, Vol. 35, No. 48, pp. 15494–15503, 1996.

[Forn 11] N. Forns, S. de Lorenzo, M. Manosas, K. Hayashi, J. M. Huguet, and F. Ritort. "Improving signal/noise resolution in single-molecule experiments using molecular constructs with short handles". *Biophysical journal*, Vol. 100, No. 7, pp. 1765–1774, 2011.

[Fran 53] R. E. Franklin and R. G. Gosling. "Molecular configuration in sodium thymonucleate". *Nature*, Vol. 171, pp. 740–741, 1953.

[Gale 10] A. Galera-Prat, A. Gómez-Sicilia, A. F. Oberhauser, M. Cieplak, and M. Carrión-Vázquez. "Understanding biology by stretching proteins: recent progress". *Current opinion in structural biology*, Vol. 20, No. 1, pp. 63–69, 2010.

[Gara 14] A. Garai, Y. Zhang, and O. K. Dudko. "Conformational dynamics through an intermediate". *The Journal of chemical physics*, Vol. 140, No. 13, p. 135101, 2014.

[Gebh 10] 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.

[Goda 96] M. Godall. *Biologia humana: fonaments biològics per a diplomatures en Ciències de la Salut*. Universitat Ramon Llull, 1996.

[Goss 02] C. Gossel and V. Croquette. "Magnetic Tweezers: Micromanipulation and Force Measurement at the Molecular Level". *Biophysical Journal*, Vol. 82 (6), pp. 3314–3329, 2002.

[Grey 64] H. Grey. "Studies on changes in the quality of rabbit-bovine serum albumin antibody following immunization". *Immunology*, Vol. 7, No. 1, p. 82, 1964.

[Gupt 11] A. N. Gupta, A. Vincent, K. Neupane, H. Yu, F. Wang, and M. T. Woodside. "Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements". *Nature Physics*, Vol. 7, No. 8, pp. 631–634, 2011.

[Hamm 55] G. S. Hammond. "A correlation of reaction rates". *Journal of the American Chemical Society*, Vol. 77, No. 2, pp. 334–338, 1955.

[Hang 90] P. Hänggi, P. Talkner, and M. Borkovec. "Reaction-rate theory: fifty years after Kramers". *Reviews of Modern Physics*, Vol. 62, No. 2, p. 251, 1990.

[Haya 10] K. Hayashi, H. Ueno, R. Iino, and H. Noji. "Fluctuation Theorem Applied to F1-ATPase". *Physical Review Letters*, Vol. 104, p. 218103, 2010.

[Heid 12] P. O. Heidarsson, I. Valpapuram, C. Camilloni, A. Imparato, G. Tiana, F. M. Poulsen, B. B. Kragelund, and C. Cecconi. "A highly compliant protein native state with a spontaneous-like mechanical unfolding pathway". *Journal of the American Chemical Society*, Vol. 134, No. 41, pp. 17068–17075, 2012.

[Heid 14] P. O. Heidarsson, M. M. Naqvi, M. R. Otazo, A. Mossa, B. B. Kragelund, and C. Cecconi. "Direct single-molecule observation of calcium-dependent misfolding in human neuronal calcium sensor-1". *Proceedings of the National Academy of Sciences*, p. 201401065, 2014.

[Heil 01] S. L. Heilman-Miller, D. Thirumalai, and S. A. Woodson. "Role of counterion condensation in folding of the Tetrahymena ribozyme. I. Equilibrium stabilization by cations". *Journal of molecular biology*, Vol. 306, No. 5, pp. 1157–1166, 2001.

[Helm 97] K. Helmerson, R. Kishore, W. D. Phillips, and H. H. Weetall. "Optical tweezers-based immunosensor detects femtomolar concentrations of antigens". *Clinical chemistry*, Vol. 43, No. 2, pp. 379–383, 1997.

[HGP 14] HGP. "All About The Human Genome Project (HGP)". *http://www.genome.gov/10001772*, 2014.

[Hilb 85] C. Hilbers, C. Haasnoot, S. De Bruin, J. Joordens, G. Van der Marel, and J. Van Boom. "Hairpin formation in synthetic oligonucleotides". *Biochimie*, Vol. 67, No. 7, pp. 685–695, 1985.

[Hinc 10] M. Hinczewski, Y. von Hansen, J. Dzubiella, and R. R. Netz. "How the diffusivity profile reduces the arbitrariness of protein folding free energies". *The Journal of chemical physics*, Vol. 132, No. 24, p. 245103, 2010.

[Hint 06] P. Hinterdorfer and Y. F. Dufrêne. "Detection and localization of single molecular recognition events using atomic force microscopy". *Nature methods*, Vol. 3, No. 5, pp. 347–355, 2006.

[Homo 03] J. Homola. "Present and future of surface plasmon resonance biosensors". *Analytical and bioanalytical chemistry*, Vol. 377, No. 3, pp. 528–539, 2003.

[Huge 01] T. Hugel and M. Seitz. "The study of molecular interactions by AFM force spectroscopy". *Macromolecular rapid communications*, Vol. 22, No. 13, pp. 989–1016, 2001.

[Hugu 10a] J. M. Huguet. *Statistical and thermodynamic properties of DNA unzipping experiments with optical tweezers*. PhD thesis, Universitat de Barcelona, 2010.

[Hugu 10b] 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.

[Humm 01a] G. Hummer. "Fast-growth thermodynamic integration: error and efficiency analysis". *The Journal of Chemical Physics*, Vol. 114, No. 17, pp. 7330–7337, 2001.

[Humm 01b] G. Hummer and A. Szabo. "Free energy reconstruction from nonequilibrium single-molecule pulling experiments". *Proceedings of the National Academy of Sciences*, Vol. 98, No. 7, pp. 3658–3661, 2001.

[Hyeo 07] C. Hyeon and D. Thirumalai. "Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments". *Journal of Physics: Condensed Matter*, Vol. 19, No. 11, p. 113101, 2007.

[Jack 06] M. B. Jackson. *Molecular and cellular biophysics*. Cambridge University Press, 2006.

[Jane 02] C. A. Janeway Jr and R. Medzhitov. "Innate immune recognition". *Science Signaling*, Vol. 20, No. 1, p. 197, 2002.

[Jarz 11] C. Jarzynski. "Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale". *Annual Review of Condensed Matter Physics*, Vol. 2, pp. 329–51, 2011.

[Jarz 97] C. Jarzynski. "Nonequilibrium equality for free energy differences". *Physical Review Letters*, Vol. 78, No. 14, p. 2690, 1997.

[Jern 55] N. K. Jerne. "The natural-selection theory of antibody formation". *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 41, No. 11, p. 849, 1955.

[Jime 03] R. Jimenez, G. Salazar, K. K. Baldridge, and F. E. Romesberg. "Flexibility and molecular recognition in the immune system". *Proceedings of the National Academy of Sciences*, Vol. 100, No. 1, pp. 92–97, 2003.

[Jime 04] R. Jimenez, G. Salazar, J. Yin, T. Joo, and F. E. Romesberg. "Protein dynamics and the immunological evolution of molecular recognition". *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, No. 11, pp. 3803–3808, 2004.

[Joub 07] S. Joubaud, N. Garnier, and S. Ciliberto. "Fluctuation theorems for harmonic oscillators". *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2007, No. 09, p. P09018, 2007.

[Juni 08] I. Junier and F. Ritort. "Unstructured intermediate states in single protein force experiments". *Proteins: structure, function and bioinformatics*, Vol. 71, pp. 1145–1155, 2008.

[Juni 09] I. Junier, A. Mossa, M. Manosas, and F. Ritort. "Recovery of free energy branches in single molecule experiments". *Physical Review Letters*, Vol. 102, No. 7, p. 070602, 2009.

[Karp 97] M. Karplus. "The Levinthal paradox: yesterday and today". *Folding and Design*, Vol. 2, pp. S69–S75, 1997.

[Kell 97] M. S. 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.

[Kerk 69] M. Kerker. *The scattering of light, and other electromagnetic radiation.* Academic press, 1969.

[Kino 00] 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.

[Koir 11] D. Koirala, S. Dhakal, B. Ashbridge, Y. Sannohe, R. Rodriguez, H. Sugiyama, S. Balasubramanian, and H. Mao. "A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands". *Nature chemistry*, Vol. 3, No. 10, pp. 782–787, 2011.

[Kosh 58] D. Koshland Jr. "Application of a theory of enzyme specificity to protein synthesis". *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 44, No. 2, p. 98, 1958.

[Kram 07] K. Kramer, A. Hubauer, R. Lausterer, J.-P. Salvador, and M.-P. Marco. "Production of Antibodies for the Quantitative Detection of the Anabolically Active Androgens 17β-Boldenone and Methylboldenone". *Analytical letters*, Vol. 40, No. 7, pp. 1461–1472, 2007.

[Kram 40] 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.

[Kuo 93] S. C. Kuo and D. A. Lauffenburger. "Relationship between receptor/ligand binding affinity and adhesion strength". *Biophysical journal*, Vol. 65, No. 5, pp. 2191–2200, 1993.

[Leff 53] J. E. Leffler. "Parameters for the Description of Transition States.". *Science (New York, NY)*, Vol. 117, No. 3039, p. 340, 1953.

[Lei 10] J. Lei and K. Huang. "Protein folding: A perspective from statistical physics". *arXiv preprint arXiv:1002.5013*, 2010.

[Levi 69] C. Levinthal. "How to fold graciously". *Mossbauer spectroscopy in biological systems*, pp. 22–24, 1969.

[Li 07] P. T. Li, C. Bustamante, and I. Tinoco Jr. "Real-time control of the energy landscape by force directs the folding of RNA molecules". *Proceedings of the National Academy of Sciences*, Vol. 104, No. 17, pp. 7039–7044, 2007.

[Lin 12] J.-C. Lin, C. Hyeon, and D. Thirumalai. "RNA under Tension: Folding Landscapes, Kinetic Partitioning Mechanism, and Molecular Tensegrity". *The journal of physical chemistry letters*, Vol. 3, No. 23, pp. 3616–3625, 2012.

[Liph 01] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco Jr, and C. Bustamante. "Reversible unfolding of single RNA molecules by mechanical force". *Science*, Vol. 292, No. 5517, pp. 733–737, 2001.

[Liph 12] J. Liphardt. "Thermodynamic limits". *Nature Physics*, Vol. 8, pp. 638–639, 2012.

[Liu 10] W. Liu, T. Meckel, P. Tolar, H. W. Sohn, and S. K. Pierce. "Antigen affinity discrimination is an intrinsic function of the B cell receptor". *The Journal of experimental medicine*, Vol. 207, No. 5, pp. 1095–1111, 2010.

[Ma 02] B. Ma, M. Shatsky, H. J. Wolfson, and R. Nussinov. "Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations". *Protein Science*, Vol. 11, No. 2, pp. 184–197, 2002.

[Mame 09] J. van Mameren, P. Gross, G. Farge, P. Hooijman, M. Modesti, M. Falkenberg, G. J. Wuite, and E. J. Peterman. "Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging". *Proceedings of the National Academy of Sciences*, Vol. 106, No. 43, pp. 18231–18236, 2009.

[Mani 00] V. Manivel, N. C. Sahoo, D. M. Salunke, and K. V. Rao. "Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site". *Immunity*, Vol. 13, No. 5, pp. 611–620, 2000.

[Mann 02] G. S. Manning. "Electrostatic free energy of the DNA double helix in counterion condensation theory". *Biophysical chemistry*, Vol. 101, pp. 461–473, 2002.

[Mann 72] G. S. Manning. "On the application of polyelectrolyte "limiting laws" to the helix-coil transition of DNA. I. Excess univalent cations". *Biopolymers*, Vol. 11, No. 5, pp. 937–949, 1972.

[Mann 78] G. S. Manning. "The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides". *Quarterly reviews of biophysics*, Vol. 11, No. 02, pp. 179–246, 1978.

[Mann 96] G. S. Manning. "Counterion condensation theory constructed from different models". *Physica A: Statistical Mechanics and its Applications*, Vol. 231, No. 1, pp. 236–253, 1996.

[Mano 05] M. Manosas and F. Ritort. "Thermodynamic and kinetic aspects of RNA pulling experiments". *Biophysical journal*, Vol. 88, No. 5, pp. 3224–3242, 2005.

[Mano 06] M. Manosas, D. Collin, and F. Ritort. "Force-dependent fragility in RNA hairpins". *Physical Review Letters*, Vol. 96, No. 21, p. 218301, 2006.

[Mano 07a] M. Manosas. *Mechanical folding/unfolding of RNA molecules: Experimental facts and theoretical models.* PhD thesis, Universitat de Barcelona, 2007.

[Mano 07b] M. Manosas, J.-D. Wen, P. T. Li, S. B. Smith, C. Bustamante, I. Tinoco Jr, and F. Ritort. "Force unfolding kinetics of RNA using optical tweezers. II. Modeling experiments". *Biophysical journal*, Vol. 92, No. 9, pp. 3010–3021, 2007.

[Mara 08] P. Maragakis, M. Spichty, and M. Karplus. "A differential fluctuation theorem". *The Journal of Physical Chemistry B*, Vol. 112, No. 19, pp. 6168–6174, 2008.

[Mart 00] J. Martí, F. S. Csajka, and D. Chandler. "Stochastic transition pathways in the aqueous sodium chloride dissociation process". *Chemical Physics Letters*, Vol. 328, No. 1, pp. 169–176, 2000.

[Math 04] D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner. "Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure". *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, No. 19, pp. 7287–7292, 2004.

[Math 99] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner. "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure". *Journal of molecular biology*, Vol. 288, No. 5, pp. 911–940, 1999.

[Mato 90] A. Matouschek, J. T. Kellis, L. Serrano, M. Bycroft, and A. R. Fersht. "Transient folding intermediates characterized by protein engineering". *Nature*, Vol. 346, No. 6283, pp. 440–445, 1990.

[Mato 92] A. Matouschek, L. Serrano, and A. R. Fersht. "The folding of an enzyme: IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure". *Journal of molecular biology*, Vol. 224, No. 3, pp. 819–835, 1992.

[McMa 02] M. T. McManus, C. P. Petersen, B. B. Haines, J. Chen, and P. A. Sharp. "Gene silencing using micro-RNA designed hairpins.". *RNA*, Vol. 8, No. 6, pp. 842–850, 2002.

[Merk 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, pp. 50–53, 1999.

[Mick 07] M. Mickler, R. I. Dima, H. Dietz, C. Hyeon, D. Thirumalai, and M. Rief. "Revealing the bifurcation inthe unfolding pathways of GFP by using

single-molecule experiments and simulations". *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 104, pp. 20268–20273, 2007.

[Mick 96] U. Micka and K. Kremer. "Persistence length of the Debye-Hückel model of weakly charged flexible polyelectrolyte chains". *Physical Review E*, Vol. 54, No. 3, p. 2653, 1996.

[Moff 08] J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante. "Recent advances in optical tweezers". *Annual Review of Biochemistry*, Vol. 77, pp. 205–228, 2008.

[Morr 11] G. Morrison, C. Hyeon, M. Hinczewski, and D. Thirumalai. "Compaction and tensile forces determine the accuracy of folding landscape parameters from single molecule pulling experiments". *Physical Review Letters*, Vol. 106, No. 13, p. 138102, 2011.

[Mosc 09] F. Mosconi, J. F. Allemand, D. Bensimon, and V. Croquette. "Measurement of the Torque on a Single Stretched and Twisted DNA Using Magnetic Tweezers". *Physical Review Letters*, Vol. 201, p. 078301, 2009.

[Moss 09a] A. Mossa, S. de Lorenzo, J. M. Huguet, and F. Ritort. "Measurement of work in single-molecule pulling experiments". *The Journal of chemical physics*, Vol. 130, No. 23, pp. 234116–234116, 2009.

[Moss 09b] 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, No. 02, p. P02060, 2009.

[Moss 09c] 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, No. 02, p. P02060, 2009.

[Moss 10] A. Mossa, J. M. Huguet, and F. Ritort. "Investigating the thermodynamics of small biosystems with optical tweezers". *Physica E: Low-dimensional Systems and Nanostructures*, Vol. 42, No. 3, pp. 666–671, 2010.

[Murp 04] M. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha. "Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy". *Biophysical Journal*, Vol. 86, No. 4, pp. 2530–2537, 2004.

[Natk 13] E. Natkanski, W.-Y. Lee, B. Mistry, A. Casal, J. E. Molloy, and P. Tolar. "B cells use mechanical energy to discriminate antigen affinities". *Science*, Vol. 340, No. 6140, pp. 1587–1590, 2013.

[Negr 07] A. Negri, E. Marco, V. García-Hernández, A. Domingo, A. L. Llamas-Saiz, S. Porto-Sandá, R. Riguera, W. Laine, M.-H. David-Cordonnier, C. Bailly, *et al.* "Antitumor activity, X-ray crystal structure, and DNA binding properties of thiocoraline A, a natural bisintercalating thiodepsipeptide". *Journal of medicinal chemistry*, Vol. 50, No. 14, pp. 3322–3333, 2007.

[Neum 04] K. C. Neuman and S. M. Block. "Optical trapping". *Review of scientific instruments*, Vol. 75, No. 9, pp. 2787–2809, 2004.

[Neum 08] 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.

[Neup 12] K. Neupane, D. B. Ritchie, H. Yu, D. A. Foster, F. Wang, and M. T. Woodside. "Transition path times for nucleic acid folding determined from energy-landscape analysis of single-molecule trajectories". *Physical Review Letters*, Vol. 109, No. 6, p. 068102, 2012.

[Olso 75] W. K. Olson. "Configurational statistics of polynucleotide chains. A single virtual bond treatment". *Macromolecules*, Vol. 8, No. 3, pp. 272–275, 1975.

[Owcz 08] R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, and J. A. Walder. "Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations". *Biochemistry*, Vol. 47, No. 19, pp. 5336–5353, 2008.

[Pala 11] 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.

[Palm 82] R. Palmer. "Broken ergodicity". *Advances in Physics*, Vol. 31, No. 6, pp. 669–735, 1982.

[Path 96] R. Pathria. "Statistical mechanics". 1996.

[Paul 46] L. Pauling. "Molecular architecture and biological reactions". *Chemical and engineering news*, Vol. 24, No. 10, pp. 1375–1377, 1946.

[Peck 11] H. E. Peckham and W. K. Olson. "Nucleic-acid structural deformability deduced from anisotropic displacement parameters". *Biopolymers*, Vol. 95, No. 4, pp. 254–269, 2011.

[Perr 02] E. Perret, A.-M. Benoliel, P. Nassoy, A. Pierres, V. Delmas, J.-P. Thiery, P. Bongrand, and H. Feracci. "Fast dissociation kinetics between individual E-cadherin fragments revealed by flow chamber analysis". *The EMBO journal*, Vol. 21, No. 11, pp. 2537–2546, 2002.

[Peyr 00] N. Peyret. *Prediction of Nucleic Acid Hybridization: Parameters and Algorithms*. PhD thesis, Wayne State University Department of Chemistry, Detroit, MI, 2000.

[Pier 13] C. A. Pierse and O. K. Dudko. "Kinetics and Energetics of Biomolecular Folding and Binding". *Biophysical journal*, Vol. 105, No. 9, pp. L19–L22, 2013.

[Poco 02] G. Pocock and C. D. Richards. *Human physiology the basis of medicine*. Oxford: Oxford University Press, 2002.

[Priv 69] P. Privalov, O. Ptitsyn, and T. Birshtein. "Determination of stability of the DNA double helix in an aqueous medium". *Biopolymers*, Vol. 8, No. 5, pp. 559–571, 1969.

[Raib 06] M. Raible, M. Evstigneev, F. W. Bartels, R. Eckel, M. Nguyen-Duong, R. Merkel, R. Ros, D. Anselmetti, and P. Reimann. "Theoretical analysis of single-molecule force spectroscopy experiments: heterogeneity of chemical bonds". *Biophysical journal*, Vol. 90, No. 11, pp. 3851–3864, 2006.

[Reco 75] M. T. Record. "Effects of Na+ and Mg++ ions on the helix–coil transition of DNA". *Biopolymers*, Vol. 14, No. 10, pp. 2137–2158, 1975.

[Reco 76] M. T. Record, C. P. Woodbury, and T. M. Lohman. "Na+ effects on transitions of DNA and polynucleotides of variable linear charge density". *Biopolymers*, Vol. 15, No. 5, pp. 893–915, 1976.

[Reco 78] M. T. Record, C. F. Anderson, and T. M. Lohman. "Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity". *Quarterly reviews of biophysics*, Vol. 11, No. 02, pp. 103–178, 1978.

[Ribe 12] M. Ribezzi-Crivellari and F. Ritort. "Force Spectroscopy with Dual-Trap Optical Tweezers: Molecular Stiffness Measurements and Coupled Fluctuations Analysis". *Biophysical journal*, Vol. 103, No. 9, pp. 1919–1928, 2012.

[Ribe 13] M. Ribezzi-Crivellari, J. M. Huguet, and F. Ritort. "Counter-propagating dual-trap optical tweezers based on linear momentum conservation". *Review of Scientific Instruments*, Vol. 84, No. 4, p. 043104, 2013.

[Ribe 14] M. Ribezzi-Crivellari and F. Ritort. "Free-energy inference from partial work measurements in small systems". *Proceedings of the National Academy of Sciences*, Vol. 111, No. 33, pp. E3386–E3394, 2014.

[Rich 03] A. Rich. "The double helix: a tale of two puckers". *Nature Structural & Molecular Biology*, Vol. 10, No. 4, pp. 247–249, 2003.

[Rief 99] M. Rief, J. Pascual, M. Saraste, and H. E. Gaub. "Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles". *Journal of molecular biology*, Vol. 286, No. 2, pp. 553–561, 1999.

[Rito 06a] F. Ritort. "Nonequilibrium fluctuations in small systems: from physics to biology". *Advanced Chemical Physics*, Vol. 137, pp. 31–123, 2006.

[Rito 06b] F. Ritort. "Single-molecule experiments in biological physics: methods and applications". *Journal of Physics: Condensed Matter*, Vol. 18, No. 32, p. R531, 2006.

[Rose 09] O. Rosen and J. Anglister. "Epitope Mapping of Antibody–Antigen Complexes by Nuclear Magnetic Resonance Spectroscopy". In: *Epitope Mapping Protocols*, pp. 37–57, Springer, 2009.

[Saen 84] W. Saenger. *Principles of nucleic acid structure*. Springer-Verlag, 1984.

[Salo 08] M. Salomo, U. F. Keyser, M. Struhalla, and F. Kremer. "Optical tweezers to study single protein A/immunoglobulin G interactions at varying conditions". *European Biophysics Journal*, Vol. 37, No. 6, pp. 927–934, 2008.

[Sant 04] J. SantaLucia Jr and D. Hicks. "The thermodynamics of DNA structural motifs". *Annual Review of Biophysics and Biomolecular Structure*, Vol. 33, pp. 415–440, 2004.

[Sant 98] 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.

[Schl 04] M. Schlierf, H. Li, and J. M. Fernandez. "The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques". *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, No. 19, pp. 7299–7304, 2004.

[Schl 06] M. Schlierf and M. Rief. "Single-molecule unfolding force distributions reveal a funnel-shaped energy landscape". *Biophysical journal*, Vol. 90, No. 4, pp. L33–L35, 2006.

[Schr 00] S. J. Schroeder and D. H. Turner. "Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA". *Biochemistry*, Vol. 39, No. 31, pp. 9257–9274, 2000.

[Schu 05] S. Schuler, T. Speck, C. Tietz, J. Wrachtrup, and U. Seifert. "Experimental test of the fluctuation theorem for a driven two-level system with time-dependent rates". *Physical Review Letters*, Vol. 94, No. 18, p. 180602, 2005.

[Seni 88] M. M. Senior, R. A. Jones, and K. J. Breslauer. "Influence of loop residues on the relative stabilities of DNA hairpin structures". *Proceedings of the National Academy of Sciences*, Vol. 85, No. 17, pp. 6242–6246, 1988.

[Seol 04] Y. Seol, G. M. Skinner, and K. Visscher. "Elastic properties of a single-stranded charged homopolymeric ribonucleotide". *Physical Review Letters*, Vol. 93, No. 11, p. 118102, 2004.

[Serr 92] L. Serrano, A. Matouschek, and A. R. Fersht. "The folding of an enzyme: III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure". *Journal of molecular biology*, Vol. 224, No. 3, pp. 805–818, 1992.

[Serr 94] M. J. Serra, T. J. Axenson, and D. H. Turner. "A model for the stabilities of RNA hairpins based on a study of the sequence dependence of stability for hairpins of six nucleotides". *Biochemistry*, Vol. 33, No. 47, pp. 14289–14296, 1994.

[Serr 97] M. J. Serra, T. W. Barnes, K. Betschart, M. J. Gutierrez, K. J. Sprouse, C. K. Riley, L. Stewart, and R. E. Temel. "Improved parameters for the prediction of RNA hairpin stability". *Biochemistry*, Vol. 36, No. 16, pp. 4844–4851, 1997.

[Shan 10] E. A. Shank, C. Cecconi, W. D. Jesse, S. Marqusee, and C. Bustamante. "The folding cooperativity of a protein is controlled by its chain topology". *Nature*, Vol. 465, pp. 637–641, 2010.

[Shir 03] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande. "Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods". *Physical Review Letters*, Vol. 91, No. 14, p. 140601, 2003.

[Shke 04] I. A. Shkel and M. T. Record. "Effect of the number of nucleic acid oligomer charges on the salt dependence of stability (Δ) and melting temperature (T m): NLPB analysis of experimental data". *Biochemistry*, Vol. 43, No. 22, pp. 7090–7101, 2004.

[Simm 93] R. Simmonst, J. T. Finer, H. Warrick, B. Kralik, S. Chu, and J. Spudich. "Force on single actin filaments in a motility assay measured with an optical trap". In: *Mechanism of Myofilament Sliding in Muscle Contraction*, pp. 331–337, Springer, 1993.

[Simm 96] R. M. Simmons, J. T. Finer, S. Chu, and J. A. Spudich. "Quantitative measurements of force and displacement using an optical trap". *Biophysical Journal*, Vol. 70, No. 4, pp. 1813–1822, 1996.

[Sims 99] D. Simson, M. Strigl, M. Hohenadl, and R. Merkel. "Focus: Breaking Single Protein Bonds". *Phys. Rev. Lett*, Vol. 83, p. 652, 1999.

[Smit 01] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante. "The bacteriophage &phis; 29 portal motor can package DNA against a large internal force". *Nature*, Vol. 413, No. 6857, pp. 748–752, 2001.

[Smit 03] S. B. Smith, Y. Cui, and C. Bustamante. "[7] Optical-trap force transducer that operates by direct measurement of light momentum". *Methods in enzymology*, Vol. 361, pp. 134–162, 2003.

[Smit 06] S. B. Smith and C. Bustamante. "Optical beam translation device and method using a pivoting optical fiber". *US Patent*, Vol. 7, pp. 274,451 B2, 2006.

[Smit 13] S. B. Smith and C. Rivetti. "http://tweezerslab.unipr.it". *web*, 2013.

[Smit 82] P. Smith, P. Maloney, and A. Ashkin. "Use of a liquid suspension of dielectric spheres as an artificial Kerr medium". *Optics Letters*, Vol. 7, No. 8, pp. 347–349, 1982.

[Smit 96] 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.

[Sola 14] A. Solanki, K. Neupane, and M. T. Woodside. "Single-Molecule Force Spectroscopy of Rapidly Fluctuating, Marginally Stable Structures in the Intrinsically Disordered Protein α-Synuclein.". *Physical Review Letters*, Vol. 112, No. 15, pp. 158103–158103, 2014.

[Solo 11] S. V. Solomatin, M. Greenfeld, and D. Herschlag. "Implications of molecular heterogeneity for the cooperativity of biological macromolecules". *Nature structural & molecular biology*, Vol. 18, No. 6, pp. 732–734, 2011.

[Stan 13] T. Stangner, C. Wagner, D. Singer, S. Angioletti-Uberti, C. Gutsche, J. Dzubiella, R. Hoffmann, and F. Kremer. "Determining the Specificity of Monoclonal Antibody HPT-101 to Tau-Peptides with Optical Tweezers". *ACS nano*, Vol. 7, No. 12, pp. 11388–11396, 2013.

[Stig 11] 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.

[Stou 01] A. L. Stout. "Detection and characterization of individual intermolecular bonds using optical tweezers". *Biophysical journal*, Vol. 80, No. 6, pp. 2976–2986, 2001.

[Stri 99] M. Strigl, D. A. Simson, C. M. Kacher, and R. Merkel. "Force-induced dissociation of single protein A-IgG bonds". *Langmuir*, Vol. 15, No. 21, pp. 7316–7324, 1999.

[Stru 00] T. Strunz, K. Oroszlan, I. Schumakovitch, H.-J. Güntherodt, and M. Hegner. "Model energy landscapes and the force-induced dissociation of ligand-receptor bonds". *Biophysical journal*, Vol. 79, No. 3, pp. 1206–1212, 2000.

[Sund 02] E. J. Sundberg and R. A. Mariuzza. "Molecular recognition in antibody-antigen complexes". *Advances in protein chemistry*, Vol. 61, pp. 119–160, 2002.

[Svob 94a] K. Svoboda and S. M. Block. "Biological applications of optical forces". *Annual review of biophysics and biomolecular structure*, Vol. 23, No. 1, pp. 247–285, 1994.

[Svob 94b] K. Svoboda and S. M. Block. "Force and velocity measured for single kinesin molecules". *Cell*, Vol. 77, No. 5, pp. 773–784, 1994.

[Take 00] J. Takei, R.-A. Chu, and Y. Bai. "Absence of stable intermediates on the folding pathway of barnase". *Proceedings of the National Academy of Sciences*, Vol. 97, No. 20, pp. 10796–10801, 2000.

[Tan 05] Z.-J. Tan and S.-J. Chen. "Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte". *The Journal of chemical physics*, Vol. 122, p. 044903, 2005.

[Tan 06] Z.-J. Tan and S.-J. Chen. "Nucleic acid helix stability: effects of salt concentration, cation valence and size, and chain length". *Biophysical journal*, Vol. 90, No. 4, pp. 1175–1190, 2006.

[Tan 07] Z.-J. Tan and S.-J. Chen. "RNA Helix Stability in Mixed Na^+ Mg^{2+} Solution". *Biophysical journal*, Vol. 92, No. 10, pp. 3615–3632, 2007.

[Tan 08] Z.-J. Tan and S.-J. Chen. "Salt dependence of nucleic acid hairpin stability". *Biophysical journal*, Vol. 95, No. 2, pp. 738–752, 2008.

[Tan 09] Z.-J. Tan and S.-J. Chen. "Predicting electrostatic forces in RNA folding". *Methods in enzymology*, Vol. 469, pp. 465–487, 2009.

[Thie 08] M. C. Thielges, J. Zimmermann, W. Yu, M. Oda, and F. E. Romesberg. "Exploring the energy landscape of antibody- antigen complexes: protein dynamics, flexibility, and molecular recognition". *Biochemistry*, Vol. 47, No. 27, pp. 7237–7247, 2008.

[Thor 07] I. F. Thorpe and C. L. Brooks. "Molecular evolution of affinity and flexibility in the immune system". *Proceedings of the National Academy of Sciences*, Vol. 104, No. 21, pp. 8821–8826, 2007.

[Tino 02] I. Tinoco Jr and C. Bustamante. "The effect of force on thermodynamics and kinetics of single molecule reactions". *Biophysical chemistry*, Vol. 101, pp. 513–533, 2002.

[Tino 10] I. Tinoco Jr. "RNA reactions one molecule at a time". *Cold Spring Harbor Perspectives in Biology*, Vol. 2, p. a003624, 2010.

[Tino 71] I. Tinoco Jr, O. C. Uhlenbeck, and M. D. Levine. "Estimation of secondary structure in ribonucleic acids". *Nature*, Vol. 230, No. 5293, pp. 362–367, 1971.

[Tino 73] I. Tinoco Jr, P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla. "Improved estimation of secondary structure in ribonucleic acids". *Nature*, Vol. 246, No. 150, pp. 40–41, 1973.

[Toan 06] N. M. Toan and C. Micheletti. "Inferring the effective thickness of polyelectrolytes from stretching measurements at various ionic strengths: applications to DNA and RNA". *Journal of Physics: Condensed Matter*, Vol. 18, No. 14, p. S269, 2006.

[Toya 10] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and N. Sano. "Experimental demonstration of information-to-energy conversion and validation of the generalizd Jarzynski equality". *Nature Physics*, Vol. 6, pp. 988–992, 2010.

[Tskh 97a] L. Tskhovrebova, J. Trinick, J. Sleep, and R. Simmons. "Elasticity and unfolding of single molecules of the giant muscle protein titin". *Nature*, Vol. 387, No. 6630, pp. 308–312, 1997.

[Tskh 97b] L. Tskhovrebova, J. Trinick, J. Sleep, and R. Simmons. "Elasticity and unfolding of single molecules of the giant muscle protein titin". *Nature*, Vol. 387, No. 6630, pp. 308–312, 1997.

[Vant 84] J. H. Van't Hoff. *Etudes de dynamique chimique*. F. Muller & Company, 1884.

[Vier 07] J. Vieregg, W. Cheng, C. Bustamante, and I. Tinoco Jr. "Measurement of the effect of monovalent cations on RNA hairpin stability". *Journal of the American Chemical Society*, Vol. 129, No. 48, pp. 14966–14973, 2007.

[Viss 96] K. Visscher, S. P. Gross, and S. M. Block. "Construction of multiple-beam optical traps with nanometer-resolution position sensing". *Selected Topics in Quantum Electronics, IEEE Journal of*, Vol. 2, No. 4, pp. 1066–1076, 1996.

[vZiv 10] J. Živković, M. Mitrović, L. Janssen, H. A. Heus, B. Tadić, and S. Speller. "Network theory approach for data evaluation in the dynamic force spectroscopy of biomolecular interactions". *EPL (Europhysics Letters)*, Vol. 89, No. 6, p. 68004, 2010.

[vZol 13] G. Žoldák and M. Rief. "Force as a single molecule probe of multidimensional protein energy landscapes". *Current opinion in structural biology*, Vol. 23, No. 1, pp. 48–57, 2013.

[Wagn 11] C. Wagner, D. Singer, O. Ueberschär, T. Stangner, C. Gutsche, R. Hoffmann, and F. Kremer. "Dynamic force spectroscopy on the binding of monoclonal antibodies and tau peptides". *Soft Matter*, Vol. 7, No. 9, pp. 4370–4378, 2011.

[Walt 94] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Müller, D. H. Mathews, and M. Zuker. "Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding". *Proceedings of the National Academy of Sciences*, Vol. 91, No. 20, pp. 9218–9222, 1994.

[Wang 02] G. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans. "Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales". *Physical Review Letters*, Vol. 89, No. 5, p. 050601, 2002.

[Wang 97] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block. "Stretching DNA with optical tweezers". *Biophysical journal*, Vol. 72, No. 3, pp. 1335–1346, 1997.

[Wats 53] J. D. Watson, F. H. Crick, *et al.* "Molecular structure of nucleic acids". *Nature*, Vol. 171, No. 4356, pp. 737–738, 1953.

[Wen 07] J.-D. Wen, M. Manosas, P. T. Li, S. B. Smith, C. Bustamante, F. Ritort, and I. Tinoco Jr. "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.

[Wilk 53] M. Wilkins and J. Randall. "Crystallinity in sperm heads: Molecular structure of nucleoprotein *in vivo*". *Biochimica et biophysica acta*, Vol. 10, pp. 192–193, 1953.

[Will 96] D. J. Williams and K. B. Hall. "Thermodynamic comparison of the salt dependence of natural RNA hairpins and RNA hairpins with non-nucleotide spacers". *Biochemistry*, Vol. 35, No. 46, pp. 14665–14670, 1996.

[Wilt 09] D. J. Wilton, R. Kitahara, K. Akasaka, M. J. Pandya, and M. P. Williamson. "Pressure-dependent structure changes in barnase on ligand binding reveal intermediate rate fluctuations". *Biophysical journal*, Vol. 97, No. 5, pp. 1482–1490, 2009.

[Wood 06a] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block. "Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid". *Science*, Vol. 314, No. 5801, pp. 1001–1004, 2006.

[Wood 06b] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block. "Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins". *Proceedings*

of the National Academy of Sciences, Vol. 103, No. 16, pp. 6190–6195, 2006.

[Wood 08] M. T. Woodside, C. García-García, and S. M. Block. "Folding and unfolding single RNA molecules under tension". *Current Opinion in Chemical Biology*, Vol. 12, pp. 640–646, 2008.

[Wuit 00] G. J. Wuite, S. B. Smith, M. Young, D. Keller, and C. Bustamante. "Single-molecule studies of the effect of template tension on T7 DNA polymerase activity". *Nature*, Vol. 404, No. 6773, pp. 103–106, 2000.

[Xia 98] T. Xia, J. SantaLucia, M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and D. H. Turner. "Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs". *Biochemistry*, Vol. 37, No. 42, pp. 14719–14735, 1998.

[Yang 12] P. M. Yangyuoru, S. Dhakal, Z. Yu, D. Koirala, S. M. Mwongela, and H. Mao. "Single-molecule measurements of the binding between small molecules and DNA aptamers". *Analytical chemistry*, Vol. 84, No. 12, pp. 5298–5303, 2012.

[Yasu 98] R. Yasuda, H. Noji, K. Kinosita Jr, and M. Yoshida. "F_1-ATPase Is a Highly Efficient Molecular Motor that Rotates with Discrete 120^o Steps". *Cell*, Vol. 93, No. 7, pp. 1117–1124, 1998.

[Yu 12] H. Yu, A. N. Gupta, X. Liu, K. Neupane, A. M. Brigley, I. Sosova, and M. T. Woodside. "Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates". *Proceedings of the National Academy of Sciences*, Vol. 109, No. 36, pp. 14452–14457, 2012.

[Zema 97] M. W. Zemansky and R. H. Dittman. *Heat and thermodynamics: an intermediate textbook*. McGraw-Hill, 1997.

[Zhan 09] X. Zhang, K. Halvorsen, C.-Z. Zhang, W. P. Wong, and T. A. Springer. "Mechanoenzymatic cleavage of the ultralarge vascular protein von Willebrand factor". *Science*, Vol. 324, No. 5932, pp. 1330–1334, 2009.

[Zhou 10] H.-X. Zhou *et al.* "Rate theories for biologists". *Quarterly reviews of biophysics*, Vol. 43, No. 2, pp. 219–293, 2010.

[Zlat 03] J. Zlatanova and S. H. Leuba. "Magnetic tweezers: a sensitive tool to study DNA and chromatin at the single-molecule level". *Biochemistry and Cell Biology*, Vol. 81 (3), pp. 151–159, 2003.

[Zlat 08] J. Zlatanova, S. M. Lindscay, and S. H. Leuba. "Single molecule force spectroscopy in biology using the atomic force microscope". *Progress in Biophysics and Molecular Biology*, Vol. 74, pp. 37–61, 2008.

[Znos 02] B. M. Znosko, S. B. Silvestri, H. Volkman, B. Boswell, and M. J. Serra. "Thermodynamic parameters for an expanded nearest-neighbor model for the formation of RNA duplexes with single nucleotide bulges". *Biochemistry*, Vol. 41, No. 33, pp. 10406–10417, 2002.

[Zuke 03] M. Zuker. "Mfold web server for nucleic acid folding and hybridization prediction". *Nucleic acids research*, Vol. 31, No. 13, pp. 3406–3415, 2003.

[Zwan 01] R. Zwanzig. *Nonequilibrium statistical mechanics*. Oxford University Press, 2001.

List of publications

- **A. Alemany** & F. Ritort. "Determination of the elastic properties of short ssDNA molecules by mechanically folding and unfolding DNA hairpins". *Biopolymers*, in press (2014).
- **A. Alemany**, N. Sanvicens, S. de Lorenzo, M. P. Marco & F. Ritort. "Bond Elasticity Controls Molecular Recognition Specificity in Antibody–Antigen Binding". *Nano letters*, 13(11): 5197-5202 (2013).
- **A. Alemany**, S. Frutos & F. Ritort Pinzas ópticas: principios y usos en el ámbito de la biofísica. *Newsletter CIBER-BBN*, 16 (2013).
- **A. Alemany**, M. Ribezzi-Crivellari & F. Ritort. "Recent progress in fluctuation theorems and free energy recovery". Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond (Wiley-VCH, Weinheim, 2012; ISBN 978-3-527-41094-1). R.Klages, W.Just, C.Jarzynski (Eds.)
- **A. Alemany**, A. Mossa, I. Junier & F. Ritort. "Experimental free-energy measurements of kinetic molecular states using fluctuation theorems". *Nature Physics*, 8(9): 688-694 (2012).
- C. V. Bizarro*, **A. Alemany*** & F. Ritort. "Non-specific binding of Na^+ and Mg^{2+} to RNA determined by force spectroscopy methods". *Nucleic acids research* 40 (14): 6922-6935 (2012). *Both authors contributed equally to the manuscript.
- **A. Alemany**, M. Ribezzi-Crivellari & F. Ritort. "Recent progress in fluctuation theorems and free energy recovery". *AIP Proceedings conference*, 1332: 96-110 (2011).
- S. Engel, **A. Alemany**, N. Forns, P. Maass & F. Ritort. "Folding and unfolding of a triple-branch DNA molecule with four conformational states". *Philosophical Magazine 91* (13-15): 2049-2065 (2011).
- **A. Alemany** & F. Ritort. "Fluctuation theorems in small systems: extending thermodynamics to the nanoscale". *Europhysics News* 41(2):27-30 (2010).

List of abbreviations

A	...	Adenine
aa	...	aminoacid
AD bead	...	Antidigoxigenin-coated bead
AFM	...	Atomic force microscope
BE	...	Bell-Evans
BSA	...	Bovine serum albumin
C	...	Cytosine
CC	...	Counterion condensation
CFR	...	Crooks fluctuation relation
DFS	...	Dynamic force spectroscopy
DHS	...	Dudko-Hummer-Szabo
DNA	...	Deoxyribonucleic acid
ds	...	double-stranded
EFR	...	Extended fluctuation relation
EJE	...	Extended Jarzynski equality
FDC	...	Force-distance curve
FEL	...	Free-energy landscape
FJC	...	Freely-jointed chain
FRET	...	Förster resonance energy transfer
G	...	Guanine
HCH	...	Horseshoe crab hemocyanin
WLC	...	Worm-like chain
Mab	...	Monoclonal antibody
MB	...	Methyl-boldenone
MT	...	Magnetic tweezers
NN	...	Nearest-neighbor
OT	...	Optical tweezers
Pab	...	Polyclonal antibody (after immunization)
PreI	...	Pre-immunological antibodies
RNA	...	Ribonucleic acid
SA bead	...	streptavidin-coated bead
ss	...	single-stranded
T	...	Thymine
TBI	...	Tightly bound ion
TS	...	Transition state
TST	...	Transition state theory
U	...	Uracil
UNZ	...	Unzipping
UO	...	Unified-oligonucleotide