Kinetics and thermodynamics of unfolding processes in DNA molecules with several conformational states: theory and experiments

Dissertationsschrift

zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
an der Universität Osnabrück
Fachbereich Physik

vorgelegt von

Dipl.-Ing. Sandra Nostheide

Betreuer: Prof. Dr. Philipp Maaß

Osnabrück, 2014

Abstract

Single-molecule experiments provide new insights into biological processes which were hitherto not accessible by use of measurements performed on bulk systems. Here we investigate the kinetics of a triple-branch DNA molecule with four conformational states employing pulling experiments with optical tweezers and theoretical modelling. In the folding and unfolding trajectories three distinct force rips related to the transitions between the conformational states are observed. Probability distributions for the first rupture forces of the distinct transitions are calculated by applying transition rate theory in combination with a free energy model. The comparison of the theoretical predictions with the experimental findings reveals a good agreement. In addition, due to our specific design of the triple-branch molecule, we recover a useful method to identify permanently frayed molecules by estimating the number of opened base pairs with the aid of the measured force jump values.

Furthermore, we investigate DNA hairpin molecules with periodic base sequences. They can be expected to exhibit a regular coarse-grained free energy landscape as a function of the number of open base pairs and the exerted mechanical force. Using a commonly applied model for the force protocol, we first analyse for which types of sequences a rather simple structure of the energy landscape is predicted, where the forward and backward energy barriers between partly unfolded states decrease linearly with force. By means of kinetic Monte Carlo simulations, stochastic unfolding trajectories for molecules of that kind are then generated. Introducing probabilities that can be sampled from these trajectories, the procedure to estimate parameters characterising the free energy landscape is outlined. Already 300 trajectories, as typically measured in single-molecule experiments, provide faithful results for the energetic parameters.

Moreover, we model pulling experiments conducted with optical tweezers of five DNA hairpin molecules with different length of the end-loop. Using the theoretically more sophisticated distance protocol, we study the complex free energy landscape including all possible intermediates going beyond the typically assumed two-state model. By dint of kinetic Monte Carlo simulations, unfolding trajectories in form of force-distance curves are generated as surrogate for experimental ones. A comparison of the associated work histograms with recent experiments showed a good agreement. The simulated results were furthermore verified by extracting free energy differences with the help of the Jarzynski equality. Eventually, based on a high number of simulated surrogate data, we provide a recipe for unidirectional processes improving the extracted free energy differences when few data is available even in the highly dissipative cases.

Contents

Chapter 1

Introduction

1.1 Single-molecule experiments (SMEs)

In recent years, single-molecule experiments (SMEs) have become of paramount importance in biophysical research due to the great progress in nano- and microscale manufacturing technologies, which facilitated the design of scientific instruments with sufficient sensitivity and precision to enable the controlled manipulation of individual molecules (for reviews see, for instance, [137, 68, 87]). In contrast to the traditionally employed bulk assays, where individual biomolecular dynamics can get masked, SMEs provide new insights into both thermodynamics and kinetics of biophysical and biochemical processes which were hitherto not accessible. They complement standard spectroscopy and microscopy methods applied in molecular biology and biochemistry. Therefore, they have to be regarded as an important source of additional information that is supporting the interpretation of biomolecular processes. Furthermore, SMEs permit the measurement of small amounts of energies and the detection of large fluctuations, which in turn are a good starting point to scrutinise the fundamental principles of statistical mechanics.

The manipulation of single molecules offers a powerful new tool in molecular and cellular biophysics and enables the exploration of processes which occur inside the cell at an unprecedented level. To instance just a few of the recently investigated biochemical processes: the transport of matter through pores or channels [109, 110, 6], interactions between DNA and proteins [91, 85, 61] or DNA and RNA [170], the motion of single-molecular motors [12, 22, 161], DNA transcription and replication [167, 100], virus infection [149, 43], DNA condensation [141] and ATP generation [169]. In addition, the structure of biological networks [158] and the viscoelastic and rheological properties of the DNA [150, 159, 38, 155, 62, 14] have been studied.

One class of the SMEs comprises unfolding and folding of DNA [29] (or other biomolecules) under application of mechanical forces by optical tweezers (OTs) [97, 28, 58, 165, 166, 117, 104, 70, 36], magnetic tweezers (MTs) [56, 33, 49,

24, 124] or an atomic force microscope (AFM) [79, 69, 172, 86, 4]. Chapter 2 introduces some basic concepts and terms common in the world of biomolecular studies, like the free energy landscape (FEL), and further information about these experimental setups. In this thesis we will focus mainly on OTs [113, 48, 65]. By means of an optical trap that is generated by a focused laser beam, this useful technique renders it possible to exert forces on micron sized objects. Sub-piconewton and sub-nanometer resolution in force and extension can be achieved, respectively [121]. Accordingly, one can study force-induced folding-unfolding dynamics and gain insight into related processes occuring in the cell and typical bond forces. The unfolding of DNA molecules is of particular interest, where the hydrogen bonds between complementary base pairs (bps) are disrupted. This so-called unzipping is connected to the DNA replication mechanism, in which in the initiation process particular DNA segments of the genome are exposed to the replication machinery. Afterwards, the advancing replication fork untwines and unzips the DNA strands. The essentials of this complex natural procedure have been investigated in the unfolding experiments discussed in ch. 3. Thereby both ends of the DNA molecule to be studied are tethered to a bead. One of these beads is held fixed, while the other one is pulled apart. The bps linking the double-strand break and the DNA molecule unfurls like a zipper.

Unzipping occurs as a stochastic process and depends on the pulling force, if applicable their rate and further microsopic details of the respective molecule [76]. Typically, DNA unfolds in the range of about 15 pN [137]. This force value depends on the respective sequence [33]. Above this value, the operation itself proceeds fast yielding a characteristic force pattern which is a fingerprint of the examined DNA sequence [86]. Consequently, repeated unzipping offers valuable information for DNA sequencing [26, 27, 7, 157, 41].

An interesting domain in biophysical studies is the research of junctions in molecules due to the manifold ways to interact with other substances, for example cations. Of great interest in this regard are three-way junctions, since metal ions such as magnesium can bind to them and alter the tertiary structure. In ch. 3 a first step of such a study is presented, where a molecule with a three-way junction alone, without cation binding, is investigated.

Many of the SMEs have so far focused on molecules with a rather simple FEL of only two states, a folded and an unfolded state, or including an additional misfolding state which is leading to distinct kinetic pathways. In ch. 3 we will treat a richer situation, where metastable states arise during the folding-unfolding route as intermediates. In this connection, we will address the following key questions:

(i) Can a corresponding molecule with such intermediate states be designed based on a suitable FEL model?

(ii) Is it possible to observe these intermediate states by perfoming pulling experiments with OTs?

(iii) Can phenomenological Bell-Evans kinetic models be used in order to describe such a folding-unfolding process which is comprising intermediate states? Specifically, how well are the experimental first rupture force distributions predicted by the kinetic theory?

Another significant aspect which we studied explicitely is the heterogeneity of molecular folding-unfolding behaviour observed in our experimental findings. The recorded force-distance curves (FDCs) of several molecules could be classified into different reproducible patterns. We found a useful method to identify irreversible molecular fraying, an unwanted phenomenon often arising in single-molecule studies.

1.2 Stochastic thermodynamics of small systems

In the last two decades, the accessible range of exact thermodynamic results has been shifted from the linear response regime, which is close to equilibrium, deep into nonequilibrium [144]. In stochastic thermodynamics, the distribution functions of typical thermodynamic quantities are treated on a mesoscopic level [116]. For example, the exchanged heat or the applied work of small systems with considerable fluctuations are defined along single trajectories which are accessible both experimentally and numerically [138, 139]. These fluctuations, which are typically non-Gaussian [144], arise from the random motions of atoms and molecules of the system of interest and the heat bath. Due to the strong influence of thermal noise, heat and work become stochastic variables and may no longer be expressed by a single value, but a distribution is required. The ensemble average though corresponds to the conventional thermodynamic quantity [47].

Typical small systems comprise colloidal particles, enzymes, molecular motors and biopolymers, such as DNA, RNA or proteins [18, 144]. A common feature is that all these systems are surrounded by an aqueous solution, a heat bath, of well-defined temperature. For these systems three types of different nonequilibrium situations exist [144]:

1. Relaxation from nonequilibrium towards equilibrium
2. Time-independent external driving, leading to a nonequilibrium steady state
3. Time-dependent external driving, e. g. by force application.

In this work we are particularly interested in the third point due to its connection to DNA replication (cp. sec. 1.1).

One important aspect is the time-scale separation between the rather slow, and thus observable, states of the system and the unobservable, fast degrees of freedom [144]. The latter arise from the heat bath or even internal degrees of freedom of the investigated biopolymers. They are forced to be always in equilibrium which is specified by the immediate values of the slow degrees of freedom. Changes in the

state of the system can be caused by the external driving or by the thermal noise and finally determine the single trajectory. The characteristics of fluctuations, external driving and the distribution of the inital state specify completely an ensemble of such trajectories.

Due to the time-scale separation, the dynamics are Markovian [144], i. e. the system has no memory of past events: only the actual state influences the next one. In case of a continuous description of the states, e. g. via position, the dynamics can be described by a Langevin equation for a single system and, on the ensemble level, by a Fokker-Planck equation. Discrete states, as used in this work, can be treated by a master equation, again for an ensemble [144].

A fundamental aspect of folding-unfolding studies (cp. sec. 1.1) is to obtain information on the free energy difference (FED) between the folded and unfolded state or, generally speaking, on the FEL, which characterises both the energy levels of the stable as well as the metastable, partly unfolded states and the energy barriers in-between them (see ch. 4). The easiest way to determine FEDs are equilibrium methods, where Boltzmann statistics is applied to the fraction of residence times of the states at different values of the control parameters [97, 117, 104, 93, 2, 136]. Yet, the energy barriers between the states may be so high that equilibrium can not be achieved within the available experimental time window [97, 140, 25, 2].

An intriguing way to measure FEDs by dint of out-of-equilibrium measurements is rendered possible by the detailed and integral work fluctuation theorems [47, 144], such as the Crooks fluctuation theorem [31] or its integral counterpart, the Jarzynski equality [80]. Similar to the universality of the Boltzmann distribution in equilibrium, these theorems hold true in an universal manner and are independent of microscopic details of the molecule and its unfolding/folding dynamics. They consider the thermodynamic work related to a certain driving protocol $\lambda(t)$ of control variables in a time interval $[t_i, t_f]$, e. g. pulling a molecule with a force $f(t)$.

According to the Crooks fluctuation theorem, the work probability distributions (WPDs) $p(W)$ for a protocol $\lambda(t)$ and $p_R(W)$ for the associated reversed protocol $\lambda_R(t) = \lambda(t_i + t_f - t)$ are related via [31]

$$\frac{p(W)}{p_R(-W)} = \mathrm{e}^{(W-\Delta G)/k_\mathrm{B}T} \,. \tag{1.1}$$

The difference $\Delta G = G_f - G_i$ of the (equilibrium) free energies $G_{i,f}$ of the molecule in the macrostates, which are specified by the control variables $\lambda_{i,f} = \lambda(t_{i,f})$, was used and $k_\mathrm{B}T$ is the thermal energy. The dissipated work, leading to an entropy increase, is defined as $W_\mathrm{diss} = \langle W \rangle - \Delta G$ [80]. This work fluctuation theorem is quite useful to extract equilibrium information out of nonequilibrium data since it is robust and converges quickly [28]. If data of both forward and reverse process is available, one can extract the FED out of the crossing point of both WPDs, i. e. when $p(W) = p_R(W)$, since they intersect at $W = \Delta G$ [28, 139, 117]. The Bennet acceptance ratio method [9] is often applied to extract the optimal estimate of the FED. An extended fluctuation relation for kinetic states, playing an important role

in misfolding that is connected to some diseases, was proposed in [107, 83] and experimentally tested in [2].

In cases where the Crooks fluctuation theorem [31] can not be applied, for instance, since the work for the reverse process can not be obtained with sufficient statistics [73, 33, 64] (cf. ch. 3), an alternative approach is to use the Jarzynski equality [173]. It also connects the FED ΔG between two equilibrium states, equivalent to the reversible work needed to transform the system from the initial to the final state, to an arbitrary nonequilibrium work W, corresponding to the thermodynamic work done on the system. This theorem [80],

$$\langle \mathrm{e}^{-W/k_\mathrm{B}T} \rangle = \int \mathrm{d}W p(W) \mathrm{e}^{-W/k_\mathrm{B}T} = \mathrm{e}^{-\Delta G/k_\mathrm{B}T} , \tag{1.2}$$

has been applied successfully in various cases [96, 140, 174, 33, 15, 75, 11, 64]. Unfortunately, it may be difficult to apply the Jarzynski equality, since $\langle \mathrm{e}^{-W/k_\mathrm{B}T} \rangle$ is dominated by rare trajectories with work values $W < \Delta G$ [80, 140, 174, 55, 78, 81, 82, 34, 88], so that a very large number of trajectories is required [140]. Extending the histograms of (typical) work values to the tail regime $W \ll \Delta G$ by using a fit to theoretical predictions of the generic behaviour in the tail regime [55, 40, 45, 122, 128] might be a solution. But even when assuming the functional form of the tail, the extraction of reliable free energy estimates from the measured data is sophisticated, since one must include the dependence of biasing effects on the number of measurements (single unfolding trajectories) correctly [163, 140, 174, 55, 52, 145, 139, 128].

In case these nonequilibrium fluctuation theorems may not be used due to insufficient statistics, one may alternatively extract the FED, or more generally, the FEL, by means of an analysis of the unfolding/folding kinetics and a comparison with predictions from theories. This method is no longer universal since specific parameters for the studied molecule need to be introduced. For instance, the FEL can be determined out of the first rupture force distributions [117]. In a DNA molecule with two states, these distributions of the unfolding and folding process are related to the distributions of the survival probabilities to stay in the unfolded and folded state, respectively. Provided that the loading/unloading rates are constant, a clever representation [117] of these survival probabilities enables the determination of both the energetic barriers and the FED between the two states.

A good theoretical description of the kinetics of the DNA unfolding and folding processes is required in order to provide faithful results. Typically, coarse-grained approaches are applied, where the molecular states are characterised by the number of open bps n, as successfully modelled in the past [165, 25, 102, 143, 99, 105, 71, 36]. An important aim is the optimisation of the model parameters by appropriate measurements and the determination of their variation when the conditions of the environment change, for instance temperature, salt concentration or pH value of the solvent.

The focus of analytical treatments so far lied on two-state folders. However, in experiments long molecules are frequently investigated, featuring various minima

in the FEL which are related to the partly unfolded states. It is advantageous that a single unfolding/folding process of these sequences includes numerous transitions, each yielding valuable information on the FEL. Apart from that, it is difficult to treat the unfolding and folding kinetics of such multistate folders theoretically, since, generally speaking, each transition between consecutive minima is characterised by distinct barriers and energy differences. It is thus useful to consider periodic DNA molecules, where the transitions between the states are characterised by the same energetic parameters. In ch. 4 we will model the unfolding kinetics of such periodic DNA sequences and study if it is possible to extract reliable information on the FEL on the basis of a small number of unfolding trajectories. Applying standard modelling for a force protocol (FP), we first illustrate that simple regular FELs can be obtained for periodic sequences, whose characteristics remain valid over a large range of pulling forces. Subsequently, kinetic Monte Carlo (KMC) simulations are used in order to generate unfolding trajectories for these sequences, as surrogate for experimental ones. By fitting the extracted survival and persistence probabilities to analytical results, we demonstrate that already about 300 trajectories, as typically recovered in SMEs, are sufficient to yield good estimates of the FEL parameters. The chapter is completed with information on how to verify this procedure experimentally with OTs as well as with MTs.

In ch. 5 we will model the unfolding kinetics of five short DNA hairpins with the same stem, but different end-loop size [128], for which experimental data is available for comparison. By dint of the Reaction Time Algorithm (RTA) [44, 66] we simulate pulling experiments performed with OTs in order to obtain force-distance curves and work histograms. We include a detailed FEL model to correctly take into account the different components of the whole molecular setup. Furthermore, we use a distance protocol (DP), as should be ideally applied for typical pulling experiments where the distance between trap and pipette is controlled [105], rather than the FP which is often used for theoretical descriptions [117, 51, 2] (cf. ch. 3). In particular, we do not restrict ourselves to a two-state model, but include all possible intermediates taking into account the topological changes in the FEL. This seems to be impossible in view of the complexity of the FEL of real molecules. Nevertheless, in our approach we make use of Bell-Evans kinetic rates [8] and take specific effects and efficient algorithms into consideration. We further investigate the simulated results by extracting the FEDs with the help of the Jarzynski equality. In particular, we focus on the highly dissipative cases where it is difficult to estimate the FEDs well in case of unidirectional processes. By dint of a high number of simulated surrogate data, we complement the recipe of [128] in order to improve the extracted free energy differences when few data is available.

Chapter 2

General background

2.1 DNA as an important biological building block

Molecular biophysics deals with the research of the structure, function and kinetics of biomolecules, the building blocks of the cell. To be specific, the biopolymers DNA, RNA and proteins play a major role in the proper operation of biological processes. Particularly, in the gene expression (see fig. 2.1) [30, 17] the first step involves the DNA replication which creates DNA duplicates. Afterwards, the inheritable information is transferred to the newly assembled RNA via transcription. In the last step, proteins are synthesised out of RNA with the help of ribosomes in a process called translation.

In this work we are especially interested in the deoxyribonucleic acid (DNA), a basic building block of all living organisms and DNA viruses, that stores the genetic material in the long term. DNA is a nucleic acid whose repeating monomeric subunits, the nucleotides, consist of three elemental components: a sugar, a phosphate group and a base for encoding genetic instructions in terms of base sequences. The four bases attached to each pentose sugar, the deoxiribose, are adenine (A), cytosine (C), guanine (G) and thymine (T). Basically, purine bases (G or A) compose complementary bps with pyrimidine bases (C or T). We only take this canonical Watson-Crick base pairing into account. In living cells the B-form DNA

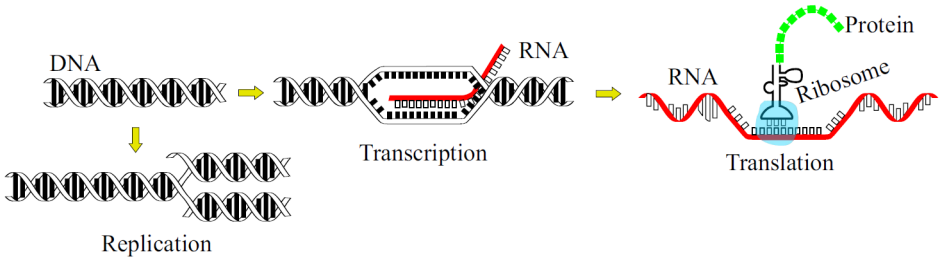


Figure 2.1: The central dogma of molecular biology [72].

predominates [160], a right handed helix with a diameter of about 2 nm. The helix twines around itself every 3.4 nm, including approximately 10 bps in every turn.

According to the classification of Linderstrøm-Lang [95], the following structural DNA[i] configurations are possible. The primary structure characterises the lowest level of structural information, i. e. the chemical composition of the nucleotide sequence. The secondary structure describes the general three-dimensional form as well as the relative arrangement of the nucleotides and is determined by hydrogen bonds between the nitrogenous bases or intramolecular double-strands, in case the segments are complementary. There exist several secondary structural motifs [101], for instance stem, junction, end-loop and interior loop.

The tertiary structure specifies the entire superordinate steric topology of the DNA, defining the atomic positions. It is stabilised by disulphide bridges as the strongest bonds, ionic linkages, hydrogen bonds as well as hydrophobic interactions[ii], which are the weakest of these bonds. Therefore, these last two bond types can be broken by the thermal fluctuations of the surrounding heat bath.

DNA typically exists in a helical conformation where two opposing strands are twisted around each other [160], stabilised by hydrogen bonds between the bases located in the interior of the helix. Due to the convolution of the double helix, the bps lie one upon the other, which leads to a further stabilisation via π-orbitals. Since these bonds can be quite easily broken and recombined, this polynucleotide is well suited for folding and unfolding experiments. This process is either induced by chemical agents, high temperature or, as studied in this work, by mechanical force. In case the double-stranded DNA (dsDNA) is forced to open by pulling on both strands from opposed ends, the process is called "shearing" [131, 118, 119]. Here we will focus on "unzipping", where the force is applied at one end of the molecule. Double-stranded DNA is thermally stable at body temperature [33]. The research of force-induced separation of dsDNA strands is thus directly connected to the unzipping process occurring in living organisms, where enzymes or further proteins act on the molecule.

2.2 Single-molecule experiments

Previously, bulk experiments have been the common way to study the physical and chemical properties of DNA, where a large number of molecules is sampled simultaneously and thus information on average molecular properties is obtained. Needless to say, this technique blocks the view towards the fundamental activities taking place at a basic level of all biological processes. In the past two decades this situation has changed vigorously. By dint of SMEs, the exploration of biological principles on the basis of single molecules is possible. Newly developed methods include the possibility to measure forces provoked by biochemical reactions and even to intervene directly in these reactions. Single-molecule methods enable the

[i] For proteins, also a quaternary structure exists.

[ii] Also known as Van der Waals forces.

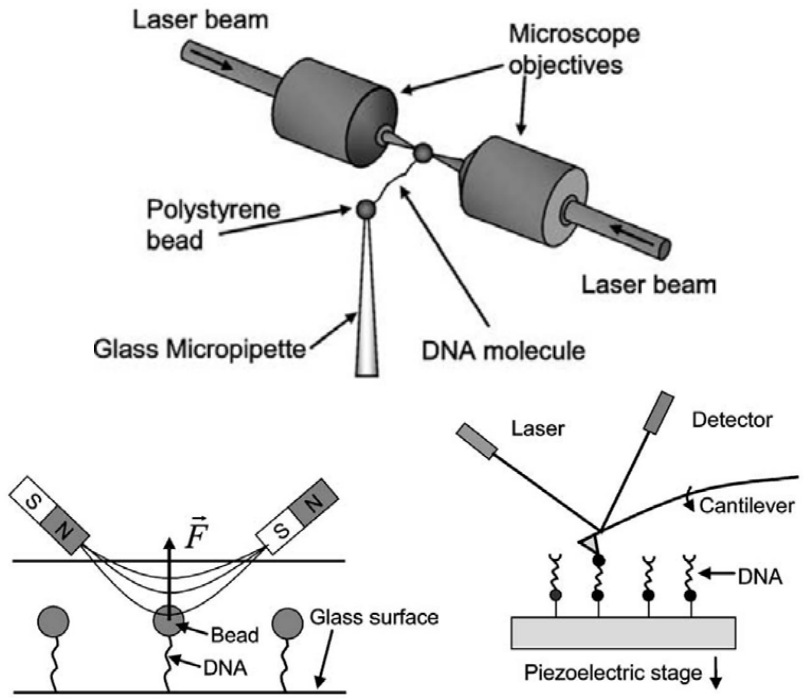


Figure 2.2: The setup of OTs (top), MTs (left) and AFM (right) [137].

investigation of the probability distribution of molecular properties. Various ways are possible to explore the folding-unfolding behaviour of dsDNA. Amongst the most often used are the ones listed below [29, 137, 59, 120, 121, 37]. An illustration of their setup is given in fig. 2.2.

- **Magnetic tweezers (MTs)**

  Especially in the force range of 10^{-2} to 100 pN [121], magnets are employed in order to pull a magnetic bead attached to one end of a DNA molecule, whose other end is fixed at a glass surface [101] by e. g. antibody-antigen interactions. With the help of magnetic traps DNA twisting is possible by rotating the magnets [56, 29, 121], like this enabling the investigation of elastic and torsional properties of DNA molecules. This versatile tool allows to work at constant force with a higher sensitivity than AFMs or OTs [121, 101] and to carry out several single-molecule measurements simultaneously [33].

- **Optical tweezers (OTs)**

  In the range of trapping forces with an order of magnitude of about 0.1 to 100 pN, OTs are well suited to apply and detect forces

on micron-sized particles in an aqueous environment [121]. The DNA molecule under study is tied to polymer spacers called handles, which in turn are, on the remaining ends, chemically linked to a bead. One of these beads is retained with the help of a pipette via air suction, the other one is optically trapped in the focus of a laser [70]. The formation of such an optical trap can be achieved by focusing an infrared laser beam onto a micron-sized spot by an objective of a microscope [120]. By dint of the pressure of radiation of the laser beam, OTs exert forces on dielectric beads [70]. The external force on the optically trapped bead can be determined by either measuring the change of its position in relation to the centre of the optical trap or, as in our case, the deflection of the trapping beam itself [72].

- **Atomic force microscope (AFM)**

  AFMs are ideally suited for the range of forces between 10 and 10^4 pN [121] and thus allow the investigation of strong to covalent interactions, like in proteins [4]. Scanning the very sharp tip of a micron-sized, flexible cantilever of known stiffness over the sample permits to measure the force acting on the tip [121, 72]. This results in a bending of the cantilever whose deflection may be monitored e. g. with the help of a laser beam [120].

2.3 Model of the experimental setup of optical tweezers

In the present work we use OTs to investigate DNA folding and unfolding processes. All experiments have been carried out with the help of a dual-beam laser instrument [70]. Further details of this specific setup can be found in [46, 72]. The schematic representation of the OT setup in fig. 2.3 shows the different components of the molecular construct[iii]. It comprises the DNA molecule under study (here the triple-branch molecule is shown, cp. ch. 3), the handles and the micron-sized polystyrene beads, captured in an optical trap and a micropipette, respectively. In-between the molecule under study and both beads so-called handles are inserted, which usually also consist of a DNA double-strand. They are polymer spacers allowing a more comfortable handling of the object under study and also avert a direct contact of the beads. Each of these polymer spacers is chemically linked to a bead. One of the beads is retained with the help of a pipette via air suction, the other one is optically trapped in a laser focus [151]. On the 5′ end of the DNA molecule, biotin is attached to enable a connection with a streptavidin-coated

[iii] Note that it is not a true-to-scale representation. Each of the two identical short dsDNA handles used for the experiments is consisting of 29 bps [51] and has a length of about 20 nm. The triple-branch molecule in this drawing is shown in the folded state. If it is totally unfolded, the molecular extension m^* is in the order of 100 nm. The beads with a diameter of 1.8 µm and 3.0 µm, respectively, are thus much bigger than shown here.

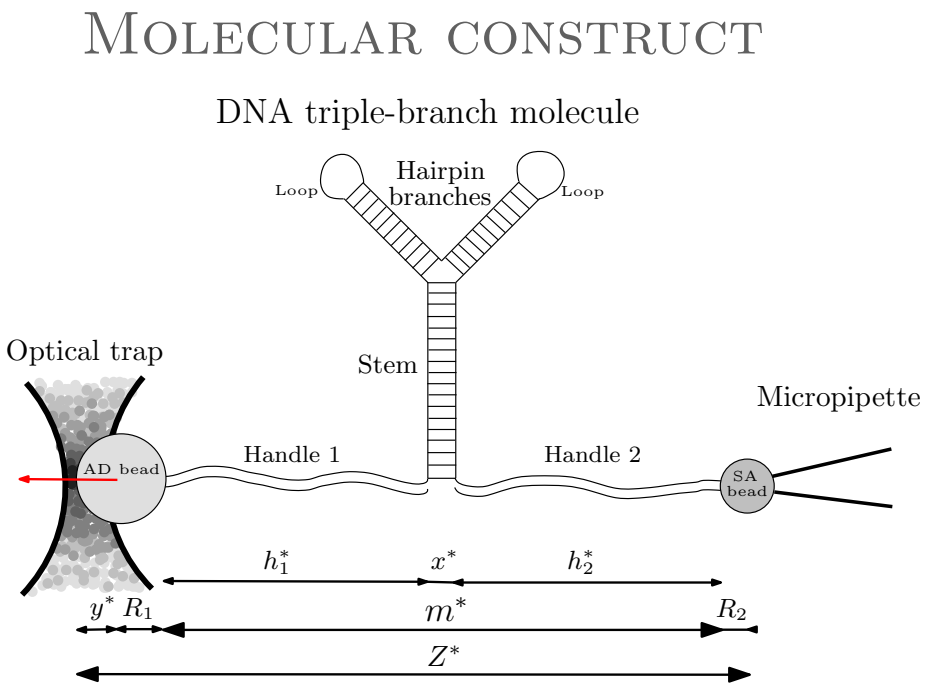


Figure 2.3: Sketch of the experimental OT setup, where the DNA triple-branch molecule is depicted as object to be studied (see ch. 3).

bead (SA bead) whose diameter is 1.8 µm. Biotin is a vitamin which establishes a strong linkage to the proteins avidin and streptavidin. The 3′ end is modified with the antigen digoxigenin, which is able to interact with an antidigoxigenin-coated bead (AD bead). The latter has a diameter of 3.0 µm.

In this simplified model [106], the configurational variables are the projections of the end-to-end distance of each element along the reaction coordinate axis, i. e. the axis along which the force is applied. An important parameter is the distance Z^* between the tip of the micropipette, fixing one polystyrene bead, and the centre of the optical trap, towards which the second polystyrene bead is pulled due to a force caused by the electric field gradients of the laser setup. This trap-pipette distance is composed of

$$Z^* = y^* + R_1 + h_1^* + x^* + h_2^* + R_2 + a^{(0)}, \tag{2.1}$$

where R_1 and R_2 are the radii of the trapped and fixed beads and $a^{(0)}$ is an arbitrary origin. The displacement of the bead in the optical trap, $y^* = f/k_\mathrm{t}$, is constructed to mediate a linear elastic behaviour in the region close to the trap centre, whereas the force is labelled as f and the trap stiffness as k_t. All variables marked by an asterisk * denote the total length of the respective element, whereas variables marked by a $^{(0)}$ designate the length of the respective element at zero force. We can thus write $y = y^* - y^{(0)}$, where $y^{(0)}$ is zero since the bead is located exactly in the centre of

the optical trap when no force is applied on it. The extension of the studied DNA molecule itself is labelled as x^*. Similarly, we have $x = x^* - x^{(0)} = u_l(f) - \delta_{n,n_{\mathrm{tot}}}\, d_0$, where d_0 is the stem diameter and $u_l(f)$ signifies the elongation of the mean end-to-end distance of the single-stranded DNA (ssDNA) in force direction. Details of the calculation of $u_l(f)$ can be found in sec. 2.6.2. h_1^* and h_2^* are the lengths of the handles which are attached to the beads and the molecule under study. The handles are assumed to have a linear elastic response to an externally applied force f, $h_i^* = f/k_{\mathrm{h},i}$, where the stiffness of each handle is $k_{\mathrm{h},i}$.

In a simplified way, one may now write eq. (2.1) as

$$Z^* = Z^{(0)} + x + h + y, \tag{2.2}$$

where we have introduced $Z^{(0)} = R_1 + R_2 + h_1^{(0)} + h_2^{(0)} + a^{(0)} + y^{(0)} + x^{(0)}$. Here $h_i^{(0)}$ denotes the handles length in the absence of a force and $h = (h_1^* - h_1^{(0)}) + (h_2^* - h_2^{(0)})$ the handle elongation due to the impact of a force. Again, $h = f/k_{\mathrm{h}}$ shows a linear elastic response and the combined stiffness of the handles k_{h} is defined as $1/k_{\mathrm{h}} = 1/k_{\mathrm{h},1} + 1/k_{\mathrm{h},2}$. We can further combine $h + y = f/k_{\mathrm{eff}}$ using the effective rigidity of bead and handles k_{eff}, given by $1/k_{\mathrm{eff}} = 1/k_{\mathrm{t}} + 1/k_{\mathrm{h}}$. Finally, we introduce $Z = Z^* - Z^{(0)} = x + y + h$ as the relative total length change in the single parts of the setup, where we have subtracted all fixed size values contained in $Z^{(0)}$,

$$Z = Z(n, f) = \frac{f}{k_{\mathrm{eff}}} + u_l(f) - d_0\, \delta_{n,n_{\mathrm{tot}}}\,. \tag{2.3}$$

Likewise, the two handles and the investigated DNA molecule have an end-to-end distance or so-called molecular extension[iv] of

$$m = m^* - m^{(0)} = x + h. \tag{2.4}$$

OTs pulling experiments permit the measurement of the force f as well as the total trap-pipette distance Z^* which can be easily converted into the relative distance Z used in the following.

2.4 Work definitions for different experimental setups

According to the first law of thermodynamics, in a closed system the change in the internal energy equals the sum of the infinitesimal change in heat δQ added to and work δW done on the system,

$$dU = \delta Q + \delta W = \left(\frac{\partial U}{\partial \rho}\right) d\rho + \left(\frac{\partial U}{\partial \lambda}\right) d\lambda\,. \tag{2.5}$$

Here λ denotes the externally controlled variable which does not fluctuate in time and ρ the configurational variable which fluctuates due to a Brownian motion [3].

[iv] The molecular extension will also be called extension and the trap-pipette distance simply distance.

Heat represents the energy supplied to the system when the configurational variable varies at fixed control parameter value, e. g. in the instant of unzipping a set of bps. In contrast, work is the energy transferred to the system by external sources when, at a given configuration, the control parameter changes. In order to choose the adequate work definition, one has to carefully take into account the experimental setup. Here we focus on MT and OT (cf. sec. 2.2). Table 2.1 gives an overview over the different work definitions which should be used for the two experimental protocols, where force and distance are controlled, respectively). Conversions between the four work quantities are also specified.

2.5 Experimental protocols and curve types of pulling experiments

Depending on the respective control parameter, two distinct experimental protocols may be considered in pulling experiments [105, 106, 72].

2.5.1 The force protocol (FP)

Using a FP, the force f and hence the position y of the centre of the AD bead relative to the centre of the optical trap are controlled by a piezo actuator. The force is thus the control parameter ($\lambda = f$), and the molecular extension $m^* = x+h+m^{(0)}$ or the trap-pipette distance Z^* are the configurational variables ($\rho = s$) (cp. sec. 2.3). Although the FP is easily theoretically accessible because the FEL depends exclusively on the DNA properties, its use is experimentally limited. It is a challenging task to work with a force-controlled protocol using single-trap OTs since the necessary force-feedback control to counterbalance force fluctuations needs a sophisticated implementation and has got a finite response time [105, 72].

The two possible representations are the distance-force curve (DFC) and the extension-force curve (EFC), where distance ($s = Z^*$) or extension ($s = m^*$) are plotted against force. For the MT setup typically the EFC is plotted (cp. e. g. [33]), while for an OT experiment both DFC and EFC are available. Both plot types are similar because only the absolute value of the ordinate is different, depending on whether distance or extension is depicted. During a disruption event the force stays at a constant value while the bps are opened, resulting in an identical length change of both extension and distance, i. e. $\Delta m^* = \Delta Z^*$, since only the hairpin extension x is changing. In this jump regime no work is done, but heat exchanged. In the stretching regime, however, work is done on the molecule: $W_f(Z^*)$ for $s = Z^*$ and $W_f(m^*)$ for $s = m^*$ (cp. tab. 2.1). The total work depends on the initial f_{in} and final f_{fin} value of the force and the configurational variable $\rho = s$ according to [3]

$$W_f(s) = \int_{f_{\text{in}}}^{f_{\text{fin}}} \delta W = \int_{f_{\text{in}}}^{f_{\text{fin}}} \left(\frac{\partial U}{\partial f}\right) df = -\int_{f_{\text{in}}}^{f_{\text{fin}}} s df \,. \tag{2.6}$$

Work definitions

Specs	Setup	FORCE PROTOCOL	DISTANCE PROTOCOL
control parameter λ		• force $\boxed{\lambda = f}$	• distance or extension [a] $\boxed{\lambda = s}$
configurational variable ρ		• distance or extension $\boxed{\rho = s}$	• force $\boxed{\rho = f}$
interesting work quantities		$\boxed{W_f(s) = -\int_{f_{in}}^{f_{fin}} s\,df}$	$\boxed{W_s(f) = \int_{s_{in}}^{s_{fin}} f\,ds}$
FEL calculation		• only contributions of the DNA hairpin	• contributions of DNA hairpin, bead in trap and handles
$\boxed{s = Z}$	OT	$\boxed{\mathbf{W_f}(\mathbf{Z}^*) = -\int_{\mathbf{f}_{in}}^{\mathbf{f}_{fin}} \mathbf{Z}^*\,\mathbf{df}}$	$\boxed{\mathbf{W_Z}(\mathbf{f}) = \int_{\mathbf{Z}_{in}}^{\mathbf{Z}_{fin}} \mathbf{f}\,\mathbf{dZ}}$ [b]
		$\boxed{\mathbf{W_f}(\mathbf{Z}^*) = \mathbf{W_Z}(\mathbf{f}) - W(Z) - W(Z^{(0)})}$	
		distance-force curve (DFC)	*force-distance curve (FDC)*
$W(Z) = Z_{fin}f_{fin} - Z_{in}f_{in}$ $W(m) = m_{fin}f_{fin} - m_{in}f_{in}$ $W(f) = \frac{f_{fin}^2 - f_{in}^2}{2k_t}$ [c] $W(Z^{(0)}) = Z^{(0)}(f_{fin} - f_{in})$ $W(m^{(0)}) = m^{(0)}(f_{fin} - f_{in})$		$\boxed{\mathbf{W_f}(\mathbf{m}^*) = \mathbf{W_Z}(\mathbf{f}) + W(f) - W(Z) - W(m^{(0)})}$ $\boxed{\mathbf{W_f}(\mathbf{Z}^*) = \mathbf{W_f}(\mathbf{m}^*) - W(f) - W(Z^{(0)}) + W(m^{(0)})}$ $\boxed{\mathbf{W_Z}(\mathbf{f}) = \mathbf{W_m}(\mathbf{f}) + W(f)}$ $\boxed{\mathbf{W_f}(\mathbf{Z}^*) = \mathbf{W_m}(\mathbf{f}) - W(m) - W(f) - W(Z^{(0)})}$	
$\boxed{s = m}$	MT	$\boxed{\mathbf{W_f}(\mathbf{m}^*) = -\int_{\mathbf{f}_{in}}^{\mathbf{f}_{fin}} \mathbf{m}^*\,\mathbf{df}}$	$\boxed{\mathbf{W_m}(\mathbf{f}) = \int_{\mathbf{m}_{in}}^{\mathbf{m}_{fin}} \mathbf{f}\,\mathbf{dm}}$ [d]
	OT [a]	$\boxed{\mathbf{W_f}(\mathbf{m}^*) = \mathbf{W_m}(\mathbf{f}) - W(m) - W(m^{(0)})}$	
		extension-force curve (EFC)	*force-extension curve (FEC)*

$Z = Z^* - Z^{(0)} = x + h + y$: relative trap-pipette distance, where all constant contributions $Z^{(0)}$ are subtracted from the absolute distance Z^*, including the molecular extension $m = m^* - m^{(0)} = x + h$ composed of the length change of the hairpin x and the handles h, and the bead displacement $y = f/k_t$ with the trap stiffness k_t
s: distance Z or extension m (or Z^* and m^*) according to the experimental setup
f_{in} and f_{fin}: inital and final values of the force f (and accordingly m, Z and s)

[a] Note that for the OT setup in the DP m is fluctuating and thus a configurational variable ($\rho = m$). [b] total accumulated work
[c] work applied to move the bead in the optical trap [d] transferred work, i. e. the work transferred to the molecular system

Table 2.1: Overview over the work definitions in force (FP) and distance protocol (DP) for two experimental setups (cp. [3] and [104]).

2.5.2 The distance protocol (DP)

The DP is the natural protocol for the experimental OT setup used throughout this work. The total distance Z^* is fixed at a certain value which evolves in time and represents the externally controlled parameter ($\lambda = s = Z^*$) [28]; both molecular extension m and force f fluctuate and are thus configurational variables, i. e. $\rho = f$ and $\rho = m$ (see [101] for a more detailed description). Compared with this, for MT the force is again a configurational variable ($\rho = f$), whereas the molecular extension is the control parameter ($\lambda = s = m$).

The force-distance curve (FDC)

During pulling experiments, which we consider in this work, the time evolution of both the force f and the trap-pipette distance Z^* between the centre of the optical trap and the tip of the micropipette are recorded. The force f can be plotted as a function of the relative distance Z, yielding a FDC (see e. g. fig. 3.3). We prefer to use this plot type since the externally controlled distance does not fluctuate, enabling a more precise analysis of data (cp. sec. 3.2).

The force-extension curve (FEC)

In the FEC, the force f is plotted as a function of the molecular extension m. The FEC is often used [97, 28, 103, 166, 53], even though m is no control parameter and therefore exposed to fluctuations.

During a disruption event, the molecular extension m increases abruptly since ssDNA is released. Consequently, the bead in the optical trap relaxes while approaching the centre of the trap, visible as the force drops by a certain value Δf while the number of opened bps is increased. Refolding the DNA molecule, however, pulls the bead away from the trap centre and leads to a suddenly rising force. For the OT setup the change in the trap-pipette distance Z is negligible during such a disruption event, so we can take $\Delta Z \simeq 0$, which is directly visible in the FDC. Using eq. (2.2), a force jump Δf can be related to the change in the hairpin extension Δx and to the change in the molecular extension Δm as

$$|\Delta x| = \frac{|\Delta f|}{k_t} \quad \text{and} \quad |\Delta m| = \frac{|\Delta f|}{k_{\text{eff}}} \,. \tag{2.7}$$

Similarly to the FP, in the jump regime of the FDC only heat is exchanged, since the trap-pipette distance is constant in the DP, whereas the work $W_Z(f)$ is only done on the molecule in the stretching regime. Only in the FEC both jump and stretching regime have to be taken into account when calculating the work $W_m(f)$. The work $W_Z(f)$ is not affected by the constant $Z^{(0)}$,

$$W_Z(f) = \int_{Z_{\text{in}}}^{Z_{\text{fin}}} f dZ = \int_{x_{\text{in}}}^{x_{\text{fin}}} f dx + \int_{h_{\text{in}}}^{h_{\text{fin}}} f dh + \frac{f_{\text{fin}}^2 - f_{\text{in}}^2}{2k_t} \,. \tag{2.8}$$

However, the situation is different with the DFC: here indeed $Z^{(0)}$ has a non negligible impact on the work value $W_f(Z)$ since the trap-pipette distance Z is plotted on the ordinate with the result that a change in the arbitrary origin $a^{(0)}$ has an effect on the surface area below the curve [129] (cp. tab. 2.1). Furthermore, one has to distinguish carefully between the accumulated work $W_Z(f)$ and the transferred work $W_m(f)$ when fluctuation theorems such as the Crooks fluctuation theorem (cp. eq. (1.1)) [31] or the Jarzynski equality (cp. eq. (1.2)) [80] are used [115]. For a detailed summary about work in the context of controlled distance and the related FDCs compare [3].

2.6 Models describing the structural conformations of polymers

In a coarse-grained description, the unfolding process of DNA (and other biomolecules) is often described through thermally activated transitions between minima of a FEL $G(n, \lambda)$ [77]. The latter may be expressed as a function of the number n of sequentially opened bps and the control parameter λ which is, depending on the experimental protocol, either the applied stretching force f or, in the case of OTs as used here, the trap-pipette distance $s = Z$ (see tab. 2.1). We can decompose the FEL as [105, 106, 117]

$$G(n, \lambda) = G_{\text{form}}(n) + G_{\text{str}}(n, \lambda), \tag{2.9}$$

where the "formation" part $G_{\text{form}}(n)$ only depends on n and quantifies the free energy release upon breaking of n bps. The "stretching" part $G_{\text{str}}(n, \lambda)$, on the other hand, also depends on the control parameter and includes, again depending on the experimental protocol, contributions of the released ssDNA, the handles, the bead in the trap and the diameter of the molecule. For the DP we consider the Helmholtz free energy, whereas for the FP the Gibbs free energy is the proper thermodynamic potential [72, 173]. Each of these contributions is treated in more detail below.

2.6.1 Free energy of formation

For the calculation of the free energy of formation $G_{\text{form}}(n)$ the nearest neighbour model [32, 39]

$$G_{\text{form}}(n) = \sum_{\mu=n+1}^{n_{\text{tot}}-1} \epsilon_{\mu,\mu+1} + G_{\text{loop}}\left(1 - \delta_{n,n_{\text{tot}}}\right), \tag{2.10}$$

is used[v], where $\epsilon_{\mu,\mu+1}$ is the interaction energy between directly adjacent bases (μ labels the bases in one of the strands). This model has been applied successfully in

[v]Note that here we consider unzipping at low temperatures below the denaturation transition of DNA molecules, where the formation of bubbles becomes negligible. At higher temperatures, where

bases	ϵ	bases	ϵ
AA, TT	-1.23	CC, GG	-1.93
AT	-1.17	CG	-2.37
TA	-0.84	GC	-2.36
AC, GT	-1.49	AG, CT	-1.36
CA, TG	-1.66	GA, TC	-1.47

Table 2.2: Nearest neighbour base pair (bp) energies (see eq. (2.10)) in kcal/mol at $T \simeq 25^{\circ}$C and 1 M monovalent salt concentration [70].

various unfolding experiments [165, 117, 25, 175, 29, 102, 71, 36, 153, 62]. The parameters of this model have been continuously improved and we use the ones recently reported in [70] (see tab. 2.2). If one applies the nearest neighbour model for instance to the sequence 5′-TCCAG...-3′ and its complementary part[vi] 3′-AGGTC...-5′, the stack energy can be written as $G_{\text{stack}} = \epsilon_{\text{TC}} + \epsilon_{\text{CC}} + \epsilon_{\text{CA}} + \epsilon_{\text{AG}} + \ldots$ For its calculation we use the nearest neighbour bp energies at $T \simeq 25^{\circ}$C and 1 M monovalent salt concentration, such as NaCl, as listed in tab. 2.2 [70]. Generally speaking, $G_{\text{form}}(n)$ is a monotonically increasing function of n with varying local rise. Yet it may decrease in the presence of entropic components such as loops.

The term G_{loop} in eq. (2.10) refers to the free energy reduction as a result of the release of the end-loop and is estimated from [143, 175], depending on the specific loop sequence. This Mfold web server provides data of calorimetric and UV absorbance bulk experiments. In case of additional structural elements, such as interior loops, further loop contributions have to be taken into account.

2.6.2 Stretching energy

Before the single contributions to the stretching energy will be discussed, two commonly used models borrowed from polymer physics, which are describing some basic structural properties of biopolymers, will be shortly introduced: the freely jointed chain (FJC) [150] and the worm-like chain (WLC) model [19, 108].

bubble formation is occuring, the standard nearest neighbour model (see eq. (2.10)) becomes less appropriate [171], and the Peyrard-Bishop-Dauxois (PBD) model [130, 35] was shown to give a better description [156]. This PBD model has been further refined by including a sequence-dependent stacking term [5]. However, a generalisation of the treatment using this model, with the goal to account for bubble formation, would require also to go beyond a modelling based on activated transition rate theory as used in this work.

[vi] In the following we will only give the sequence from 5′ to 3′ end, without further specifying the regular complementary part.

The freely jointed chain (FJC) model

The FJC model, with an extra term suggested in [150], yields for the force-dependent end-to-end distance $u_l(f)$

$$u_l(f) = l\left(1 + \frac{f}{Y}\right)\left[\coth\left(\frac{bf}{k_\mathrm{B}T}\right) - \frac{k_\mathrm{B}T}{bf}\right], \tag{2.11}$$

where Y denotes the Young modulus and b the Kuhn length. Typical values of the Kuhn length and the Young modulus are $b = 1.42\,\mathrm{nm}$ and $Y = 812\,\mathrm{pN}$ [150] or, as published recently, $b = 1.15\,\mathrm{nm}$ and $Y = \infty$ [70], respectively. We use the latter values. The contour length $l = l(n)$ of released ssDNA from n bps is

$$l = 2\,d\,n + \delta_{n,n_\mathrm{tot}}\,d\,n_\mathrm{loop}\,. \tag{2.12}$$

Here, $d = 0.59$ nm/base [38, 165, 71, 2] is the equilibrium distance between consecutive bases in the strand of the ssDNA and n_loop the number of bases per loop. For easier reading we do not explicitely indicate the dependence of the contour length on n.

The worm-like chain (WLC) model

In the WLC model [19, 108], the elastic force due to an elongation u_l is

$$f(u_l) = \frac{k_\mathrm{B}T}{4P}\left[\frac{1}{(1 - u_l/l)^2} - 1 + 4\frac{u_l}{l}\right], \tag{2.13}$$

and to obtain $u_l(f)$, this equation has to be inverted. The persistence length P, quantifying the stiffness of the polymer (cp. [46]), lies in a typical range of 1.0 to 1.5 nm for ssDNA [38]. In contrast, dsDNA has a persistence length of about 50 nm or 150 bps (1 bp = 0.34 nm) [150]. Compared to other biological polymers such as actin or microtubule molecules, where $P = 15\,\mu\mathrm{m}$ [168] and 6 mm [111], respectively, DNA is thus quite flexible. The alignment of polymer segments of $P = 50\,\mathrm{nm}$ at room temperature requires forces of the order $k_BT/P = 0.1\,\mathrm{pN}$. Thus, in the typical experimental force range of 5 to 50 pN the experimentally recorded data is nearly not influenced by such an alignment [20].

The stretching energy of the ssDNA

The strain energy of the unfolded single-stranded part [115] takes the decrease of entropy of the ssDNA into account and can be calculated from the work needed to stretch the unpaired bases. Since this work depends on the experimental protocol (cp. sec. 2.5), we have to treat both protocols separately. Because the length of the unfolded single-stranded part, $u_l(f)$, is monotonically increasing with force, it has a unique inverse $f_l(u) = u_l^{-1}(f)$, which is the force exerted by a ssDNA chain with contour length l, if its mean end-to-end distance is elongated by u. Both polymer models introduced above can be used to account for the elastic response of the

released ssDNA (cp. e. g. [70, 150] for the FJC and [102, 2] for the WLC model). Note that in general we chose to work with the FJC model (see eq. (2.11)) with the parameters of [70], despite the fact that both approaches give similar good results (cp. tab. 3.1 and sec. 3.3 for a detailed discussion, see also [70]).

Force protocol

In the FP, we can write for the stretching energy of the ssDNA [154]

$$G_{\mathrm{str}}^{\mathrm{ssDNA}}(n,f) = \int_0^{u_l} \mathrm{d}u'\, f_l(u') - f\, u_l(f) = -\int_0^{f} \mathrm{d}f'\, u_l(f')\,, \tag{2.14}$$

where we performed an integration by parts. The term $-f u_l(f)$ describes the energetical preference of unfolded states with larger force. Using the FJC model with a Young modulus $Y = \infty$ [70], we can further write [72]

$$G_{\mathrm{str}}^{\mathrm{ssDNA}}(n,f) = -\frac{l\, k_{\mathrm{B}}T}{b}\left[\ln\left(\sinh\left(\frac{bf}{k_{\mathrm{B}}T}\right)\right) - \ln\left(\frac{bf}{k_{\mathrm{B}}T}\right)\right]. \tag{2.15}$$

Distance protocol

One can easily obtain the stretching energy of the ssDNA for the DP out of eq. (2.14) with the help of a Legendre transformation [72, 173],

$$G_{\mathrm{str}}^{\mathrm{ssDNA}}(n,u_l) = f\, u_l(f) - \int_0^{f} \mathrm{d}f'\, u_l(f')\,, \tag{2.16}$$

where we can again replace the integral by eq. (2.15). Here we split the contributions of ssDNA and stem which are connected by $x = u_l(f) - \delta_{n,n_{\mathrm{tot}}}\, d_0$.

Stem orientation

Since the stem is aligned in force direction when the whole molecule unfolds, we also included the energetic cost of the stem orientation in the DP as [117]

$$G_{\mathrm{str}}^{\mathrm{stem}}(n,d_0) = -\delta_{n,n_{\mathrm{tot}}}\, f\, d_0\,. \tag{2.17}$$

We set $d_0 \simeq 2\,\mathrm{nm}$ in accordance with the diameter of the B-DNA helix (cp. sec. 2.1) [38, 165]. Note that in the FP, the stem contribution has a positive sign [117]. For an alternative way to model the stem contribution as a dipole see [51, 2].

The contribution of the bead in the optical trap

We include the energy contribution of the bead in the optical trap only in the DP [106, 117]. For small displacements, the optical trap creates a nearly harmonic potential of a one-dimensional spring whose rest position is at $y = 0$ [72, 117],

$$G_{\mathrm{str}}^{\mathrm{bead}}(y) = \frac{1}{2}\, k_t\, y^2\,, \tag{2.18}$$

where k_t is the trap stiffness. The restoring force, $f = -dG_{\mathrm{str}}^{\mathrm{bead}}(y)/dy = -k_t y$, can be determined by measuring the displacement of the bead from the centre of the trap focus.

The stretching energy of the handles

Also the stretching energy of the handles is included only in the DP [106]. In analogy to eq. (2.18), we assume it as a harmonic potential [117]

$$G_{\mathrm{str}}^{\mathrm{handles}}(h) = \frac{1}{2}\, k_h\, h^2\,, \tag{2.19}$$

where k_h is the trap stiffness. Alternative approaches to model the handles use the WLC model $G_{\mathrm{str}}^{\mathrm{handles}}(h) = \int_0^h dh'\, f(h')$ (see eq. (2.13) using h instead of $u_l(f)$) [105, 162, 72].

Summary of the stretching contributions

Summing up, for the FP one can thus write

$$G_{\mathrm{str}}(n, f) = G_{\mathrm{str}}^{\mathrm{ssDNA}}(n, f) + G_{\mathrm{str}}^{\mathrm{stem}}(n, d_0)\,. \tag{2.20}$$

Note that we did not include $G_{\mathrm{str}}^{\mathrm{stem}}(n, d_0)$ in our calculations in sec. 3.4. In the DP we need to consider

$$G_{\mathrm{str}}(n, Z) = G_{\mathrm{str}}^{\mathrm{ssDNA}}(n, u_l) + G_{\mathrm{str}}^{\mathrm{stem}}(n, d_0) + G_{\mathrm{str}}^{\mathrm{bead}}(y) + G_{\mathrm{str}}^{\mathrm{handles}}(h)\,. \tag{2.21}$$

Chapter 3

Pulling experiments of a triple-branch DNA molecule using a distance protocol

Parts of this chapter are published as:

> Sandra Engel, Anna Alemany, Nuria Forns, Philipp Maass, and Felix Ritort. Folding and unfolding of a triple-branch DNA molecule with four conformational states. Phil. Mag., 91(13):2049–2065, March 2011.

3.1 Description of the pulling experiments

Based on Mfold folding predictions [143, 175] and taking FEL considerations into account (see sec. 2.6), we designed and synthesised a DNA molecule which consists of three parts and is hence referred to as a triple-branch molecule [46]. It is composed of a stem as introduced in [115] with 21 bps and two nearly identical hairpin branches formed of 16 bps and an end-loop of four bases (GAAA), thus comprising a total number of 114 bases (see fig. 3.1). To avoid misfolding, the second hairpin branch differs at two positions from the first one. The triple-branch molecule is inserted between two identical short dsDNA handles of 29 bps [51], leading to a total number of 172 bases in the whole DNA molecule.

Three different kind of experimental protocols [101] have been developed with the objective of exploring the folding-unfolding transition in DNA molecules with OTs, namely hopping, force-jump and pulling[i] experiments (see also [46, 93]). In the following we will concentrate on pulling experiments. Our pulling experiments are carried out with a miniaturised dual-beam laser OTs apparatus [70] at room temperature ($\simeq$ 25 °C) and at a salt concentration of 1 M NaCl aqueous buffer with neutral pH (7.5) stabilised by Tris HCl and 1 M EDTA. The dual-beam OTs collect

[i] Also known as force-ramp experiments [93].

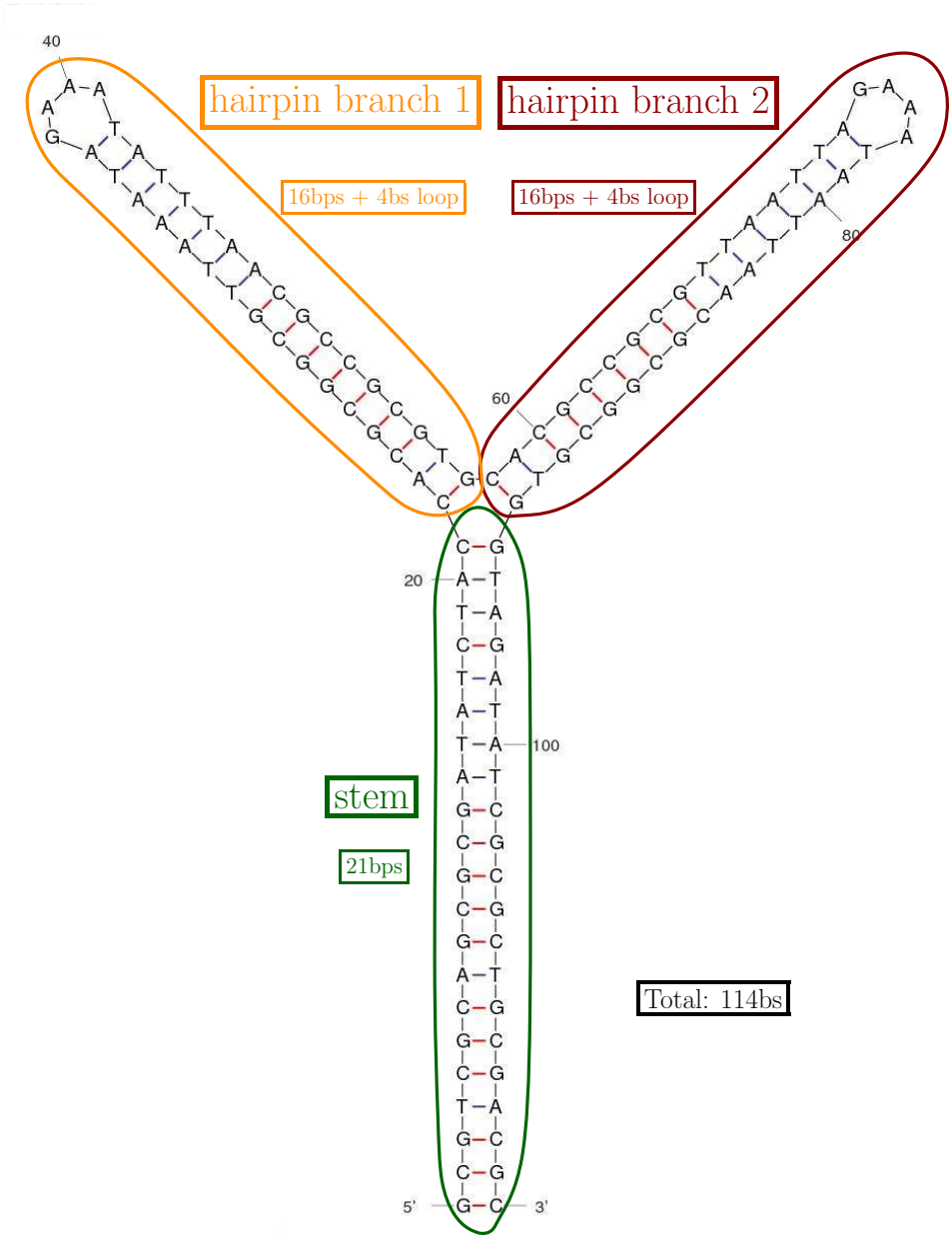


Figure 3.1: Structure of the triple-branch DNA molecule.

data at 4 kHz and can operate with a feedback rate of 1 kHz. Spatial resolution constitutes 0.5 nm with a maximal distance range of $\simeq$ 10 µm. Forces up to 100 pN can be achieved, whereas the force resolution is 0.05 pN.

Optical tweezer pulling experiments permit the measurement of the force f as well as the total distance Z^* between the centre of the optical trap and the tip of the micropipette (see fig. 2.3). As discussed in sec. 2.3, we will work with the relative trap-pipette distance Z in the following. In the experiment we vary the distance $Z(t)$ with a constant speed $v = dZ/dt$ in the range of 45 nm/s to 200 nm/s, which corresponds to a constant average loading rate r of 3.0 pN/s and 13.4 pN/s in between rip events, respectively. Pulling speed v and loading rate r are related via the effective stiffness k_{eff} of the molecular construct, which is extracted from the average slope of the FDCs,

$$r = k_{\text{eff}}\, v. \tag{3.1}$$

The experiment consists of loading cycles which in turn are divided into an unfolding part (during "pulling") and a folding part (during "pushing")[ii].

The loading cycles are repeated as long as the tether connection is unbroken. Otherwise a new connection has to be established, sometimes even a new molecule must be linked to a new bead (see sec. 3.5). In sec. 3.3 we will discuss different patterns which we found in the measured FDCs and analyse two representative molecules in a detailed way. We chose them among seven molecules featuring a first and among five molecules exhibiting a second pattern. Due to a logarithmic dependence of the first rupture force with velocity, the different pulling speeds, ranging from 45 to 200 nm/s, influence the experimental findings only weekly. The data was compatible for the two sets of similar molecules. In the theoretical analysis of the data in sec. 3.4, we will focus on a molecule recorded at a speed of 200 nm/s with 82 loading cycles. The second molecule investigated in detail comprises 55 cycles.

3.2 Analysis of the unfolding and folding trajectories

Based on the design of the triple-branch molecule, we have to distinguish between four conformational states (see fig. 3.2), corresponding to the four minima of the FEL (see sec. 3.4):

1. a completely folded molecule.
2. a completely unfolded stem, but both hairpin branches still folded.
3. a completely unfolded stem and either hairpin branch 1 or 2.
4. a completely unfolded molecule.

[ii] Also known as stretching and releasing process [115] or unzipping and rezipping [72].

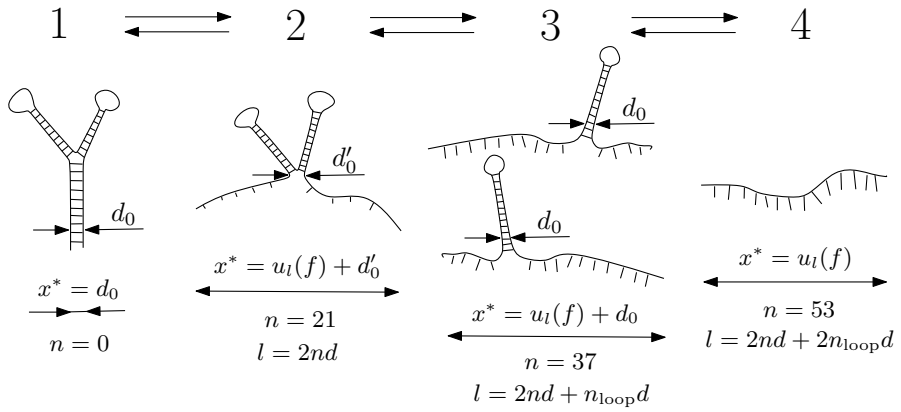


Figure 3.2: The four stable or metastable states of the triple-branch molecule: 1 - folded molecule, 2 - unfolded stem, 3 - stem and one hairpin branch unfolded, 4 - unfolded molecule. The values of x^* refer to the end-to-end distance given in eq. (3.3). The number of opened bps n corresponds to the single parts of the triple-branch molecule (cp. fig. 3.1) and the contour length l is calculated according to eq. (3.4).

Figure 3.3(a) depicts a typical unfolding curve and fig. 3.3(b) a typical unfolding and folding trajectory in the form of a FDC, recording the evolution of the force as a function of time and relative trap position Z (cp. sec. 2.5). With rising Z, the force f first increases almost linearly according to the elastic response of the handles. In our case these linkers consist of dsDNA and do not influence the FDC in the force range where the triple-branch molecule unfolds or folds, since the overstretching transition occurring typically at 65 pN [150] lies at much higher forces than we explore. At a first rupture force f_1 a sudden decrease, a so-called "jump", Δf_1 can be observed, that is caused by the unfolding of the stem. This unfolding goes along with an abrupt length change when ssDNA is released. As a consequence, the bead in the optical trap approaches the centre of the trap, leading to a suddenly smaller force value in the FDC. After this jump Δf_1, the curve increases again linearly up the next force rip at a first rupture force f_2. Here one of the hairpin branches unfolds, which in turn leads to the force jump Δf_2. Finally, the second hairpin branch unzips at a first rupture force f_3 with a jump Δf_3. The linear regime following this last force rip[iii] corresponds to the stretching of the whole molecular construct, including handles and the already unfolded triple-branch molecule.

Starting in the completely unfolded state, during the folding part of the cycle, when Z decreases, the force first follows closely the corresponding unfolding part of the curve. However, the first backward transition from state 4 to 3 does not occur

[iii] Some hairpin molecules repeatedly unfold and refold during one unfolding process (cf. [97, 115, 104, 72]). This so-called hopping depends on the height of the kinetic barrier, the applied force and the loading rate.

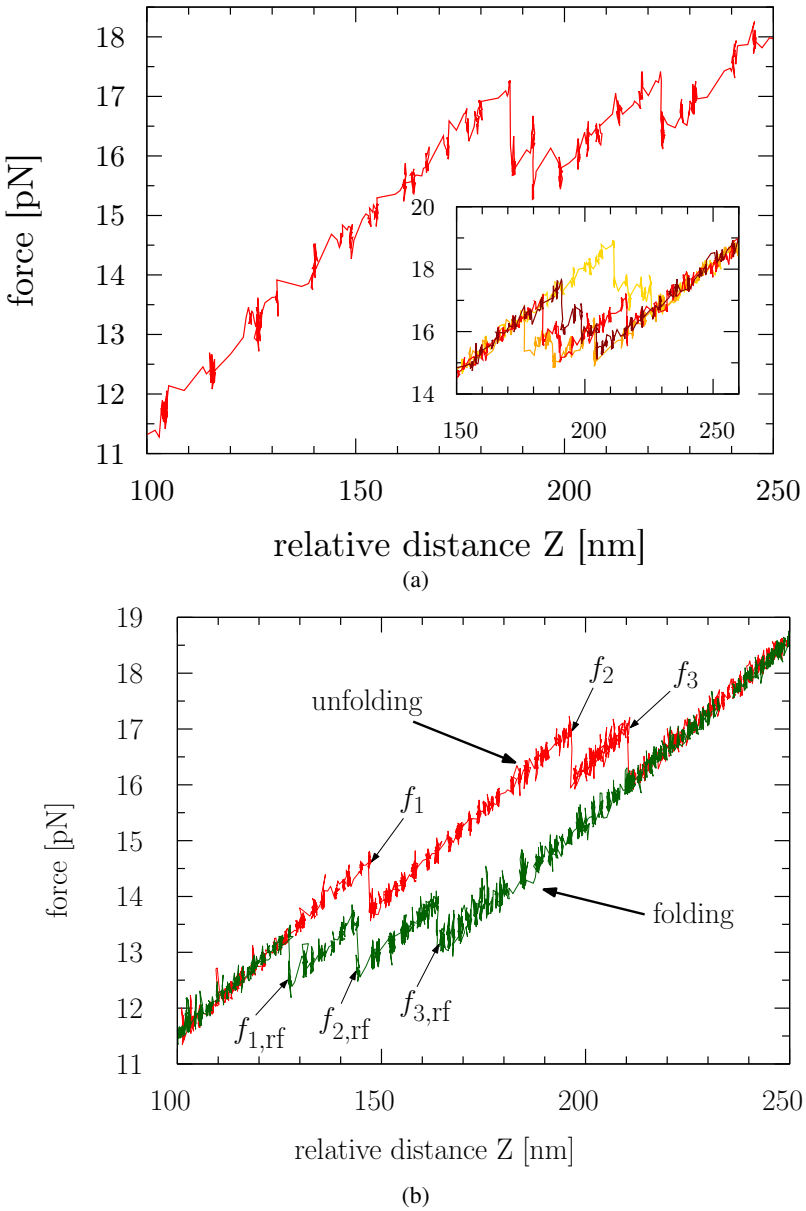
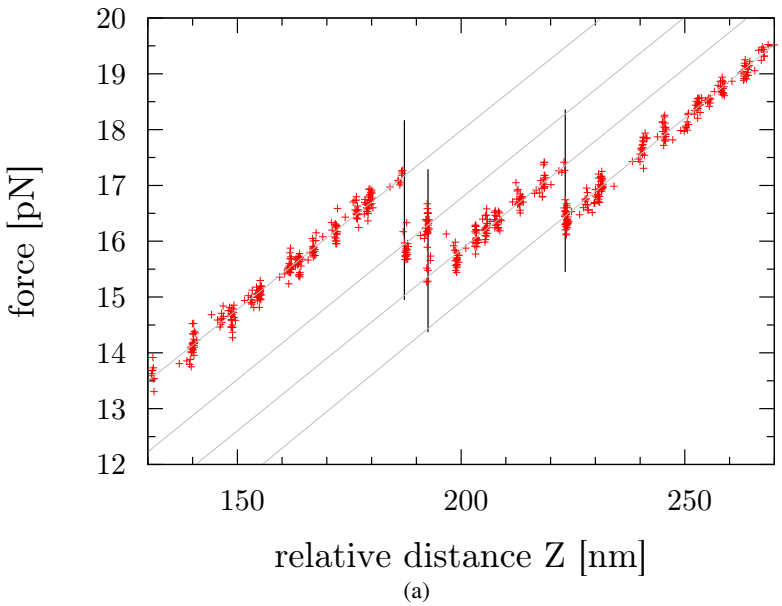


Figure 3.3: Force as a function of change in the trap-pipette distance (a) for a typical unfolding trajectory of the first investigated molecule (first pattern) and (b) for a typical unfolding and folding trajectory of the second investigated triple-branch molecule (second pattern). Indicated are the three first rupture forces belonging to the transitions 1 (f_1), 2 (f_2) and 3 (f_3) and the respective refolding forces, labelled $f_{i,\mathrm{rf}}$. The stochastic nature of the unzipping process is depicted in the inset in (a), which is showing four further unfolding curves for the first investigated molecule.

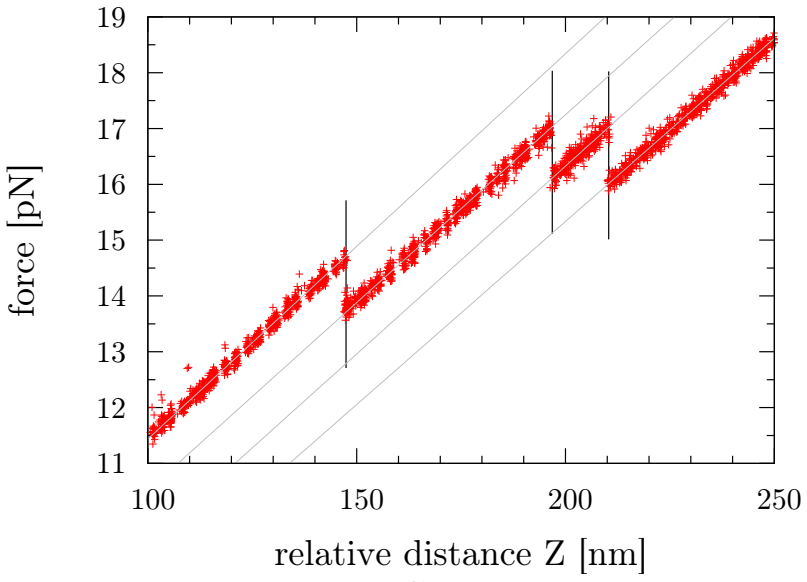
at $(f_3 - \Delta f_3) \simeq 16.2\,\mathrm{pN}$, but at a considerably lower value which is $(14.0 \pm 0.8)\,\mathrm{pN}$. This hysteresis effect [117] is clearly visible in fig. 3.3(b). Furthermore, investigating a large number of folding trajectories revealed that the bases of the hairpin branches do not always pair conjointly in well-defined events during a short time interval. In contrast to this, the respective force rips during unfolding indicate a cooperative behaviour of the biomolecule [116], since the breakage of all hydrogen bonds that stabilise the DNA occurs almost simultaneously. A second transition, after which both hairpin branches are unfolded, takes place at $(13.3 \pm 0.8)\,\mathrm{pN}$ instead of $(f_2 - \Delta f_2) \simeq 15.5\,\mathrm{pN}$. Eventually, one can identify a last sharp transition to the folded state 1 around $(11.2 \pm 0.8)\,\mathrm{pN}$, which is again considerably lower than $(f_1 - \Delta f_1) \simeq 13.7\,\mathrm{pN}$. The unfolding process is thus delayed and occurs at higher forces, whereas folding takes place at lower forces. This hysterese effect is more pronounced the higher the pulling speed and the longer the molecule and thus the higher the kinetic barrier [117].

During unfolding, one can assume that the breakage of hydrogen bonds follows the sequence of the triple-branch molecule. Therefore it is useful to calculate the FEL depending on the number of broken bonds n, which is used as a state variable (see sec. 2.6 and 3.4). Typically, the folding exhibits less sharp transitions (see fig. 3.3(b)), since in particular at the beginning of the folding process a huge number of secondary structures are accessible. This implies that the kinetic pathways are less predefined and accordingly, the transitions get smeared out. Regarding a theoretical treatment, a description in terms of a simple state variable n becomes unlikely to be sufficient. Therefore, in a refined analysis, many more configurations should be included as relevant states in a coarse-grained description [103]. However, this goes beyond the scope of this work and we thus will concentrate on the unfolding process in the following.

There are plenty of possible ways to analyse FDCs in order to extract first rupture forces and force jumps, for instance using a Bayesian approach [72]. In our procedure we arranged the normalised data, i. e. the relative trap-pipette distance Z and the force f, in windows of a certain size of data points. Next, we calculated for all consecutive windows the slope of the considered data points, the span, i. e. the maximum distance between the lowest and the highest force value, and the mean of the relative distance $\overline{Z_j}$ as a moving average. We assumed that transitions between the conformational states take place when the slope is minimal and the span maximal under the condition that a certain number of contiguous windows is connected. After the values of $\overline{Z_j}$ of the three force rips have been recovered, slope and axis intercept are calculated by linear regression for each conformational state. Now one can easily compute the first rupture forces as the intersection points with the four fitted lines and extract the force jump values. This procedure is exemplified in fig. 3.4 for both molecules, whose unfolding trajectories are the same as in figs. 3.3(a) and 3.3(b), respectively. Since the data acquisition rate is constant, at higher pulling speeds fewer data points are collected. Therefore the extracted values of the first molecule are more broadly distributed. The three first rupture forces f_i as well as the force jumps Δf_i are subject to stochastic fluctua-



(a)



(b)

Figure 3.4: The first rupture forces and the force jump values of the unfolding trajectories displayed in fig. 3.3 are extracted as indicated here from the intersection of the fitted grey lines and the transition points (vertical lines). The procedure is exemplified for the molecule used in the theoretical analysis in sec. 3.4 in (a) and for the molecule with permanent fraying behaviour in (b).

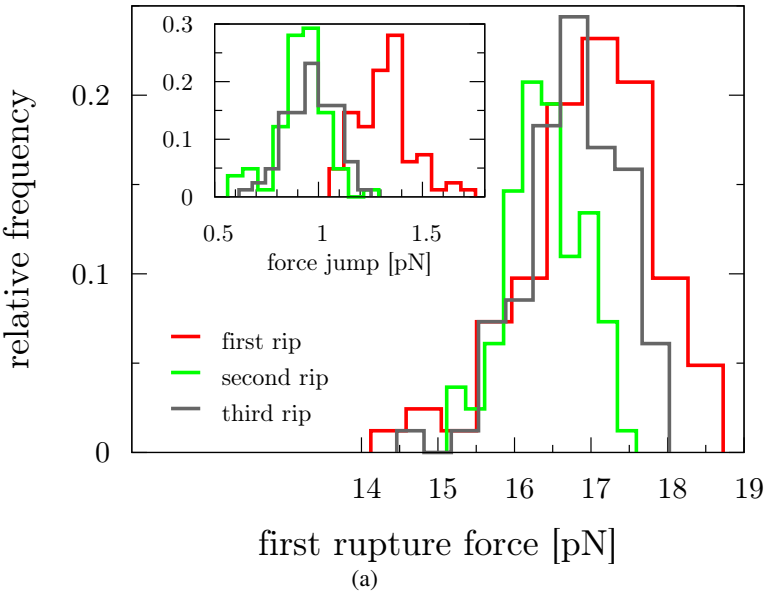
tions (cp. fig. 3.3(a)), where the inset depicts four unfolding trajectories belonging to different pulling cycles of the same molecule. An analysis revealed that the fluctuations of Δf_i are approximately ten times smaller than the fluctuations of f_i. We will therefore in sec. 3.4 neglect the fluctuations in the force jumps Δf_i and only use their averages $\overline{\Delta f_i}$, that we discuss in more detail in the following section.

3.3 Different unfolding patterns related to the number of opened bps

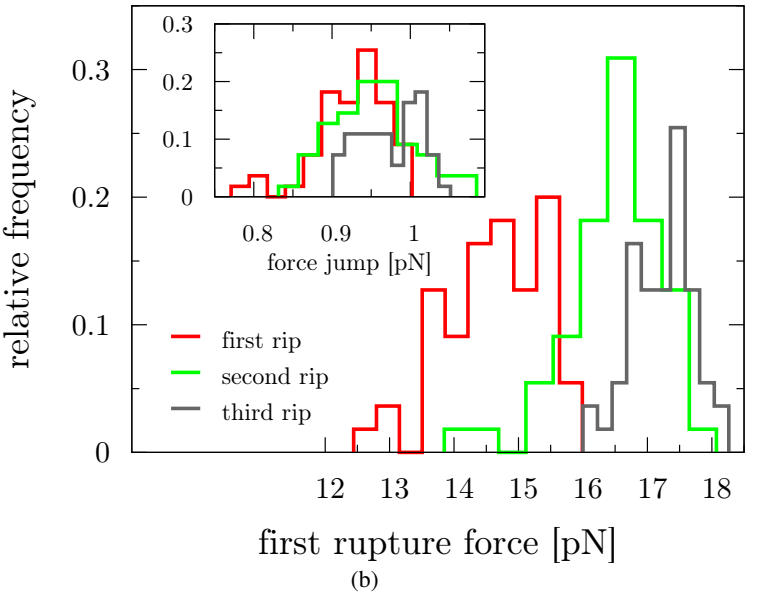
Applying the above introduced procedure (see sec. 3.2) to analyse the experimental data of several molecules, we discerned two predominating patterns in the unfolding trajectories. These patterns are reflected in the distributions of the first rupture forces f_i as follows. Figure 3.5(a) depicts the first pattern, where the distributions have a similar shape for all three force rips. The histograms indicate the existence of a maximum slightly below 17 pN. The average values for the three first rupture forces are $\overline{f_1} = (17.0 \pm 0.9)$ pN, $\overline{f_2} = (16.4 \pm 0.6)$ pN and $\overline{f_3} = (16.8 \pm 0.7)$ pN. In contrast, fig. 3.5(b) displays a second pattern, where the first rupture force of the first rip $\overline{f_1} = (14.6 \pm 0.8)$ pN is strikingly lower than in the first pattern. The remaining two rip forces lie mainly in the same range, i. e. between 16 and 18 pN. As in the first pattern, the second rip tends to have a slightly smaller first rupture force, $\overline{f_2} = (16.5 \pm 0.8)$ pN, than the third rip, $\overline{f_3} = (17.2 \pm 0.5)$ pN.

The small $\overline{f_1}$ of the second molecule suggests the occurrence of permanent molecular fraying. Obviously less force is needed to unfold the stem than typically (cp. fig. 3.3(a) with 3.3(b)), because some bps at its basis are partly or completely melted. In other words, during folding this molecule does not reach an entirely folded state, but a few bps of the stem next to the handles remain irreversibly and permanently open. This phenomenon has been observed previously in several pulling experiments [165, 1, 117, 70, 51]. A possible reason for irreversible fraying is the formation of reactive oxidative species due to the impact of the laser light of the optical trap, which leads to a degradation of the DNA bases [89]. In this way singlet oxygens are generated, which are known to oxidise certain nucleic acids, such as guanine and thymine, irreversibly. This could explain why once a molecule showed fraying in our experimental data it did not exhibit normal behaviour in later pulling curves. We work with polystyrene microspheres which are more prone to photodamage than the DNA bases themselves. Their wide ranging interaction with the bases might be reduced by replacing polystyrene by silica beads, which exhibit considerably minor irreversible oxidative damage. Further similar experiments with silica beads might reduce the amount of frayed molecules.

The corresponding force jump distributions of both investigated molecules are depicted in the insets in figs. 3.5(a) and 3.5(b). The first molecule has a large first force jump $\overline{\Delta f_1} = (1.3 \pm 0.2)$ pN and two smaller force jumps at the second and third rip, $\overline{\Delta f_2} = (0.9 \pm 0.2)$ pN and $\overline{\Delta f_3} = (1.0 \pm 0.2)$ pN. In contrast, the frayed molecule exhibits three force jumps of approximately the same value,



(a)



(b)

Figure 3.5: Histograms of the three first rupture forces during unfolding for two representative triple-branch molecules: (a) the molecule used in the theoretical analysis in sec. 3.4 and (b) the molecule with permanent fraying behaviour. Note that the average force value of the first rip in (b) is smaller than in (a). The insets display the respective histograms for the force jumps.

i. e. $\overline{\Delta f_1} = (0.93 \pm 0.05)\,\mathrm{pN}$, $\overline{\Delta f_2} = (0.95 \pm 0.06)\,\mathrm{pN}$ and $\overline{\Delta f_3} = (0.97 \pm 0.04)\,\mathrm{pN}$, respectively. This illustrates clearly the influence of irreversible fraying in the latter case since, roughly estimated, the same number of bps is expected to open in all three rips. This should not be the case in a completely folded molecule (cp. sec. 3.4).

As depicted in fig. 3.4, during a force rip, the relative trap-pipette distance Z is constant and thus $\Delta Z = 0$. Therefore the change in the relative molecular extension,

$$\Delta m = \Delta Z - \Delta f / k_\mathrm{t} = \Delta x\,, \tag{3.2}$$

is only related to the force jump Δf and the trap stiffness $k_\mathrm{t} \simeq 0.08\,\mathrm{pN/nm}$ for our experiments (cp. sec. 2.3). For a certain force and elastic model for the released ssDNA (cp. sec. 2.6.2), the number of opened bps n is related univocally to the equilibrium end-to-end distance of the investigated DNA molecule x^*. Its change $\Delta x^* = \Delta x$ in turn equals Δm. By means of this relation, it is possible to estimate the change in the number of bps Δn_i which are opened sequentially during each force rip. Focussing on the configurations of the four conformational states of the triple-branch molecule, x^* can be decomposed into two parts (see fig. 3.2). The first part is the elongation $u_l(f)$ of the mean end-to-end distance of the ssDNA along the force direction. It accounts for the ideal elastic response of the ssDNA with a contour length l. The second part contains the contribution of the diameter of stem and hairpin branches, respectively. We thus can write

$$x^* = u_l(f) + \begin{cases} 0, & n = 21 + 16 + 16\,\mathrm{bps} \\ d_0, & n = 0 \text{ and } 21 + 16\,\mathrm{bps} \\ d_0', & n = 21\,\mathrm{bps} \end{cases}, \tag{3.3}$$

where $d_0 \simeq 2\,\mathrm{nm}$, in accordance with the diameter of the B-DNA helix (cp. sec. 2.1) [38, 165]. The exact value of the diameter contribution d_0' of both hairpin branches, when the stem is unfolded, depends on the orientation of the branches. We set $d_0' = 2d_0$ as a working value[iv].

The contour length l (cp. sec. 2.6.2), depending on the number of opened bps n, can be written for the configurations of the four conformational states as

$$l = 2nd + \begin{cases} 0, & n = 0 \text{ and } 21\,\mathrm{bps} \\ n_\mathrm{loop} d, & n = 21 + 16\,\mathrm{bps} \\ 2n_\mathrm{loop} d, & n = 21 + 16 + 16\,\mathrm{bps} \end{cases}. \tag{3.4}$$

The interphosphate distance d is 0.59 nm/base [38, 165, 71, 2] and $n_\mathrm{loop} = 4$ is the number of bases in the end-loop. For every force rip f_i the corresponding number of opened bps n_i is calculated separately, taking into account the differences of the contour length l between the conformational states.

[iv] Due to this simplification, Δn_1 of the first rip is likely to be slightly underestimated and Δn_2 of the second rip overestimated. However, this will not affect the change in the total number of opened bps Δn_tot because we consider the change of x^*. Therefore the contributions d_0' will cancel each other out.

		molecule 1				molecule 2			
model	parameter	change in no. of opened bps [bps]							
	b & P [nm], Y [pN]	Δn_1	Δn_2	Δn_3	Δn_{tot}	Δn_1	Δn_2	Δn_3	Δn_{tot}
FJC	b = 1.42, Y = 812 [150]	18 (2)	14 (2)	15 (2)	46 (3)	13 (1)	15 (1)	15 (1)	42 (2)
	b = 1.15, $Y = \infty$ [70]	19 (3)	15 (2)	16 (2)	50 (3)	14 (1)	16 (1)	16 (1)	46 (2)
WLC	P = 1.0	20 (3)	16 (3)	17 (3)	54 (3)	14 (1)	17 (1)	17 (1)	49 (2)
	P = 1.3 [142]	20 (3)	15 (2)	16 (2)	51 (3)	14 (1)	16 (1)	16 (1)	46 (2)
	P = 1.5	19 (3)	15 (2)	16 (2)	50 (3)	13 (1)	16 (1)	16 (1)	45 (2)
expected values		21	16	16	53	21	16	16	53

Table 3.1: Overview of the change in the number of opened bps using different polymer models and parameters for the two representative molecules. (The numbers in brackets are the standard deviations.)

Depending on the respective model and parameters, the results for the estimated change in the number of opened bps lie in an acceptable range, when the errors are taken into account (see tab. 3.1). It is not evident which model and parameters can be considered as the best ones. Good results are found for the FJC model using recent values of [70] (b = 1.15 nm and $Y = \infty$) and for the WLC model with P = 1.3 nm [142] (cp. sec. 2.6.2). In the following, we chose to work with the FJC model with the parameters of [70].

As expected from fig. 3.5(b), the second molecule indeed reveals a considerably smaller Δn_1 of 13 or 14 bps, depending on the respective model. This means that 7 or 8 bps are not closed when the folding process is completed. Performing SMEs without knowing the exact influence of permanently frayed bps can thus lead to fairly misinterpreted results. One can estimate the number of irreversibly frayed bps at the basis of the stem by checking the appropriate parameters for the polymer models by dint of the change in the number of opened bps for the hairpin branches, since no fraying occurs in the branches. This method can be in principle also applied to non-permanent, reversible molecular fraying.

3.4 Theory for the kinetics of the unfolding process

On a coarse-grained level the kinetics of the unfolding process can be described based on the Gibbs free energy $G(n, f)$, which is a function of the number of sequentially opened bps n and the applied force f (see sec. 2.6). This FEL has a shape as sketched in fig. 3.6(a) for small forces, including $f = 0$. At zero force, the FEL only consists of $G_{\text{form}}(n)$ and increases monotonically with n (see sec. 2.6.1). In case of the triple-branch molecule local minima occur at the metastable states 2, 3 and 4. This is related to the increase of entropy due to the release of additional degrees of freedom when the stem-hairpin-junction and end-loops of the hairpin branches are unfolded. The FEL is tilted with rising force and the energies of the metastable states are lowered (see sec. 2.6.2). Once the stem is completely unfolded, several structures compatible with a given n can occur, refering to dif-

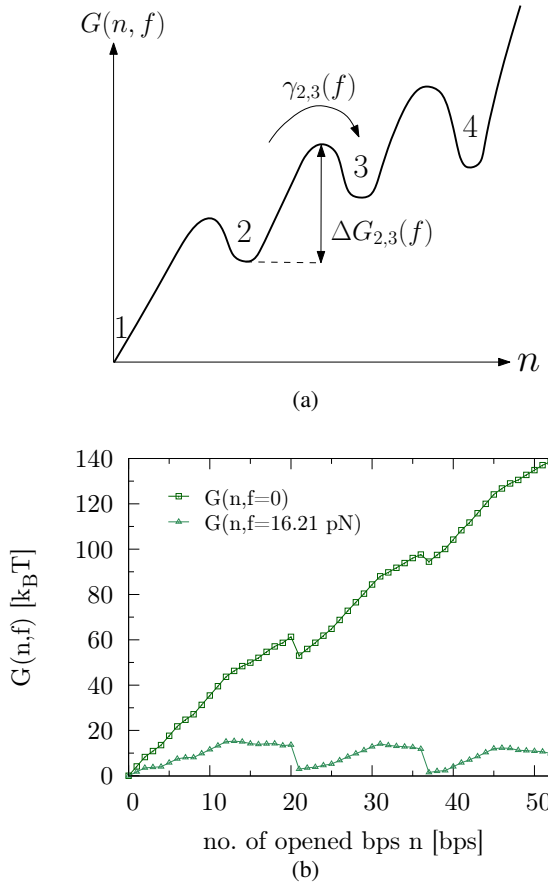


Figure 3.6: (a) Sketch of the expected FEL $G(n, f)$ for the triple-branch molecule at zero force as a function of the number of opened bps n. (b) $G(n, f)$ calculated using eqs. (2.9), (2.10), (2.20) and (3.5) at two forces, given in units $k_\mathrm{B}T$.

ferent possibilities of breaking the bps in the two hairpin branches. Therefore we introduce a "degeneration" parameter α, leading to $G(n, \alpha, f)$. Taking the different cases of partly unfolded branches into account, this FEL is calculated analogously to eq. (2.9). For the calculation of the stretching contributions we use eq. (2.20), which models our experiments in a satisfactory manner (see discussion in sec. 5.1). At a given n, we must calculate the (restricted) partition sum over the configurations α to get $G(n, f)$ using

$$G(n, f) = -k_\mathrm{B}T \ln \sum_{\alpha} \exp\left(-\frac{G(n, \alpha, f)}{k_\mathrm{B}T}\right). \tag{3.5}$$

Figure 3.6(b) depicts the FEL $G(n, f)$ for $f = 0$ and $f = 16.21$ pN. As anticipated in fig. 3.6(a), at zero force it has minima at the stable/metastable states and gets tilted with rising force. The levels of the minima become comparable when ap-

proaching the force regime where rips occur (cp. fig. 3.3(a)). The DNA sequence given in fig. 3.1 has been designed deliberately to yield the multiple-state structure in terms of the FEL (see fig. 3.6(b)). This gives us some confidence in the model underlying the construction of the FEL.

Based on the FEL, we can apply standard transition rate theory [8] and write for the transition rate from state i to $i+1$

$$\Gamma_{i,i+1}(f) = \gamma_i^0 \, \gamma_{i,i+1}(f) \,. \tag{3.6}$$

γ_i^0 is an attempt rate and $\gamma_{i,i+1}(f)$ the Boltzmann factor corresponding to the activation barrier $\Delta G_{i,i+1}(f)$ that has to be surmounted,

$$\gamma_{i,i+1}(f) = \exp\left(-\frac{\Delta G_{i,i+1}(f)}{k_\mathrm{B}T}\right) . \tag{3.7}$$

As indicated in fig. 3.6(a), the activation energy $\Delta G_{i,i+1}(f)$ is calculated from the FEL by determining the energy $G_i^{\min}(f)$ of the local minimum belonging to state i and the saddle point energy $G_{i,i+1}^{\mathrm{saddle}}(f)$ of the ith transition between state i and $(i+1)$, that is to say $\Delta G_{i,i+1}(f) = G_{i,i+1}^{\mathrm{saddle}}(f) - G_i^{\min}(f)$. Further, we can compute the transition probability $W(f_i|f_{i-1})$ for the first rupture force f_i of the ith transition provided that the first rupture force was f_{i-1} in the $(i-1)$th transition (see [46] for details). The result is

$$W(f_i|f_{i-1}) = \frac{\gamma_i^0}{r} \, \gamma_{i,i+1}(f) \, \exp\Big[-\frac{\gamma_i^0}{r} \int_{f_{i-1}^*}^{f} df' \, \gamma_{i,i+1}(f')\Big], \tag{3.8}$$

where $f_i^* = f_i - \overline{\Delta f_i}$ and $f_0 = f_0^* = 0$. In addition, we use the loading rate r (see sec. 3.1) and γ_i^0 and $\gamma_{i,i+1}(f)$ as defined in eqs. (3.6) and (3.7). The attempt rate γ_i^0 was used as the only fitting parameter. The best fit (see fig. 3.7) was obtained for the attempt rates $\gamma_1^0 = 9 \cdot 10^6$ Hz, $\gamma_2^0 = 1.1 \cdot 10^6$ Hz and $\gamma_3^0 = 3.5 \cdot 10^5$ Hz. A general discussion of the attempt frequency can be found in sec. 5.2.3 and ch. 6. Hence, we obtain for the joint probability density of the three first rupture forces

$$\Psi_3(f_1, f_2, f_3) = W(f_1|0) \, W(f_2|f_1) \, W(f_3|f_2) \,. \tag{3.9}$$

Figure 3.7 displays the histograms for the three rips given in fig. 3.5(a) in comparison with the distributions calculated from our theory. In view of the available statistics (82 cycles, see sec. 3.1), the agreement is quite satisfactory.

3.5 Comparison of three samples of the same molecule

Let us have a closer look at further data of molecules of pattern 1, that is without permanent molecular fraying, and which effects may occur during a pulling experiment. After having performed a certain number of pulling cycles, either the connection between one bead and the attached DNA breaks or the bead is no longer

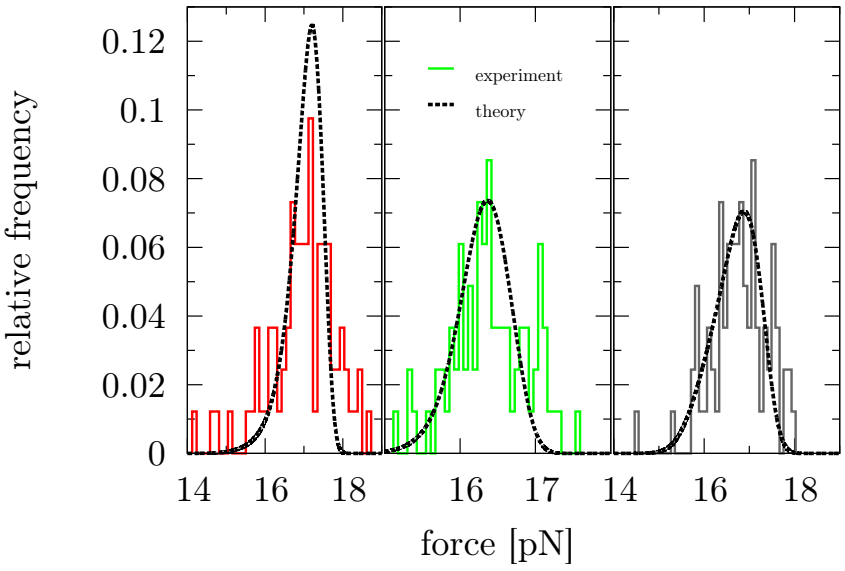


Figure 3.7: Comparison of the first rupture force distributions from fig. 3.5(a) with the theory for the first (f_1), the second (f_2) and the third (f_3) transition (see left, middle and right panel). The best fit was obtained for the attempt rates $\gamma_1^0 = 9 \cdot 10^6$ Hz, $\gamma_2^0 = 1.1 \cdot 10^6$ Hz and $\gamma_3^0 = 3.5 \cdot 10^5$ Hz.

captured in the trap (cf. sec. 2.3). Then one has to reestablish the broken connection or to capture the bead again. Sometimes it is even necessary to restart the whole procedure of tethering using a new DNA molecule and new beads [72]. All these different effects happened while we recorded data in our pulling experiments so that we got several individual samples of the same pulling speed. However, it was possible to pull one molecule at three different pulling speeds before any interrupting effect occurred. It is thus revealing to have a look at this molecule where three samples are available to exclude systematic deviations based on a slightly different molecular setup. Such an analysis can give information about the intrinsic deviations from the expected behaviour for one and the same molecule.

The three samples of the same molecule were recorded at pulling speeds of 97.0, 137.9 and 166.0 nm/s. This corresponds to loading rates of 5.7, 8.2 and 9.9 pN/s. For unfolding we detected 104, 129 and 114 trajectories and for folding 76, 113 and 122. Note that the number of trajectories for unfolding and folding is not the same concerning one sample due to several effects. The most prominent one is that we classified the FDC into several transition types, according to which part(s) of the molecule open at once during one rip. Here we focus on the analysis of three separate rips corresponding to the opening or closing of the three parts of the molecule, i. e. the stem and the two hairpin branches. We also eliminated curves with hopping, e. g. when the already opened stem of the molecule refolds

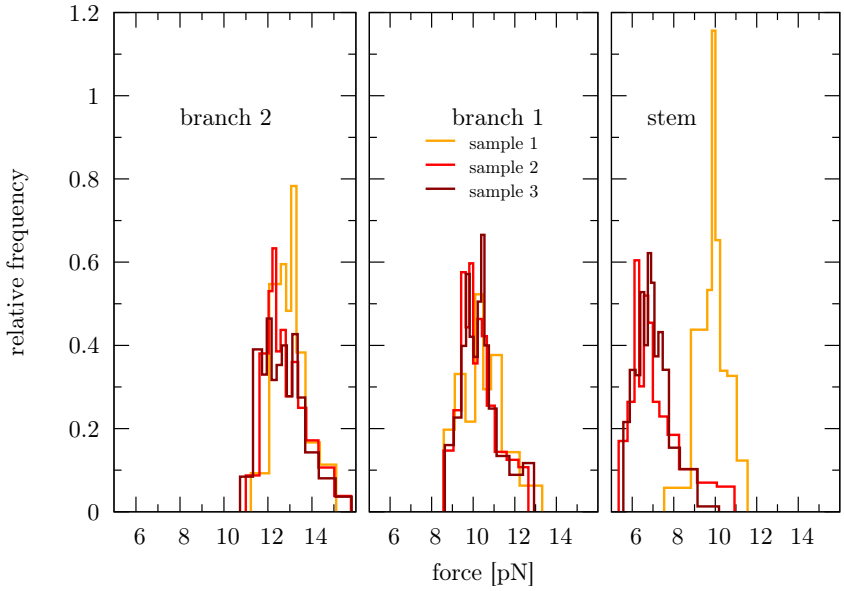


Figure 3.8: Folding force distributions for three samples of the same molecule for hairpin branch 2, hairpin branch 1 and stem (see left, middle and right panel).

and unfolds again (cf. sec. 3.2). Sometimes a molecule does not fold entirely before the unfolding cycle is executed anew (non-permanent, reversible molecular fraying (see sect. 3.3)). Those curves have also been removed from analysis. Molecules exhibiting permanent molecular fraying (cp. sect. 3.3) were excluded from the very beginning.

The unfolding distributions of the first rupture forces overlap well and their mean lies roughly at the same value. Only the distribution of the third rip, i. e. when the second hairpin branch unfolds, of the molecule which was pulled fastest lies at sligthly higher forces. The force jump distributions for unfolding match well for each rip. As expected, the stem exhibits a larger force jump than the hairpin branches (for details cp. sec. 3.3). The force jump distribution of the third rip is shifted to a little higher values compared to the second rip.

The folding distributions of the three rips also overlap very well for both first folding force and force jumps (see left and middle panel of fig. 3.8 for the first folding forces of the two hairpin branches). There is only one exception, namely when the stem folds (see right panel of fig. 3.8). Here the first folding force distribution of the sample which was pulled slowest lies at roughly the same values as the hairpin branch which folded directly before the stem. Only the remaining two samples show a clear tendency, i. e. a reduction of the mean and a shift of the whole distribution by approximately 3 pN for every new rip, i. e. closing of bps during the folding process. The deviation of the slowest sample could possibly arise from two different ways of how the stem folds: either directly after the second branch has

folded (slowest sample) or with some rather large difference in time and thus force evolution (two faster samples). The force jump distributions of all three samples and all three rips lie in the same range, even though we expect the stem (third rip) to have a higher value (see sec. 3.2 and 3.3). Apart from this, there are no major deviations to be seen, i. e. by repeating the same kind of pulling experiment several times with the same molecule and the same connection to the bead we expect to get a quite similar distribution of first rupture forces and force jumps for both unfolding and folding part of the pulling cycle.

Chapter 4

Unfolding kinetics of periodic DNA hairpins using a force protocol

Parts of this chapter are published as:

> Sandra Nostheide, Viktor Holubec, Petr Chvosta, and Philipp Maass. Unfolding kinetics of periodic DNA hairpins. J. Phys.: Condens. Matter, 26:205102, 2014.

4.1 Free energy landscapes in the force protocol

In this chapter we consider the FEL in the FP, i. e. where the force is the control parameter (cp. sec. 2.6). At zero force, the free energy $G(n,0) = G_{\mathrm{form}}(n)$ (cp. eq. (2.10)) of long unbranched hairpin molecules without any additional structures such as, for instance, interior loops, increases monotonically with n. The local rise $\Delta G_{\mathrm{form}}(n) = G_{\mathrm{form}}(n+1) - G_{\mathrm{form}}(n)$ varies according to the respective sequence. Considering unzipping experiments with a constant loading rate r, where the force increases linearly with time,

$$f(t) = f_0 + rt\,, \tag{4.1}$$

the FEL gets tilted and local minima ("states") form at values of n, where $\Delta G_{\mathrm{form}}(n)$ changes considerably, that is where $\Delta G_{\mathrm{form}}(n-1)$ and $\Delta G_{\mathrm{form}}(n)$ are quite different. After a certain amount of tilting, the values of the free energy minima become similar compared to the energy of the completely folded state. Then thermally activated transitions from one minimum to the next one drive the molecule into the unfolded state. In general, each of these transitions is characterised by different forward and backward barriers, that in addition evolve differently in time. Therefore, theoretical treatments become more and more difficult when the number of states increases.

Since we pursue the goal to make a theoretical treatment of the kinetics tractable, we are searching periodic sequences where the forward and backward barriers decrease linearly with force and, in addition, the energy levels of the states decrease linearly with fn_α. Here we denote the respective value of n associated with the state α with n_α. In particular, we write for the forward barriers $\Delta(f)$, the backward barriers $\Delta'(f)$ and the state free energies $G_\alpha(f) \equiv G_{n_\alpha}(f)$:

$$\Delta(f) = \Delta_0 - \Delta_1 f\,, \tag{4.2}$$

$$\Delta'(f) = \Delta'_0 - \Delta'_1 f\,, \tag{4.3}$$

$$G_\alpha(f) = \alpha(g_0 - g_1 f)\,. \tag{4.4}$$

Since $G_{\alpha+1}(f) - G_\alpha(f) = \Delta(f) - \Delta'(f)$, it holds $g_0 = \Delta_0 - \Delta'_0$ and $g_1 = \Delta_1 - \Delta'_1$. The FEL is thus specified by four independent parameters. It is sufficient that eqs. (4.2)-(4.4) are obeyed in the range where the energy levels begin to be comparable to the minimum of the folded state, so that transitions can start to occur. Is it possible to find periodic DNA sequences satisfying eqs. (4.2)-(4.4) on the basis of eq. (2.9) using eqs. (2.10) and (2.20)? This will be the topic of the next section.

4.2 Free energy landscapes of periodic DNA sequences

First of all, we analyse unbranched DNA hairpin molecules with periodic bp sequences, but without interior loops. If the periodicity length L of the units is small enough, eqs. (4.2)-(4.4) can be fulfilled in a large force range. However, the minima are not very pronounced and the barriers constitute only a few multiples of the thermal energy $k_\mathrm{B}T$. FEDs of these sequences should therefore be measurable from equilibrium occupancy statistics, too [97, 117, 104, 93, 2, 136]. Considering interior loops attached to the basic units in a second step, that typically lead to high barriers between the minima, the requirements imposed by eqs. (4.2)-(4.4) can be well fulfilled. Here we need to use additional parameters specifying the change of the FEL by the interior loops [143, 175]. However, they are less accurate than those defining the bp interactions [70] (see tab. 2.2 in sec. 2.6.1).

4.2.1 Hairpin molecules without interior loops

Figure 4.1 sketches a dsDNA with periodicity length $L = 5$ bps and an arbitrarily chosen sequence in the period. In principle, such hairpin molecules can be branched or form other structural elements (cf. sec. 2.1). To exclude such motifs we rely on folding predictions of the program Mfold [143, 175]. FELs of unbranched hairpin molecules generally do not satisfy the requirements of eqs. (4.2)-(4.4). When force is increased, minima and barriers form regularly in the FEL (see sec. 2.6.1). Beyond a certain force value, however, new minima and barriers can form in-between. The overall structure of the regular FEL is thus not stable, but one regular type goes over to another one. The "new barrier effect" leads to force intervals where eqs. (4.2)-(4.4) are satisfied, but jumps in $\Delta(f)$, $\Delta'(f)$ and $G_\alpha(f)$

arise when going from one interval to the next one. A second way how the regular FEL is modified is connected to a shift of the positions of minima and saddle points. This "shift effect" leads to force intervals where eqs. (4.2)-(4.4) are satisfied, but jumps in the slopes Δ_1, Δ_1' and g_1 emerge between the intervals. It is difficult to avoid the "new barrier effect" for long periodicity lengths, even when sequences are specifically optimised in order to circumvent this problem. For sufficiently short periodicity lengths ($L \lesssim 7$), the "new barrier effect" is usually not appearing. Certainly, the "shift effect" affects the regular behaviour of the FEL also for small periodicity lengths. Sequences have to be optimised in order to avoid it.

Before we focus on optimised sequences, let us have a short look onto random sequences with different periodicity length L. Figure 4.2 comparatively depicts the FEL $G(n, f)$ as a function of the number n of opened bps: for $L = 5$ bps the sequence reads AATCG, for $L = 10$ bps AATCG AAGGT and for $L = 15$ bps AATCG AAGGT CGTAA. Depending on the applied force, the FEL comprises up to one (two or three) barrier(s) per $L = 5$ (10 or 15) bps section. At 17 pN there is, on average, one barrier every 5 bps. Since the structure is repetitive, the barrier height remains constant, i. e. for $L = 5$ bps exists all in all one forward barrier height and for $L = 10$ (15) bps there are up to two (three) different, alternating barrier heights observable. These forward barrier heights $\Delta_i(f)$ (see fig. 4.3) are linearly decreasing with force, with jumps occurring whenever the force is high enough to form a second (or third) barrier for $L = 10$ (15) bps due to the "new barrier effect". For $L = 5$ bps, thus, none of these jumps is present. We also give the fraction of number of barriers per number of bps in the stem. Besides, the slope of the forward barrier heights $\Delta_i(f)$ is changing for all three depicted sequences due to the "shift effect".

An optimisation among all 4^5 sequences for $L = 5$ bps with regard to eqs. (4.2)-(4.4) yields the smallest regression coefficients for the sequence consisting of TGCAA (see fig. 4.5(b)) which we will refer to as sequence I. The respective intercept and slope parameters of the linear functions for sequence I are given in tab. 4.1. The associated FEL is shown in fig. 4.5(a) for three force values: $f = 14$ pN (minima have not yet formed), $f = 16$ pN (minima have formed, the energy levels increase with n) and $f = 17.7$ pN (equal energy level of the minima). In the inset

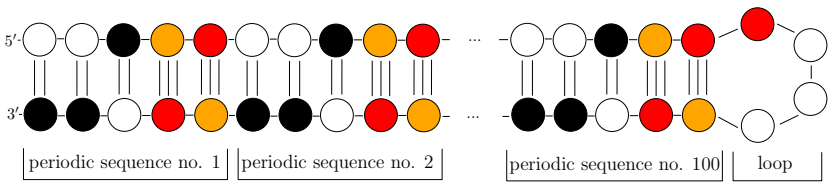


Figure 4.1: Sketch of a periodic, random DNA hairpin structure with a periodicity length of $L = 5$ bps. The bases are coloured as follows: adenine (white), cytosine (orange), guanine (red) and thymine (black).

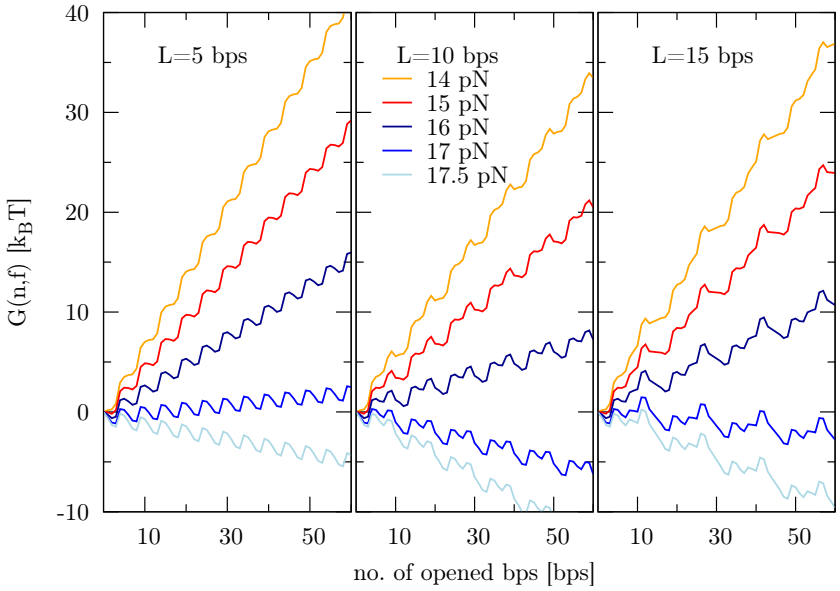


Figure 4.2: Comparison of the FEL of three random DNA sequences with a periodicity length of $L = 5$ bps (left panel), $L = 10$ bps (middle panel) and $L = 15$ bps (right panel) for several magnitudes of force.

two barriers zoomed out from the FEL for $f = 16$ pN are displayed. Since these barriers are merely 2.8 $k_B T$ high, they can be easily overcome at room temperature and one can apply equilibrium methods.

4.2.2 Hairpin molecules with interior loops

For longer basic units, eqs. (4.2)-(4.4) can be satisfied by attaching interior loops to each unit. In such a case even equal type of bases, for instance AAA... paired with TTT..., in the double-stranded parts yield local minima in the FEL. The reason is that the ssDNA of the interior loops is more flexible than the dsDNA of the strands (cp. sec. 2.6.2). This leads to an entropy rise and hence a drop in $G(n, f)$ associated with additional G_{loop} terms in eq. (2.10). Such a kind of sequences with equal types of bases in the double-stranded parts could be particularly useful to determine improved G_{loop} values by means of SMEs.

Let us now have a look at a double-stranded part consisting of bases A and complementary bases T [23], and an interior loop with TTT attached to the A bases and CCC connected to the T bases. With the help of the folding predictions of Mfold [143, 175] the linear folding structure of this molecule was checked. The contribution of the loop to the free energy, G_{loop}, is 3.24 kcal/mol. The height of the barriers can be regulated by varying the length of the double-stranded part. For instance, for a 10 bps long double-stranded part the barrier height is 8.1 $k_B T$ at

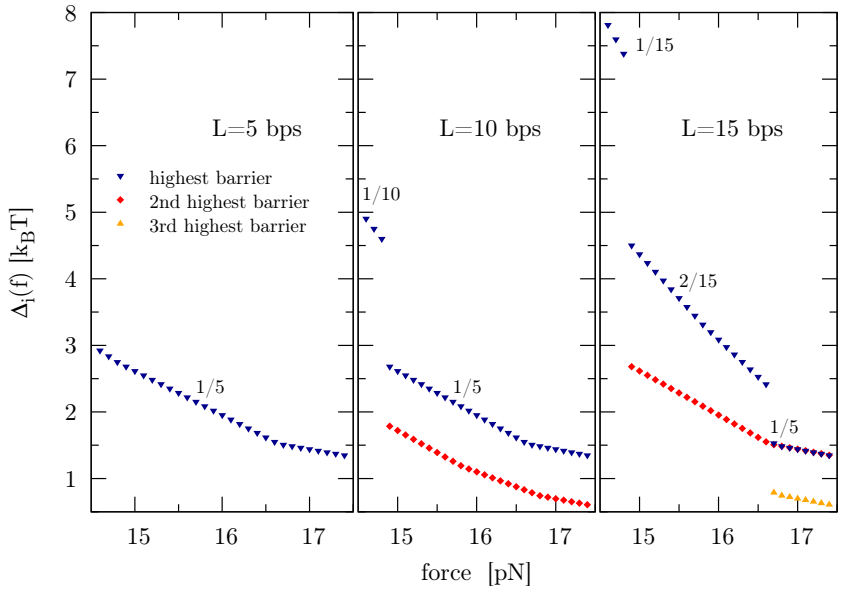


Figure 4.3: Comparison of the forward barrier heights $\Delta_i(f)$ of three random DNA sequences with a periodicity length $L = 5$ bps (left panel), $L = 10$ bps (middle panel) and $L = 15$ bps (right panel). The fraction denotes the number of barriers per number of bps in the stem.

$f = 10.4$ pN, where all minima are at equal level. For 20 bps we find a barrier height of $9.7\,k_BT$ at the corresponding force $f = 12.4$ pN. The sequence with 10 bps in the double-stranded part is referred to as sequence II. Its FEL is given in fig. 4.6(a) for three force values. Figure 4.6(b) shows that it follows the linear behaviour according to eqs. (4.2)-(4.4). Table 4.1 unifies the respective intercept and slope parameters of these functions.

To give an example for mixed types of bases in the double-stranded part, we use one of the branches of the triple-branch molecule [2] (see ch. 3). We employ the same loop as for the molecule considered above with equal types of bases, but reverse the order, so that the CCC bases are located in the strand closer to

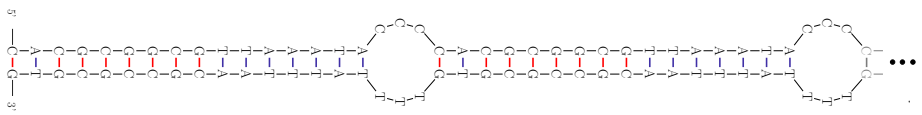


Figure 4.4: DNA hairpin molecule with sequence III. One basic unit consists of a double-stranded part with mixed types of bases and a ssDNA loop attached to it.

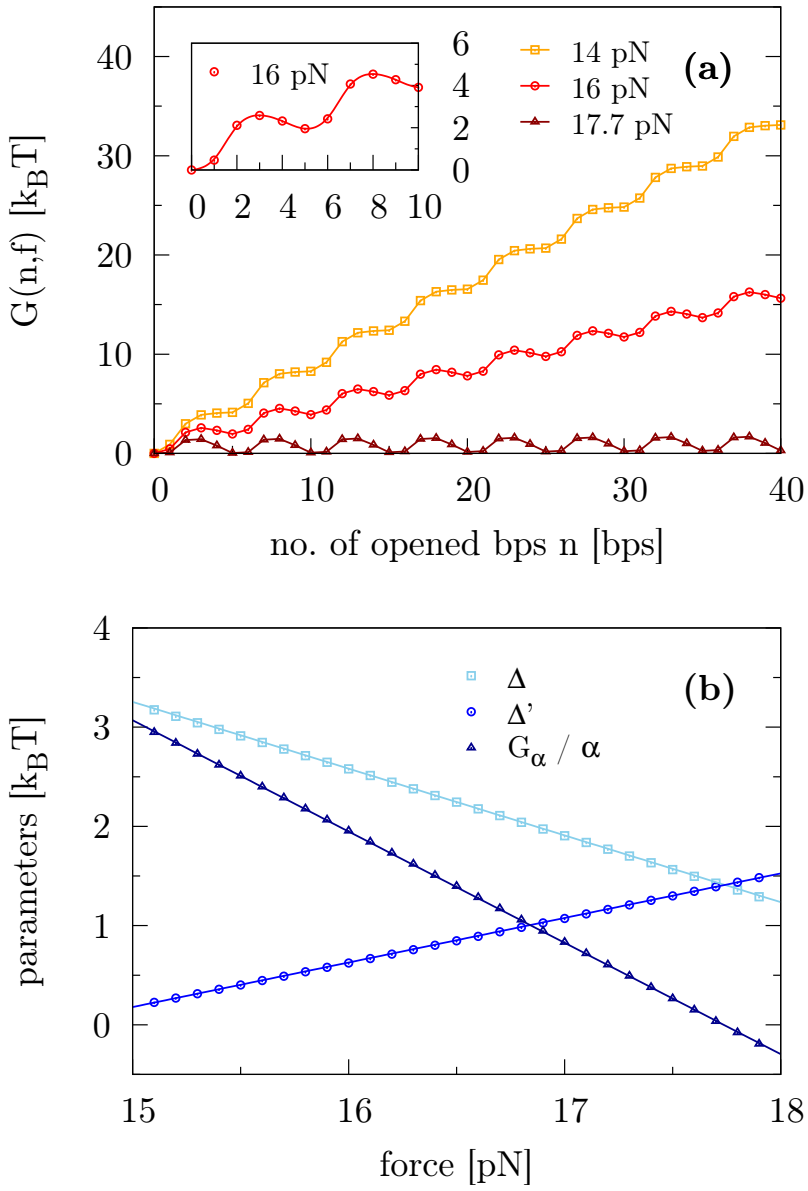


Figure 4.5: (a) FEL for the periodically continued sequence I for three magnitudes of force. The sequence consists of TGCAA as the basic unit and no interior loops. The inset depicts a zoom of the first two barriers for a force of 16 pN. (b) Forward barriers Δ, backward barriers Δ' and (order-normalised) state energies G_α/α as a function of force for the sequence I. The regression lines demonstrate the linear behaviour according to eqs. (4.2)-(4.4).

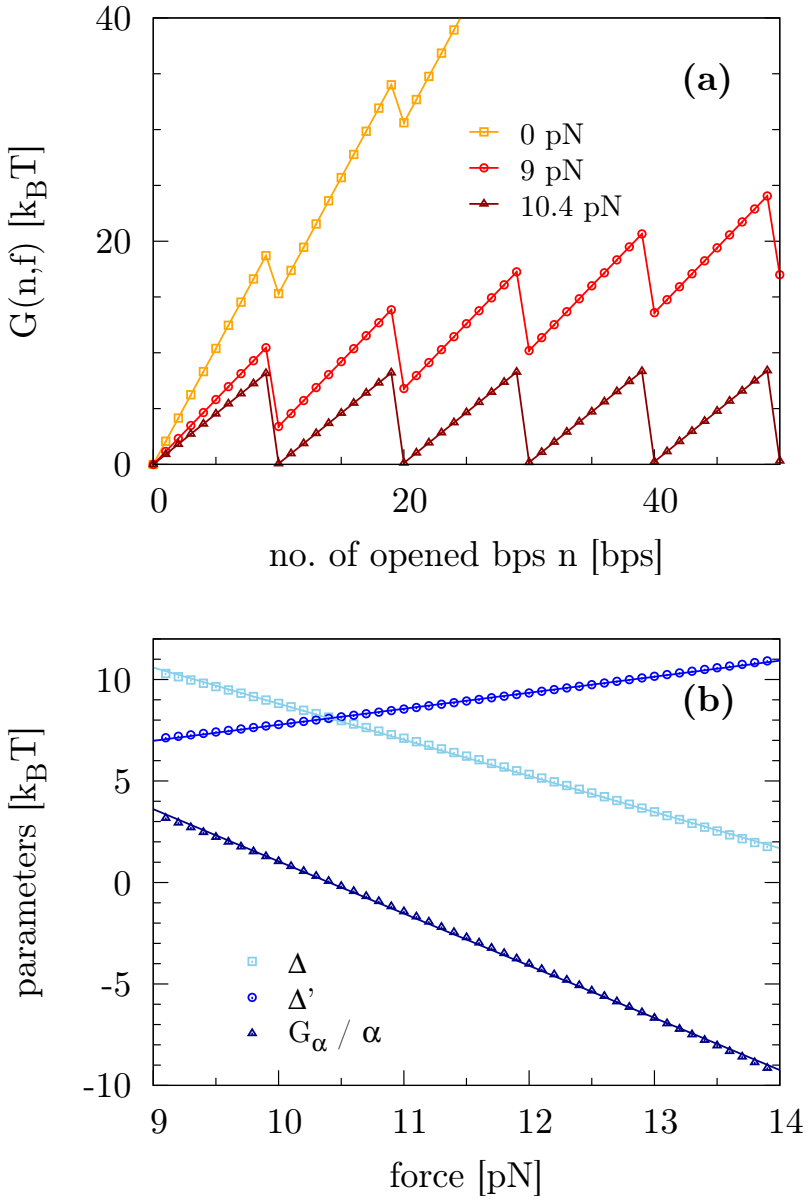


Figure 4.6: (a) FEL of the sequence II for three magnitudes of force. The sequence is composed of basic units which consist of 10 bases A and complementary bases T and an interior loop, where TTT is attached to the A bases and CCC to the T bases. (b) Forward barriers Δ, backward barriers Δ' and (order-normalised) state energies G_α/α as a function of force for the sequence II. The regression lines demonstrate the linear behaviour according to eqs. (4.2)-(4.4).

parameter	unit	sequence I	sequence II	sequence III
Δ_0	$k_\mathrm{B}T$	13.35	26.61	41.53
Δ_1	$\frac{k_\mathrm{B}T}{\mathrm{pN}}$	0.67	1.78	1.97
Δ_0'	$k_\mathrm{B}T$	−6.55	−0.13	−20.91
Δ_1'	$\frac{k_\mathrm{B}T}{\mathrm{pN}}$	−0.45	−0.79	−2.27
g_0	$k_\mathrm{B}T$	19.90	26.74	62.45
g_1	$\frac{k_\mathrm{B}T}{\mathrm{pN}}$	1.12	2.57	4.23

Table 4.1: Energetic parameters of eqs. (4.2)-(4.4) for the three sequences I-III.
I: Optimised sequence with TGCAA as a basic unit, without interior loops.
II: Basic units with 10 A bases paired with T, separated by interior loops.
III: Basic units with mixed types of bases of 16 bps length, separated by interior loops.

the 5' end. The corresponding G_loop value is 3.23 kcal/mol. Figure 4.4 displays the molecule labelled as sequence III and its FEL is given in fig. 4.7(a) for three force values. It follows eqs. (4.2)-(4.4), as depicted in fig. 4.7(b). The intercept and slope parameters of the linear functions are listed in tab. 4.1. At a force value of $f = 14.7$ pN, where all minima are at equal energy level, the barrier height is $12.6\,k_\mathrm{B}T$.

4.3 Kinetics and Monte Carlo simulations

According to standard transition rate theory [8], the rate to overcome the forward barrier is

$$\lambda(f) = \nu\ \exp\left[-\beta\Delta(f)\right]\,, \tag{4.5}$$

where ν is an attempt frequency and $\beta = 1/k_\mathrm{B}T$. Similarly, this equation can be written for the backward transition rate $\lambda'(f)$ with the backward barrier $\Delta'(f)$ in the exponential. Replacing f by t via eq. (4.1), one can write

$$\lambda(t) = \nu\ \exp[-\beta(\Delta_0 - \Delta_1 f_0)]\ \exp(\beta\Delta_1 rt)\,, \tag{4.6}$$

$$\lambda'(t) = \lambda(t)\ \exp[\beta(g_0 - g_1 f_0)]\ \exp(-\beta g_1 rt). \tag{4.7}$$

Considering a molecule with m periodic units, the probabilities $p_\alpha(t)$, $\alpha = 0, \ldots, m$, of being in state α evolve according to

$$\frac{\mathrm{d}p_0(t)}{\mathrm{d}t} = -\lambda(t)\ p_0(t) + \lambda'(t)\ p_1(t)\,, \tag{4.8}$$

$$\frac{\mathrm{d}p_\alpha(t)}{\mathrm{d}t} = -\lambda(t)\ \left[p_\alpha(t) - p_{\alpha-1}(t)\right] + \lambda'(t)\ \left[p_{\alpha+1}(t) - p_\alpha(t)\right], \quad \alpha = 1, \ldots, m-1\,, \tag{4.9}$$

$$\frac{\mathrm{d}p_m(t)}{\mathrm{d}t} = \lambda(t)\ p_{m-1}(t) - \lambda'(t)\ p_m(t)\,. \tag{4.10}$$

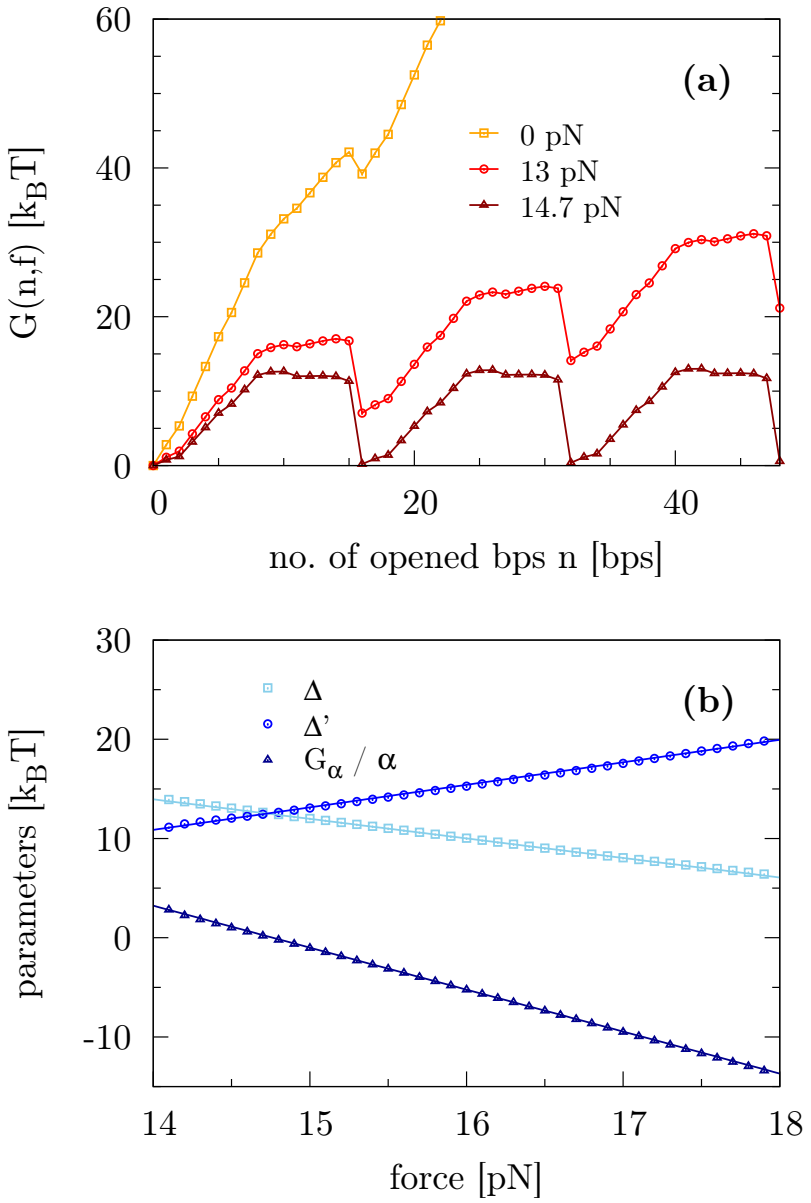


Figure 4.7: (a) FEL of the sequence III given in fig. 4.4 for three force values. (b) Forward barriers Δ, backward barriers Δ' and (order-normalised) state energies G_α/α as a function of force for the sequence III. The regression lines demonstrate the linear behaviour according to eqs. (4.2)-(4.4).

While for a two-level system ($m = 1$) an explicit solution is given in [16, 152], analytical treatments are more and more difficult for larger m. A solution for arbitrary m is developed in [67], where the backward transitions are neglected. In general, the cumulative rate

$$\begin{aligned}\Lambda(t, t_0) &= \int_{t_0}^{t} \mathrm{d}t' \, \lambda(t') \\ &= \frac{\nu \, \exp[-\beta(\Delta_0 - \Delta_1 f_0)]}{\beta \Delta_1 r} \left[\exp(\beta \Delta_1 r t) - \exp(\beta \Delta_1 r t_0)\right] \end{aligned} \tag{4.11}$$

enters the solutions, and similarly also $\Lambda'(t, t_0)$. Since we want to identify the parameters in eqs. (4.2)-(4.4), it is convenient to introduce the probability $\Psi_0(t)$ that, when starting in the folded state ($\alpha = 0$), no transition occurs until time t,

$$\Psi_0(t) = \exp[-\Lambda(t, 0)] \, . \tag{4.12}$$

Here only the forward rate is included, so that via a fit to experimental data the two parameters Δ_0 and Δ_1 can be extracted. In order to recover the two remaining parameters Δ_0' and Δ_1' (or g_0 and g_1) of eqs. (4.2)-(4.4), we introduce the probability $\Psi_1(t)$ that, when starting in the folded state ($\alpha = 0$), at least one forward, but no backward transition occurs until time t,

$$\Psi_1(t) = \int_0^t \mathrm{d}t_1 \, \exp[-\Lambda'(t, t_1)] \lambda(t_1) \exp[-\Lambda(t_1, 0)] \, . \tag{4.13}$$

The term $\lambda(t_1) \exp[-\Lambda(t_1, 0)]$ is the probability that the first transition occurs at a time t_1, and the term $\exp[-\Lambda'(t, t_1)]$ is the probability that thereafter no backward transition occurs. For any loading rate, $\Psi_1(t)$ goes through a maximum and approaches a finite value in the long-time limit. This maximum can become unnoticeable at high loading rates because backward transitions become the less likely the larger r.

In order to test the procedure, we perform Monte Carlo simulations of the stochastic process using the parameter values listed in tab. 4.1, henceforth refered to as the "true values". We consider molecules with $m = 4$ periodic units in all cases. The probabilities $\Psi_0(t)$ and $\Psi_1(t)$ are sampled from a set of N unfolding trajectories. Applying the First Reaction Time Algorithm (FRTA) [44, 66], the simulations can be performed in an efficient manner. Using the preceding transition time t_0, one generates a random attempt time t for a subsequent transition from a uniformly distributed number η in the unit interval by

$$\Lambda(t, t_0) = -\ln(1 - \eta) \, , \tag{4.14}$$

yielding

$$t = \frac{1}{\beta \Delta_1 \eta} \ln\left(\exp(\beta \Delta_1 \eta t_0) - \frac{\beta \Delta_1 \eta}{\nu \mathrm{e}^{-\beta(\Delta_0 - \Delta_1 f_0)}} \ln(1 - \eta)\right) . \tag{4.15}$$

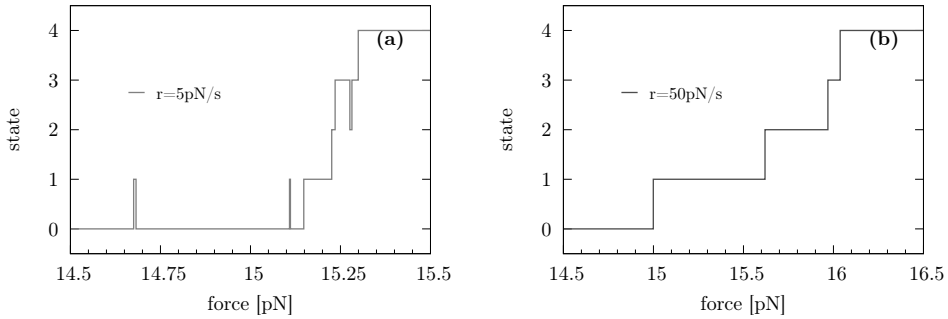


Figure 4.8: Molecular state as a function of the time-dependent force $f(t) = f_0 + rt$ for two different loading rates r in the relevant time (force) regime of unfolding. For smaller or larger times, the molecule is either in the folded ($\alpha = 0$) or unfolded ($\alpha = m$) state, respectively.

Analogously, an attempt time t' for a backward transition is generated from another random number η' (where Δ_0' and Δ_1' replace Δ_0 and Δ_1). The transition, taking place after the antecedent transition at time t_0, occurs at the smaller of both time values, that is $\min(t, t')$. Finally, the respective forward or backward transition is performed.

Below we will describe in detail the procedure applied to sequence III with mixed types of bases in the basic units including interior loops. The remaining two sequences I and II are treated analogously. As already stated in sec. 4.2.1, for sequence I without interior loops a description of the unfolding kinetics based on jump dynamics is not really appropriate, since the barriers between minima are too small. Keeping this in mind, we nevertheless include an analysis of this sequence for the sake of completeness.

Figure 4.8 displays typical simulated trajectories for loading rates of (a) $r = 5$ pN/s and (b) $r = 50$ pN/s, which are plotted as a function of the time-dependent force $f(t) = f_0 + rt$ (see eq. (4.1)). As anticipated from fig. 4.7(a), the first transitions take place around $f = 15$ pN, and afterwards the molecule quickly unfolds. Furthermore, backward transitions become the less frequent the higher the loading rate r is. For instance, in fig. 4.8(a) three backward transitions occur before reaching the unfolded state, while in fig. 4.8(b) no backward transition is observed. If nearly no backward transitions are included in the unfolding process, little information on the backward barriers is contained in the sampled $\Psi_1(t)$. On account of this, estimates for Δ_0' and Δ_1' are the less accurate the larger r. If the long-time limit of $\Psi_1(t)$ lies between zero and one, the statistics may be considered as sufficient (compare the discussion below).

Figure 4.9(a) depicts Ψ_0 which is sampled from $N = 1000$ trajectories for three different loading rates. The solid lines are a nonlinear fitting of eq. (4.12) to the

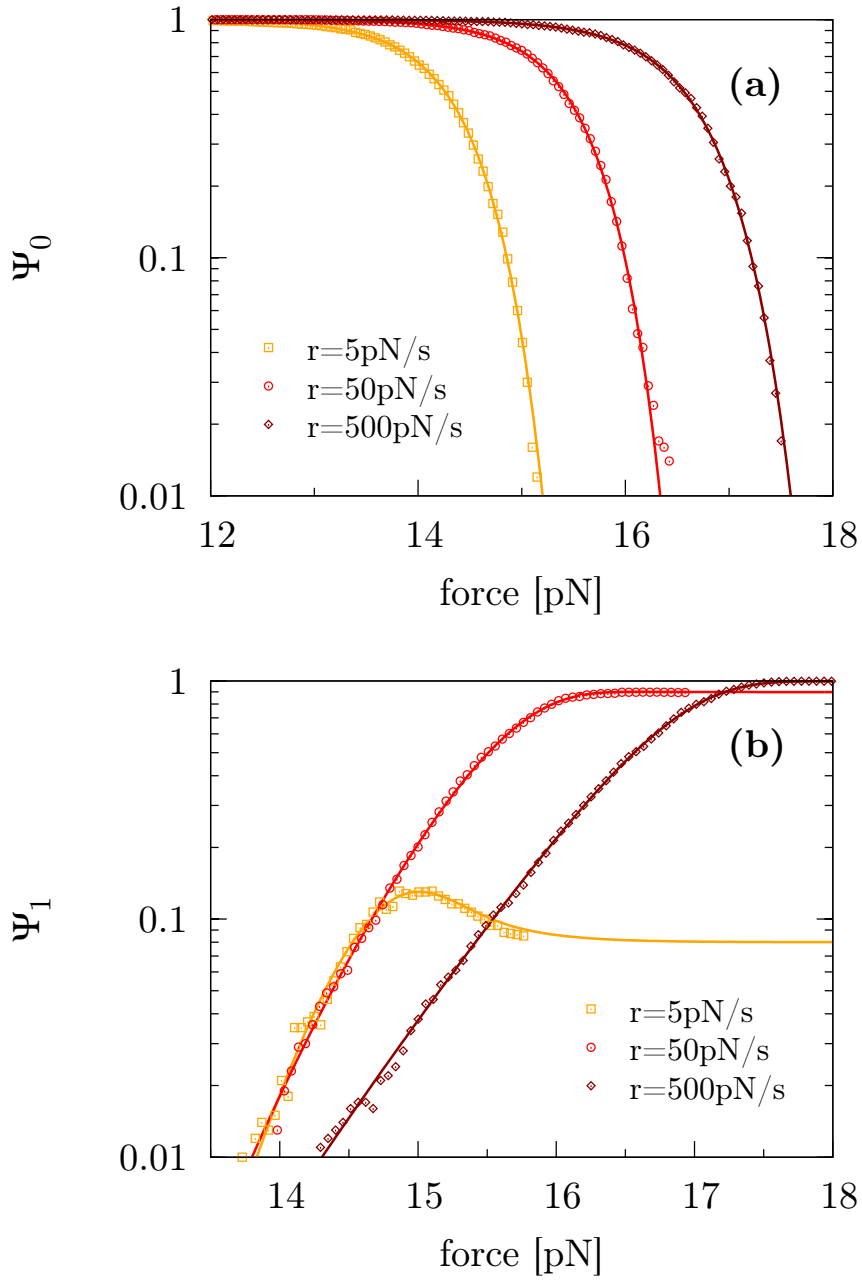


Figure 4.9: (a) Ψ_0 and (b) Ψ_1 as function of the time-dependent force $f(t) = f_0 + rt$. The symbols refer to KMC simulations of $N = 1000$ trajectories for three different loading rates. The solid lines are fits of eqs. (4.12) and (4.13) to the simulated data using the Levenberg-Marquardt algorithm.

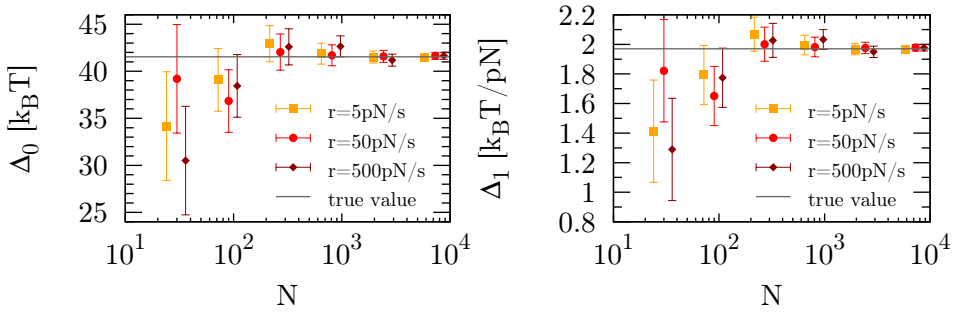


Figure 4.10: Estimates of Δ_0 (left panel) and Δ_1 (right panel) based on the analysis of N simulated unfolding trajectories for three different loading rates. The N values for r = 5 pN/s and r = 500 pN/s have been slightly shifted for better visibility.

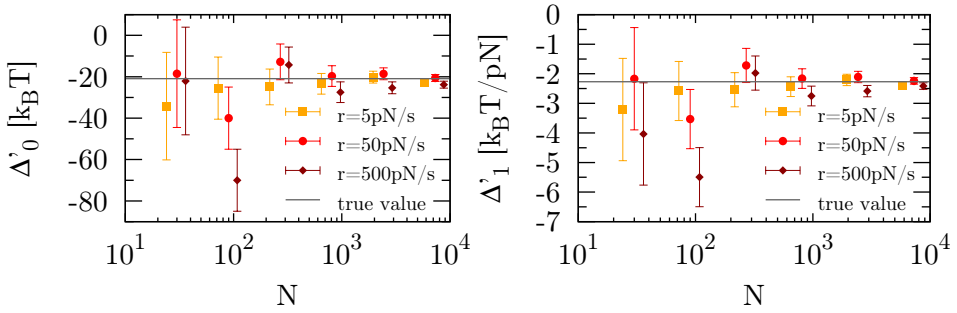


Figure 4.11: Estimates of Δ'_0 (left panel) and Δ'_1 (right panel) based on the analysis of N simulated unfolding trajectories for three different loading rates. The N values for r = 5 pN/s and r = 500 pN/s have been slightly shifted for better visibility.

data sets by dint of the Levenberg-Marquardt algorithm [133].[i] The simulated data follows closely the predicted behaviour and the respective fit parameters are given in fig. 4.10 as a function of the number of generated trajectories N. In SMEs it is already an ambitious task to record 1000 trajectories, but this is not out of reach. Approximately 300 trajectories are sufficient to ensure that the fitted Δ_0 and Δ_1 are quite close to the true values (see fig. 4.10). In simulations, larger numbers of trajectories can be easily generated and the fit parameters converge to the true values for $N \to \infty$.

With respect to applications, it is encouraging that already for 300 trajectories

[i] In the Levenberg-Marquardt algorithm, in addition to the estimated $\Psi_j(t)$ values, $j = 0, 1$, their standard deviations $\sigma_j(t)$ are required. In the simulation, they can be obtained by repeating the procedure for estimating $\Psi_j(t)$ many times for the same number of trajectories. The standard deviations have a pronounced time-dependence and their maximum lies in the "unfolding regime" (cf. fig. 4.8). Fortunately, using time-independent σ_j has a negligible effect on the best fit parameters Δ_j. Therefore, constant σ_j can be used in an experimental realisation, where only one set of N trajectories is available.

the fitted Δ_0 and Δ_1 have a deviation of only a few percent, even when the fit starts with guessed input values deviating by 40 percent from the true parameters. The error bars in fig. 4.10 were obtained by repeating many times the analysis for a certain number N of trajectories. However, in an experiment the estimation of the standard deviation would be difficult.

As outlined above, in a second step we can determine the backward barriers Δ_0' and Δ_1' by analysing $\Psi_1(t)$. Therefore, the forward barriers are fixed to their fitted values. In fig. 4.9(b) the fits of eq. (4.13) to the simulated data are displayed. The whole procedure is accomplished completely analogous to the determination of Δ_0 and Δ_1. Figure 4.11 depicts the respective fit results of the energetic parameters. Actually, we attained equivalent findings with respect to the quality of the fitting. Particularly, already 300 trajectories, as typically recorded in SMEs [52, 123], are enough to result in estimates of the energetic parameter deviating by only a few percent from the true values.

Similar analyses have been performed for the other two sequences, I and II. The respective results for $N = 1000$ trajectories are listed in tab. 4.2, along with the values for sequence III. For sequence I, it was only possible to extract Δ_0 and Δ_1, since the backward barriers are so small that backward transitions take place very rapidly after the first forward transition. Consequently, $\Psi_1(t)$ is nearly zero for all t. The same applies to sequence II for the loading rate $r = 5$ pN/s. A way to overcome this problem could be to generalise the technique by introducing probabilities that at least $k > 1$ backward transitions occur up to time t. Note that $1 - \psi_1(t)$ is the probability that up to time t at least one backward transition takes place. Unfortunately, the complexity of analytical expressions for such probabilities increases with k, so that a fit to the sampled data will be hard to handle in practise. As already discussed in sec. 4.2.1, for sequences with small energetic barriers the application of equilibrium methods is more appropriate. Good estimates are attained in all cases where it was possible to extract the energetic parameters of eqs. (4.2)-(4.4). As illustrated above, these estimates are the less accurate the higher r, since then $\Psi_1(t)$ approaches values close to one at long times. In order to get good estimates, one should therefore tune the loading rate in such a way that the long-time limit $\Psi_1(t)$ is roughly one half. In doing so, the "level of irreversibility" which is given by r is matched to the time scale of molecular relaxation.

4.4 Force-controlled pulling experiments

In our model of the FP, which is used in this chapter, we assume that the force can be controlled during the whole experiment, in particular during those instants when bps break. Indeed there exist two promising experimental ways to apply a linearly increasing FP (cp. eq. (4.1)) using single-molecule force spectroscopy methods, namely MTs and OTs. Spatial resolution can go down to 2 and 0.1 nm, whereas measurable forces can be as small as 0.01 and 0.1 pN, respectively [121].

A well-suited electromagnetic tweezers apparatus, which was developed by

sequence	r	Δ_0	Δ_1	Δ_0'	Δ_1'	g_0	g_1
unit	pN/s	$k_\mathrm{B}T$	$\frac{k_\mathrm{B}T}{\mathrm{pN}}$	$k_\mathrm{B}T$	$\frac{k_\mathrm{B}T}{\mathrm{pN}}$	$k_\mathrm{B}T$	$\frac{k_\mathrm{B}T}{\mathrm{pN}}$
I	true	13.35	0.67	-6.55	-0.45	19.90	1.12
	5	13.34	0.65	*	*	*	*
	50	13.35	0.66	*	*	*	*
	500	13.28	0.64	*	*	*	*
II	true	26.61	1.78	-0.13	-0.79	26.74	2.57
	5	26.63	1.79	*	*	*	*
	50	26.76	1.79	-0.05	-0.78	26.81	2.57
	500	26.59	1.77	0.22	-0.74	26.37	2.51
III	true	41.53	1.97	-20.91	-2.27	62.45	4.23
	5	42.09	2.01	-24.66	-2.51	66.75	4.52
	50	42.41	2.03	-17.43	-2.04	59.84	4.07
	500	39.31	1.83	- 15.96	-1.97	55.27	3.80

Table 4.2: Estimates of the energetic parameters of eqs. (4.2)-(4.4) for the sequences I-III based on sampling $\Psi_0(t)$ and $\Psi_1(t)$ from $N = 1000$ unfolding trajectories. The results are given for three different loading rates r. An asterisk indicates that the determination of Δ_0' and Δ_1' from $\Psi_1(t)$ was not possible (see discussion in main text).

Fisher *et al.* [49], combines six thin-foil pole plates for a three-dimensional position control with a laser-based particle tracking system. It is enabling a position feedback with a bandwidth of 10 kHz, instead of typically 60 Hz employing video-based tracking [120]. With a force resolution of 0.01 pN [29], a force bandwidth in excess of 3 kHz [49] and a maximal achievable force of 700 pN when using a 1 μm paramagnetic particle, this system fulfills all requirements to provide a linear force ramp, even though the force needs to be adjusted instantly in the course of each disruption event, i. e. unfolding of a set of several bps, to maintain the desired protocol. Since the lateral resolution is 2.4 nm [49], this setup is very convenient to record EFCs (see sec. 2.5).

MTs naturally operate in the so-called force clamp mode where the force is kept constant (”clamped”) and the response of the molecule under study, e. g. a DNA hairpin, is measured regarding the position change of the force probe, i. e. the paramagnetic particle [50]. Due to their relatively low stiffness, MTs need an active feedback to work in the position clamp mode, which in turn is the natural system for AFM and OTs. In this modality the force response is recorded, while the force is adjusted in order to maintain a constant position. To transform such a spring-based system with a high spring constant, operating as a passive position clamp, to force clamp mode, a force feedback is required. It is possible to achieve this either actively [90] with a restricted bandwidth or even passively [58] without such a limitation, so that the experimental setup of [58] is ideally suited for the investigation of DNA hairpin unfolding, where both a high bandwidth and a high accuracy are needed. Instead of using video microscopy for particle tracking, as

well the resolution of OTs is greatly improved by applying a quadrant photodiode, thereby gaining sub-nanometer spatial resolution and detection bandwidths of more than 100 kHz [121].

With the help of a dynamic position control which can be incorporated into a feedback loop [90] one can adjust the force on the bead in the optical trap. One way is to move the fluidic chamber and thus the glued on micropipette using a piezoelectric motor stage [121]. This method is comparatively slow with a response time of about 10 ms, but on the other hand allows a full three-dimensional control over a rather large range of motion ($\sim$ 100 µm) without additional forces acting on the bead in the trap, since the fluid inside the chamber is not set in motion ("slip conditions") [3]. By means of capacitive position detection, the accuracy in position of such a feedback loop is at least within the range of nm [120].

The second possibility to govern force is to steer the trapping beam. Making use of an acousto-optic deflector [146], this movement can be done very fast, with a response time of only approximately 10 µs [121]. Although this response is sufficiently fast to take the measurement of biochemical kinetics [90], motion is restricted to a few μm in only one axial plane [121]. Trap-steering operates in "stick conditions" because the liquid inside the chamber is dragged and therefore the bead in the optical trap is exposed to an additional Stokes force [105, 165, 3].

For a good resolution in time, position and force an optical trap which is softer than the handles should be used, with a stiffness preferably below 0.1 pN/nm [106]. Independently of the experimental design of the OT setup, there is an intrinsic limitation due to the corner frequency of the bead in the trap. Its inverse is the time the bead needs to relax in the trap, typically lying in the range of 1 to 10^{-2} ms [106], i. e. all force changes occuring faster than this pass unnoticed by the bead (cp. sec. 5.1.1).

Chapter 5

Unfolding kinetics of DNA hairpins using a distance protocol

5.1 Calculation of the free energy landscape in the distance protocol

In this chapter we want to calculate the FEL by applying a more detailed model than before (cp. ch. 3), including not only the contributions due to the formation of the dsDNA and the stretching of ssDNA, but also the contributions of stem diameter, bead and handles. The different components of the whole molecular setup were described in sec. 2.3 and the FEL calculation was outlined in sec. 2.6. In contrast to ch. 4 where the FP was considered, we refer here to typical pulling experiments where the distance between the optical trap and the micropipette is controlled [105]. Therefore, we work with the DP in the FEL calculation, which should be ideally applied. However, due to the reduced calculation effort and since it is a correct description when the fluctuations are small [173], the FP is often used for theoretical descriptions even in the DP case (see ch. 3 and [117, 51, 2]).

Considering unzipping experiments with a loading rate r, where the distance increases linearly with time,

$$Z(t) = \frac{rt}{k_{\mathrm{eff}}}, \tag{5.1}$$

the energy landscape gets tilted with increasing $Z(t)$. Since we here investigate short DNA hairpins [128], two local minima ("states") are located at $n = 0$ and $n = n_{\mathrm{tot}}$, where the local rise of $G_{\mathrm{form}}(n)$ changes significantly (see sec. 2.6.1 and 4.1). In-between them, more minima can form ("new barrier effect", see fig. 5.1) and shift their position ("shift effect", see sec. 4.2.1). We thus deal with a multi-state folder. Once the tilting is large enough that the level of the minimum of the unfolded state becomes comparable to the energy of the folded state, a thermally activated transition becomes likely, leading to the unzipping of the molecule.

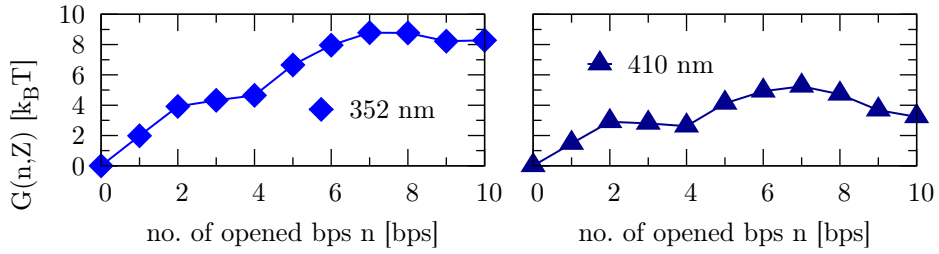


Figure 5.1: The "new barrier effect". In the left panel, the FEL $G(n, Z)$ is depicted for $Z = 352\,\text{nm}$ where only one maximum is visible (at $n = 7\,\text{bps}$). In the right panel, at $Z = 410\,\text{nm}$, an additional minimum at $n = 4\,\text{bps}$ emerged, so that two barriers are formed, with maxima at $n = 2\,\text{bps}$ and $n = 7\,\text{bps}$.

During the simulation which we will describe in sec. 5.2.1, it is sufficient to know the positions of the neighbouring barriers at the actual time or distance (see eq. (5.1) for the relation between time and distance). Then one can calculate the barrier height, which has to be overcome at the present moment in order to change the state. It is, however, much more convenient to explore the whole FEL before performing the simulation and to store the essential properties (see sec. 5.1.2), which we will need in the course of the simulation (see sec. 5.2.1). Since we deal with an adiabatic process (see sec. 5.1.1), where the perturbation of the system is slow, it is possible to calculate the full FEL in advance.

5.1.1 Adiabatic process

Like during the opening of a zipper, when a sufficiently high force is applied, the bps are opened and stretched at the same time, because the whole molecule relaxes very fast. The typical relaxation time of handles, ssDNA and bps is much lower than the relaxation time of the bead [106]. During the unfolding process, the single bps from $n = 0$ to n_{tot} are opened in a sequential manner, i. e. bp after bp. Force is only applied to the first bp closest to the fork. When it is opened and the ssDNA is stretched, force can be applied to the adjacent, still closed bp. While the molecular extension $m = h + x$ (see eq. (2.4) in sec. 2.3) remains unchanged during the opening of one bp, the single components of the whole molecule adapt their lengths very fast [106] in order to equilibrate at the new value of n, while the force stays constant. When a few bps opened, it can happen that the force applied to the first bp closest to the fork is no longer large enough to open further bps or even the whole molecule. Then the unfolding process stops for a while, during which the bead in the optical trap has time to relax to a new equilibrium force. For a given distance Z and number of opened bps n, one can extract this equilibrium force out of eq. (2.3). When several bps opened, we can clearly observe a decrease

in the measured equilibrium force due to the increase of the ssDNA length, which has to be compensated by a lower force since the value of Z stays constant during the force rip (cp. sec. 3.3). Using the DP of eq. (5.1), the distance is increased linearly. Typically this process occurs much slower than the bead relaxation [106]. The increase of the distance leads again to a higher equilibrium force, which in turn favours the opening of further bps in the same way as described above. Due to the large separation of timescales between the relaxation of the single parts of the system and the unfolding kinetics it is thus possible to simulate a pulling experiment as an adiabatic process [105, 54], where the whole FEL can be computed in advance.

5.1.2 Technical details of the storage of the free energy landscape properties

The following two sets of data can be, for convenience, prepared before starting the simulation. We always need to store the distance and the total number of barriers. Firstly, we need the height of all forward and backward barriers for a very fine mesh of distances ("barrier heights file"). The minimal height of barriers needs to be specified in our procedure. The dynamics of thermally activated transitions breakes down when the barriers reach a minimal barrier height of about $1\,k_BT$. For smaller barrier heights the use of Brownian dynamics is required. However, the jump dynamics is still valid and correctly reproduces the Brownian dynamics below this limit. Yet barriers smaller than $1\,k_BT$ may no longer be called true barriers. Accordingly, we imposed a limit of $0.2\,k_BT$ on the minimum barrier height.

Secondly, we store a list of the topological changes of the FEL ("topological changes file") including the bp position n of all barriers (i. e. maxima) and minima at the distances at which the topology of the FEL changes. Furthermore, we add the actual number of the respective minimum connecting one line of the topological changes file with the next one ("marker"). We will illustrate this concept of "relabelling" by an exemplary topological changes file (see tab. 5.1 and fig. 5.2).

distance Z [nm]	total no. of barriers	n of max.	n of min.	marker
200.00	1	20	0 21	1
285.00	2	11 20	0 14 21	2
349.18	2	11 20	0 13 21	1 2
496.06	1	20	0 21	0 1

Table 5.1: Example of a part of a topological changes file.

In a first case let us consider the "new barrier effect" (see fig. 5.1 and sec. 4.2.1). Imagine that in line 1 of the topological changes file, i. e. at $Z = 200.00\,\text{nm}$, the total number of barriers is one. Hence, we have one minimum worth mentioning, apart from the minimum number 0, which we do not refer to further since it always stays at $n = 0$ (see part (1) of fig. 5.2). In line 2, i. e. at $Z = 285.00\,\text{nm}$, a second barrier has formed. Consequently, there exist two minima, whereas the

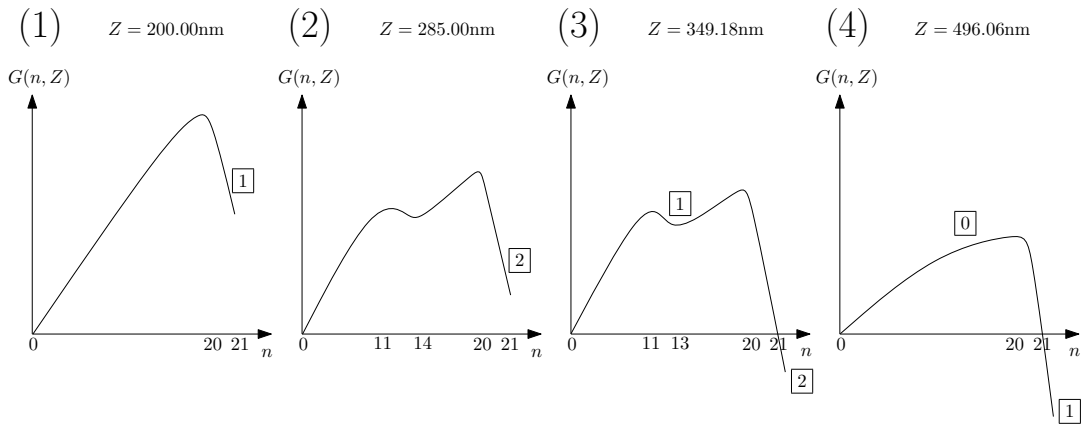


Figure 5.2: Illustration of the topological changes file of tab. 5.1. The FEL $G(n, Z)$ is sketched for four different distances Z. We further label the bp positions n of minima and maxima. The numbers in boxes indicate the markers for the changes of the minima from one FEL to the next one.

new minimum is created on the left side of the previous minimum 1 (see part (2) of fig. 5.2). Then the marker of the old minimum 1 will be relabelled to minimum 2, since the new minimum has to be labelled as 1, respecting the order of the minima according to the number of opened bps n. At the end of line 1 of the topological changes file we thus read "1", at the end of line 2 we find a naked "2" (see also the numbers in boxes of part (1) and (2) of fig. 5.2). This means that only the right minimum in our example is included in this relabelling marker for a simply reason: the molecule can only be in minimum 0 or 1 at the start of the simulation. When the distance specified in line 2 is reached, the molecule can perform a transition in which it arrives at the new minimum 1. This is a direct transition which is not connected to the relabelling introduced above. The molecule knows nothing about this new minimum 1 before reaching the distance of the topological change of line 2 and can access it only via a transition.

A second case is when only the position of at least one minimum or maximum changes, but not the total number of barriers (see part (3) of fig. 5.2). In our example, in line 3, i. e. at $Z = 349.18$ nm, we then write for the relabelling markers simply 1 for minimum 1 and 2 for minimum 2, since no real relabelling takes place, but a so-called "shift process" (cp. sec. 4.2.1). In this case the position of a minimum or maximum is changed, e. g. here minimum 1 changes its position from bp 14 to 13. Of course a shift can also occur at the same time of a pure topological change where the total number of barriers changes.

Let us now consider the third case, the relabelling when a minimum vanishes (see part (4) of fig. 5.2). Let in line 4, i. e. at $Z = 496.06$ nm, the topological change consist of a reduction of one minima, thus back to one barrier. We then write 0 and 1 as relabelling markers, since the old minimum 1 is vanishing completely,

while the minimum 2 is again labelled as 1. What happens if the molecule is in a minimum that is vanishing? It simply shifts to the neighbouring minimum at the side where the barrier vanished, a further example for a "shift process". It has then a lower free energy value compared to the old minimum position.

Having stored all this information in the files barrier heights and topological changes, one can perform the simulation faster since the underlying FEL with its respective changing barrier heights and positions of barriers and minima is well known.

5.2 Kinetics and Monte Carlo simulations

5.2.1 Reaction Time Algorithm

Out of the stored forward (backward) barriers adjacent to the actual minimum the molecule is in, we calculate the forward (backward) transition rate $w_+(t)$ ($w_-(t)$) according to standard transition rate theory [8], in case the molecule can perform a forward (backward) jump, since it is not in the final (initial) state $n = n_{\mathrm{tot}}$ ($n = 0$). Otherwise $w_+(t)$ ($w_-(t)$) is set zero. For the forward transition rate we thus can write, replacing Z by t,

$$w_+(t) = \nu \, \exp\left[-\beta\Delta(t)\right] \tag{5.2}$$

where $\Delta(t)$ is the forward barrier, ν is an attempt frequency (see discussion in sec. 5.2.3) and $\beta = 1/k_{\mathrm{B}}T$. A corresponding equation applies to the backward transition rate $w_-(t)$ with the backward barrier $\Delta'(t)$ in the exponential.

The total transition rate for a transition at time t is $w(t) = w_-(t) + w_+(t)$. Let $W(t, t_0) = \int_{t_0}^{t} dt' \, w(t')$ be the cumulative rate, using the time t_0 of the previous transition or the last topological change. Every topological change which alters $w(t)$ via $w_-(t)$ and / or $w_+(t)$ is relevant. The others are not important, that means when they concern transitions that are not directly (i. e. within one transition) reachable from the actual minimum at time t. However, as explained above, in the overview file "topological changes" they are also listed since they may be relevant for the simulation of another trajectory.

These Monte Carlo simulations can be performed in a very efficient manner using the Reaction Time Algorithm (RTA) described in [44, 66]. Let t_0^* be the time of the next topological change after the time t_0. Then one has to perform the following steps:

(1) Draw a random time t_r via the transformation formula

$$W(t_r, t_0) = -\ln(1 - u_1)\,, \tag{5.3}$$

where u_1 is a uniformly distributed number in the unit interval. In case the molecule is in the last state, the forward transition rate vanishes, $w_+(t) = 0$, and the molecule can only perform a backward transition with the rate $w_-(t)$.

At a constant loading rate, this backward transition rate is decreasing so fast, that $W(\infty, t_0) = \lim_{t\to\infty} W(t, t_0) < \infty$. We thus have to use the generalised transformation formula

$$W(t_r, t_0) = -\ln\left(1 - N(t_0)\, u_1\right), \tag{5.4}$$

where

$$N(t_0) = 1 - \exp(-W(\infty, t_0)). \tag{5.5}$$

(2) If $t_r > t_0^*$, set $t_0 = t_0^*$ and continue with (1). In this case relabelling and perhaps a change in position ("shift process") need to be accomplished (see explanations concerning the topological changes above). Finally, the rate $w(t)$ has to be updated. If $t_r < t_0^*$, continue with (3).

(3) Draw another independently distributed random number u_2 out of the unit inverval. If $w_-(t_r)/w(t_r) > u_2$, perform a backward, otherwise a forward transition, that means change the actual minimum number and update the actual n of the minimum. Then set $t_0 = t_r$ and continue with (1).

5.2.2 Force-distance curves (FDCs)

In order to obtain a FDC, the transition time of the α^{th} transition, t_α, is converted via eq. (5.1) into the distance Z_α at which the α^{th} transition ("rip") takes place. Using eq. (2.3), the force f_α at the beginning of the transition can be computed. Note that during this transition, the distance Z_α stays constant, since the transition takes place much faster than Z is increased by the DP (see sec. 5.1.1). Therefore, at the same value of Z, we get two values of force for the two values of n (the minima the molecule is before and after the transition), whose difference Δf_α gives the sudden de- or increase ("jump") in force. Δf_α is related to the respective hairpin sequence. As described in sec. 3.2, in a typical unfolding curve, the force increases linearly, determined by the elastic response of the DNA handles and the already unfolded ssDNA. Sudden transitions from one state to another induce a jump in force, since either parts of the hairpin molecule unfold or fold. Due to this abrupt length change, the bead changes its position in the optical trap and causes the force jump (see sec. 5.1.1).

Figure 5.3 depicts typical simulated FDCs for hairpins A (left panel) and E (right panel), where the force is plotted as a function of the linearly increasing distance $Z(t) = rt/k_{\text{eff}}$ (cf. eq. (5.1)). Note that we used the loading rates of [128], i. e. $r = 16\,\text{pN/s}$ for hairpin A and $r = 5.2\,\text{pN/s}$ for hairpin E. After a first short opening of about the first half of hairpin A, the molecule stays trapped in the completely folded state for a long time. Between 15 and 16 pN hairpin A unfolds and folds again, at least partly, several times. The grey line with triangles shows the pure results of our simulation, including many fast transitions. Since we include more than two minima into our considerations, the hairpin repeatedly hops between two neighbouring minima. From time to time it reaches the entirely unfolded state, in

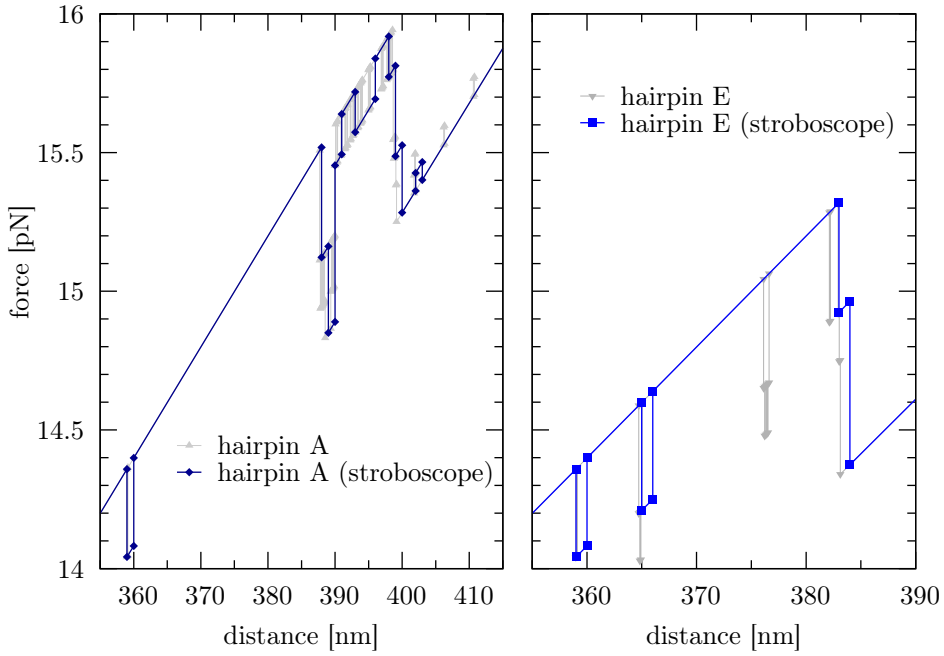


Figure 5.3: Typical simulated FDCs for two of the DNA hairpins used in [128]. In addition to the pure simulation results using a multi-state folder, we plot the same FDC investigated with a stroboscope (see main text).

which it stays for good only at high distance, when the last backward barrier is getting higher and higher, so that the molecule can not escape any more from the completely unfolded state. The dark blue line with diamonds depicts the stroboscope version of the same trajectory, that means when we only have a look at the FDC at a sequence of times separated by a small time step. We chose to glance at the FDC every 1 nm, according to the distance resolution given in [128]. Then, in our example trajectory, we see only two times a complete unfolding with a few intermediates. Due to the force fluctuations encountered in pulling experiments, our results are comparable to typical experimental FDCs (see [128]). Especially in the stroboscope version of the FDC the tendency of a two-state behaviour is visible, since the intermediate states can be overlooked due to thermal noise.

Hairpin E shows a few more short unfolding transitions in the range of about 360 to 380 nm as compared to hairpin A. This is not attributed to hairpin E and could have happened in the same way in the trajectory of hairpin A since the barrier heights are the same except for the last backward barrier. However, hairpin A was pulled at a higher loading rate so that less of these first short unfolding transitions are likely to occur. Nevertheless, also hairpin E mainly stays in the completely folded state up to 380 nm. The grey line with triangles again depicts

General parameters	
salt concentration	0.5 M NaCl
temperature	room temperature (≃ 25 °C)
k_{eff}	0.04 pN/nm
k_{t}	0.05454 pN/nm
k_{h}	0.15 pN/nm

Hairpin specific parameters				
hairpin	pulling speed [pN/s]	loading rate [nm/s]	no. of curves	bases of loop
A	400	16	760	4
B	200	8	203	6
C	65	2.6	289	12
D	200	8	526	16
E	130	5.2	182	20

Table 5.2: Overview of the parameters used in order to compare our simulated data to the experimental findings of [128].

the pure simulation results, where the molecule unfolds entirely much faster and with less backward transitions compared to hairpin A. This behaviour with only one complete unfolding is also reflected in the stroboscope plot (blue line with squares). Due to the much higher last backward barrier (about 3.5 times the barrier height of hairpin A) which has to be surmounted once the hairpin reached the totally unfolded state, hairpin E typically stays trapped in this state. Since the end-loop is much longer for hairpin E, i. e. 20 instead of 4 bases for hairpin A (see [128] and tab. 5.2), the contour length for $n = n_{\mathrm{tot}}$ is much larger (see eq. (2.12)) and thus the force right after the transition to the completely unfolded state is lower.

5.2.3 Work histograms

The work calculation is performed for every trajectory once all transitions are known according to $W = W_Z(f) = \int_{Z_{\mathrm{in}}}^{Z_{\mathrm{fin}}} f\, dZ$, i. e. the area below each FDC (see eq. (2.8) and sec. 2.5). Out of all work data we generated work histograms and give the respective parameters used in the simulation in tab. 5.2. We compare our simulated results with the experimental data reported in [128] for all five hairpins (see fig. 5.4). Open symbols refer to our simulated work histograms for the same number of unfolding cycles as recorded in the respective experiments (filled symbols). The overall agreement is quite good. Note that here only one sample was measured for every hairpin A to E [128]. In order to decrease the high variability among different samples of the same molecule, one typically averages over several samples [117, 2, 72]. The different shape of the experimental work histogram of

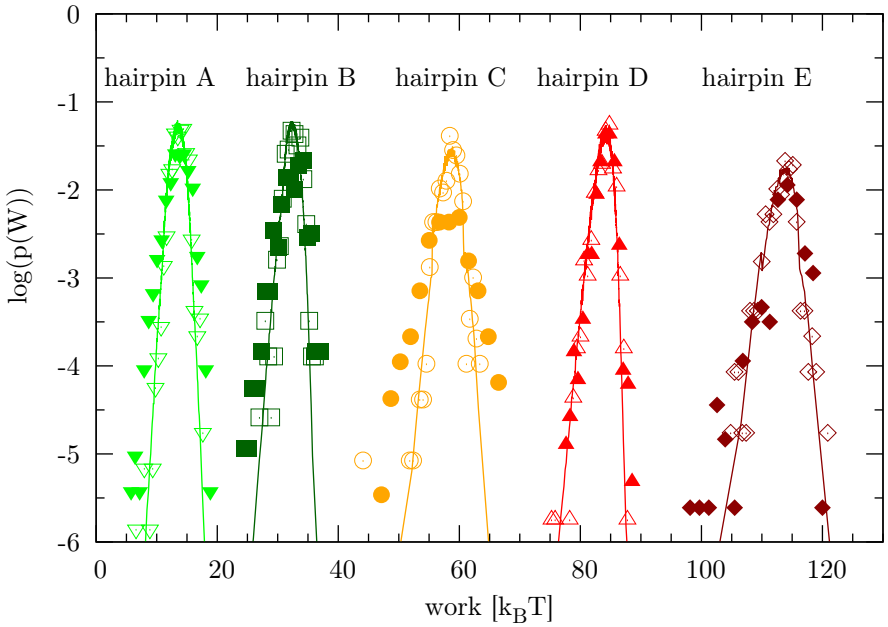


Figure 5.4: Comparison of the work histograms of our simulated data with the experimental findings of [128]. For each of the five DNA hairpins used in [128] we plot the experimental results (filled symbols) and our simulated work histograms (open symbols) for the same number of unfolding cycles using equidistant binning. In addition, we depict 10^6 simulated unfolding curves (lines) using adapted binning.

hairpin C might be attributed to this sample variability.

The attempt frequency ν was used as a fitting parameter in order to get a comparable width of the work histograms. We used $\nu = 5 * 10^6$ Hz for the simulations of hairpin D, $\nu = 5 * 10^4$ Hz for hairpins A and B, $\nu = 1 * 10^4$ Hz for hairpin E and $\nu = 5 * 10^3$ Hz for hairpin C. Most of these values lie a bit lower than typical values [25, 98, 102, 10] (cf. sec. 3.4). Recent experiments show that ν varies for different molecular structures [94]. Note that the work calculated for the trajectories using a stroboscope (see 5.2.2) is the same as using the full number of transitions. This means that even when the fast transitions are not detectable in the experiment, the measured work is not affected.

We compare furthermore in fig. 5.4 the experimental work histograms of [128] with 10^6 simulated unfolding curves (lines). Here we did not use an equal bin size (equidistant binning) for the histogram of the simulated data which is often employed (see e. g. [28, 117, 128]), but rather used an equal number of entries per bin (adapted binning), resulting in a different size for each bin. We thus obtain

less, but more significant data points in the tails. On the other hand, we get a larger number of data points near the average work, which scatter more than when using equal bin size since we use far less data points per bin. Our way of generating the work histograms is thus especially advantageous when the total number of work values is limited, e. g. in experiments, and when the tails are of utmost importance (see secs. 5.3.2 and 5.3.3).

5.3 Application

5.3.1 Free energy difference calculation

In the following we express all energies in units of $k_\mathrm{B}T$ referring to room temperature. For a certain number N of measured work values W, we first calculate the Jarzynski estimator via [80, 128]

$$\Delta G_N = -\log \frac{1}{N} \sum_{i=1}^{N} \exp\left(-W_i\right). \tag{5.6}$$

In a second step, we use the improved Jarzynski estimator suggested by Palassini and Ritort [128]

$$\Delta G_N^* = \Delta G_N - \langle B_N \rangle \tag{5.7}$$

for the unidirectional case, in which the bias expectation $\langle B_N \rangle$ is included (cp. sec. 5.3.3). In order to finally recover the free energy of formation ΔG_0, we need to subtract the stretching energy $G_\mathrm{str}^\mathrm{sim}$ of the ssDNA, bead and handles from ΔG_N or ΔG_N^*. Therefore, we define a certain range in distance used for the work calculation in such a way that the molecule always starts in the completely folded state ($n = 0$) and always ends in the entirely unfolded state ($n = n_\mathrm{tot}$). We set our starting and final point of the work calculations to $Z_\mathrm{in} = 200\,\mathrm{nm}$ and $Z_\mathrm{fin} = 500\,\mathrm{nm}$. In our case, the initial force f_in can be recovered by using $n = 0$ and $Z_\mathrm{in} = 200\,\mathrm{nm}$. Similarly, the final force f_fin may be calculated for $n = n_\mathrm{tot}$ and $Z_\mathrm{fin} = 500\,\mathrm{nm}$. The work due to stretching of the ssDNA is given by [72]

$$W_{u_l} = f_\mathrm{fin}\, u_l(f_\mathrm{fin}) - \int_0^{f_\mathrm{fin}} df\, u_l(f). \tag{5.8}$$

According to [117], we can write for the reversible work that is needed to stretch the handles and to displace the bead in the trap

$$W_{yh} = \frac{f_\mathrm{fin}^2 - f_\mathrm{in}^2}{2\,k_\mathrm{eff}}. \tag{5.9}$$

Summing up, the sequence-dependent stretching energy that we have to subtract is

$$G_\mathrm{str}^\mathrm{sim} = G_\mathrm{str}^\mathrm{sim}(n_\mathrm{tot}, Z_\mathrm{fin}) - G_\mathrm{str}^\mathrm{sim}(0, Z_\mathrm{in}) = W_{u_l} + W_{yh}. \tag{5.10}$$

Finally, we can extract ΔG_0 out of the Jarzynski estimator via [117, 2]

$$\Delta G_0 = \Delta G_N - G_{\mathrm{str}}^{\mathrm{sim}} . \tag{5.11}$$

Analogously, the improved free energy of formation ΔG_0^* can be extracted when the improved Jarzynski estimator ΔG_N^* is made use of [128].

5.3.2 General results

An estimation of the JE via the unidirectional variance estimator [163, 55] yields in case of a Gaussian WPD

$$\Delta G_\sigma = \langle W \rangle - \sigma^2/2 , \tag{5.12}$$

where we used the average work $\langle W \rangle$ of N simulated curves and the respective standard deviation σ. We checked this variance estimator for a large number of curves N at a very low loading rate of $r = 0.01\,\mathrm{pN/s}$ for all hairpins. Note that this is experimentally not possible due to disturbing drift effects which are closely related to the extremely large time which is needed to record a single trajectory [72]. However, the larger the end-loop, the lower the loading rate should be in order to reach equilibrium. Therefore, we get deviations of ΔG_N from ΔG_σ of up to $1\,\mathrm{k_B T}$, which decrease further when the loading rate is lowered. As expected, the variance estimator ΔG_σ yields very good results at $r = 0.01\,\mathrm{pN/s}$ for hairpins A and B, which have a small end-loop. In particular, for hairpin A even for a low number of curves and for $r = 16\,\mathrm{pN/s}$ as in the experiments of [128], the variance estimator yields very good results for ΔG_N.

Now we can subtract the deterministic parts $G_{\mathrm{str}}^{\mathrm{sim}}$ and compare ΔG_0 to $\Delta G_0^{\mathrm{true}}$, the true value of the free energy of formation for each hairpin, namely $G_{\mathrm{form}}(n_{\mathrm{tot}}) - G_{\mathrm{form}}(0)$ (see sec. 2.6.1), at the loading rates used in the experiments of [128]. In agreement with the results of these experiments, for hairpin A and B the deviation of ΔG_0 from $\Delta G_0^{\mathrm{true}}$ is small, i. e. below $2\,\mathrm{k_B T}$, already for a low number N of simulated curves and decreases further at higher N. For the hairpins with larger end-loops the deviation increases and culminates in a deviation of typically $17\,\mathrm{k_B T}$ for hairpin E, again similar to the experiments. For $N = 10^6$ this deviation can be reduced by $4\,\mathrm{k_B T}$ and thus remains still far away from equality.

5.3.3 Improving the results for the highly dissipative case

For hairpin E which is furthest from equilibrium we apply exemplarily a correction by improving the Jarzynski estimator (see eq. (5.7)) [128]. In this case the Gaussian approximation of the left tail which may be used for the hairpins with smaller end-loops may no longer be applied. This makes the proper extraction of the free energy of formation quite sophisticated for hairpin E. In [128] additional information of the reverse process was used in order to extract the fit parameters of the left tail well (see eq. (2) in [128]). Here, we truly focus on the unidirectional

case for which the use of the Jarzynski estimator is required, because the Crooks fluctuation theorem [31] may not be applied (see sec. 1.2). Since we know the true value of the free energy of formation ΔG_0^{true}, we can establish a procedure of how to extract the improved free energy of formation ΔG_0^* well even for a small number of measured curves.

In a first step, we concentrate on the 10^6 simulated curves which allow us to perform a good fit of the left tail. As already explained in sec. 5.2.3, we refrain from using equidistant binning since especially in the tails the data points are less and less significant. Instead, we opt for the adapted binning, where the number of entries per bin should be quite small. In the following we work with five curves per bin throughout. Before we give further details about how we proceeded, we provide the recipe suggested in [128] including our supplements.

1. Use the N measured or simulated work values W to compute
   (a) the Jarzynski estimator ΔG_N (see eq. (5.6))
   (b) the average work $\langle W \rangle$
   (c) the standard deviation σ
   (d) the work histogram, using preferably the adapted binning method described in sec. 5.2.3 with a low number of curves per bin, e. g. 5.
2. Use the following shape parameters as input values for the fit according to eq. (2) in [128]
   (a) $\alpha = 0$
   (b) the normalisation constant $q = \Pi^{-1/2}$
   (c) δ should lie in the range of 1 to 2
   (d) the tail width $\Omega = \sqrt{2}\,\sigma$ (see 1c)
   (e) a characteristic work value $W_c = \langle W \rangle$ (see 1b).
3. Compute the improved Jarzynski estimator ΔG_N^* (see eq. (5.7)) using for the calculation of $\langle B_N \rangle$ either eqs. (5) or (6) given in [128], according to the respective value of λ (cf. eq. (3) in [128]).
4. Verify the result graphically by plotting the work histogram and the fit to the left tail $W \ll W_c$ (see eq. (2) in [128]) in a log $p(W)$ plot, where $p(W)$ is the work distribution.
5. Calculate ΔG_0^* using eq. (5.11) and compare with the possibly known values of ΔG_0^{true}.

It proved useful to carry out the fit according to eq. (2) in [128] in a range of 2σ, starting from the leftmost data point of the work histogram. In our case we fit about

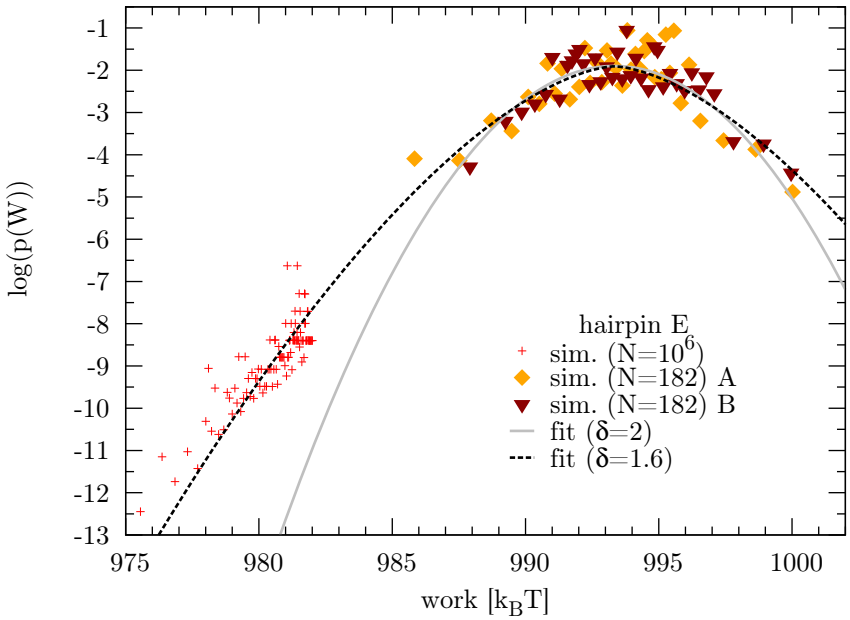


Figure 5.5: Simulated work histograms for hairpin E. The left tail of the simulation for $N = 10^6$ (red crosses) is plotted together with two simulated sets for $N = 182$ (filled diamonds and triangles). In addition, the tail behaviour is given for $\delta = 2$ and 1.6 by a grey solid and a black dashed line, respectively.

50 data points using $\sigma = 2.66$ by means of the Levenberg-Marquardt algorithm [133]. We found a pronounced dependence of the fit results on the values of δ and Ω (see 2c and 2d). Due to the marginal influence of the other shape parameters, we will concentrate on δ and Ω in the following. We recovered the best result with less than 0.01 k_BT deviation of ΔG_0^* from ΔG_0^{true} at values close to $\Omega = \sqrt{2}\,\sigma = 3.77$ and $\delta = 1.60$. To be precise, $\Omega = 3.80$ and $\delta = 1.598$ yielded this best result. For this case where a large number N of curves is available, the standard deviation proves to be a successful estimation for $\Omega = \sqrt{2}\,\sigma$. In order to find an appropriate value of δ even when no reference value ΔG_0^{true} is known, we suggest to perform fits for different start values of δ, e. g. 2.0, 1.9, 1.8 and so on, and to verify graphically the fit result (cp. fig. 5.5). The optically most convincing fit yields the best result for ΔG_0^* for large N of far less than 1 k_BT deviation.

In a second step, we combine the knowledge obtained from $N = 10^6$ surrogate curves with the data of $N = 182$ trajectories (see tab. 5.2) in order to improve the extraction of the free energy of formation. In fig. 5.5 we demonstrate that the behaviour of the left tail can be deduced from the $N = 10^6$ simulated trajectories (red crosses). The tail behaviour of eq. (2) in [128] is shown for $\delta = 2$ and 1.6 using

$\Omega = \sqrt{2}\,\sigma$ by a grey solid and a black dashed line, respectively. However, it is not evident how the left tail continues regarding the two sets of simulated data with $N = 182$ (filled diamonds and triangles). For such a low number of curves it is thus difficult to estimate especially the parameter δ sufficiently well (see the discussion in [128]). Furthermore, it is difficult to decide which fitting range should be used. On the one hand one should respect $W \ll W_c$, on the other hand one needs a few data points in order to be able to perform a fit. We thus found a fit range of less than 2σ most appropriate in this case, depending on the characteristics of the respective work histogram. It is helpful to start the fit with the values which we recovered for Ω and δ as best results for $N = 10^6$. Then the deviation of ΔG_0^* from ΔG_0^{true} lies typically within $1.5\,k_BT$. Even better results may be obtained by directly using the best shape parameters of $N = 10^6$ for the calculation of $\langle B_N \rangle$ without any further fit, where we usually get a deviation below $1\,k_BT$. Such a deviation lies much below the typical free energy of formation of one bp which is $2.5\,k_BT$ and is consistent with the best results for hairpin E obtained with the help of the reverse process in [128].

Since the fitting of a low number of curves N is highly error-prone as explained above, we propose to combine the analysis of highly dissipative experimental data with the simulation of a high number of curves N. Like this, the shape parameters of the fitting function may be precisely estimated and the free energy of formation well extracted. We can thus confirm the method to recover an improved Jarzynski estimator developed by Palassini and Ritort [128]. Including our complements based on the simulation of a huge set of surrogate data, a good estimation of the free energy of formation is possible even in cases far from equilibrium that are difficult to access in unidirectional processes.

Chapter 6

Conclusions and future lines of research

Due to new developments in nanoscience and nanotechnology in recent years, single-molecule techniques may nowadays be considered as standard. They complement data obtained by bulk experiments and reveal additional information about the processes occuring in biomolecules, such as the nucleic acids DNA and RNA, or proteins, including enzymes and molecular motors.

This work focusses on OTs which are a versatile tool to manipulate tiny objects such as DNA hairpin molecules. With the aid of this device the central dogma of molecular biology [30] can be scrutinised by, for example, investigating the DNA replication [17]. This unzipping mechanism was studied here in terms of a force-induced process both experimentally and theoretically, including kinetic and thermodynmic aspects.

An important field of biophysical studies deals with junctions in molecules, because they facilitate manifold ways to interact with other substances, such as cations. Especially interesting in this regard are three-way junctions since metal ions, for example magnesium [97, 125, 105, 28, 60, 57, 92, 10], can bind to them and alter the tertiary structure. In ch. 3 we presented a first step of such an investigation, where we analysed a molecule with a three-way junction alone, without *in vivo* relevant substances.

One of our goals was to check if both the construction and the kinetics of more complex DNA molecules with a richer folding-unfolding behaviour can be characterised by appropriate extensions of theories, which have been developed for two-state systems and proven to work successfully. Our findings indicate that this is indeed possible, at least for the unfolding curves. Therefore, a triple-branch molecule has been specifically designed, yielding a four-state system on the basis of a common FEL model. Due to the successful design, we were able to prove the existence of these four conformational states by the emergence of corresponding force rips in pulling experiments. Besides, the first rupture forces have been sys-

tematically measured and their distributions have been computed. A transition rate theory based on the FEL successfully describes these distributions.

We detected two patterns in the recorded unfolding trajectories: one indicating the expected unfolding behaviour and the other one pointing to the occurrence of irreversible molecular fraying. Such a characterisation was enabled by a connection of the extracted force jump values to the change in the number of opened bps at each transition. In the course of this estimation, we tested the validity of two polymer models (FJC and WLC) for the elastic response of the ssDNA as well as several sets of parameters to get the best agreement with the expected values. Such kind of analysis is helpful to compare the elastic properties measured in DNA unzipping experiments with those obtained by stretching ssDNA polymers [70].

For one type of molecules a smaller force than expected was needed to unfold the stem, because some bps at its basis were partly or completely melted due to photodamaging. Permanent molecular fraying is a normally unwanted, but frequently occuring effect in SMEs and deserves special attention so that its distorting influence on experimental results can be reduced. It is thus necessary to find ways to avoid irreversible fraying, since not completely closed molecules change the measured unfolding-folding trajectories. For this reason an average over all molecules, including permanently frayed ones, can lead to vast deviations of the real values and hence to fairly misinterpreted results. In order to improve the statistics of the results, it is thus required to detect irreversibly frayed molecules and to remove them from the analysis. In ch. 3 we worked out a useful method to identify permanent molecular fraying. Using appropriate parameters for the polymer models, one can estimate the number of irreversibly frayed bps at the basis of the stem. Further experiments performed with silica instead of polystyrene beads may clarify under which conditions permanent molecular fraying can be reduced.

Due to the possibility of distorting curves in the experimental data, as, for example, due to permanent molecular fraying, human supervision is needed during the analysis of FDCs. In view of refined experiments yielding a higher number of data, an automatic analysis will be of immeasurable value.

All this knowledge paves the way to further interesting studies such as the folding problem [84, 164, 103, 21]. Short hairpins with canonical base pairing reveal typically a two-state behaviour and fold cooperatively [25, 106, 117]. Nevertheless, an entropic correction term could be required in order to capture the full kinetics of the folding process [103]. In longer molecules with several conformational states, such a refined analysis might be even more important and should be addressed in a future work. Additional research on the folding problem may deepen our knowledge about diseases connected to misfolding [164].

Furthermore, in a similar study as carried out here, the binding of metal ions to the three-way junctions could be examined in order to investigate structural changes due to the formation of tertiary contacts and therefore altered kinetics of the unfolding process [105].

The translocation motion of helicases that unwind dsDNA before the replication process can start could be addressed as another interesting subject, including

the investigation of how they move along bifurcation points. Besides, helicases might help to improve the DNA unzipping resolution to one bp via direct force application at the fork [72]. While in MTs a resolution of one to two nucleotides has already been achieved [124], up to date, it is a challenging problem to increase the precision of OTs in order to resolve less then ten bps [13, 137, 72]. Apart from purely technical limitations, the fluctuations in the measured data impede to extract reliable information from DNA sequencing [137]. These noise fluctuations can be reduced by increasing the trap stiffness [54], by using very rigid handles [51] and by removing the soft ssDNA released during unzipping from the fork, where the force should be applied locally [72]. In addition, an enhanced detection resolution is expected for the application of two optical traps [112] or a combination of OTs with single-molecule fluorescence techniques [63, 61].

Eventually, the kinetic approach for the prediction of the probability distributions of the first rupture forces could be a well-suited starting point for future in-depth modelling in this domain on a more microscopic basis.

In a second part (see ch. 4), we showed that the FEL for different types of DNA hairpins with periodic base sequences changes regularly when an external mechanical force f is applied. Such a regular change manifests itself in the forward and backward barriers between consecutive states, which are decreasing linearly with f over a large range of forces, including the regime where the molecules unfold entirely. With the help of KMC simulations, stochastic unfolding trajectories were generated as surrogate for experimental curves. In order to identify the energetic parameters of the FEL, we introduce the probability $\Psi_0(t)$ that a molecule remains in the folded state until time t and the probability $\Psi_1(t)$ that until time t a molecule undergoes at least one forward, but no backward transition. Theoretical calculations yield straightforward expressions for these two probabilities, which can be sampled from the trajectories. Like this the FEL parameters can be determined by a nonlinear fitting procedure. We showed that already about 300 unfolding trajectories, as typically measured in SMEs [52, 123], are sufficient to obtain good parameter estimates. Best results can be obtained when the loading rate is well chosen, that is to say if the long-time limit of $\Psi_1(t)$ is neither close to zero nor close to one. For small barriers which are only a few $k_\mathrm{B}T$ high, the loading rate may become too high to be realisable in experiment.

From a theoretical point of view, the possibility to generate simple functional forms of FELs, depending on the number of open bps and the applied force, provides an interesting basis for future research, where exact results could be obtained for kinetic and energetic properties. Analytical results for WPDs will be particularly interesting to gain a deeper knowledge about tail regimes.

In view of future studies, it is important to note that the treatment presented in this work is limited to unzipping at low temperatures below the denaturation transition of the DNA hairpin molecules, where bubble formation can be safely neglected. In the presence of bubble formation at higher temperatures, the approach we used here on the basis of the standard nearest neighbour model becomes less

appropriate [171]. Instead, the Peyrard-Bishop-Dauxois (PBD) model [130, 35] was shown to provide a better description [156] and has been further refined by including a sequence-dependent stacking term [5]. However, a generalisation of the treatment applying the PBD model, with the aim to account for bubble formation, would also require to go beyond a modelling based on activated transition rate theory which we used.

Concerning possible applications, our results provide a promising means to obtain improved values of the free energies of interior loops, e. g. by use of OTs or MTs. Periodic sequences can be synthesised and analysed by attaching loops at regular intervals to homogeneous double strands. Good counting statistics should be achieved when using several periods.

In a third part (see ch. 5), we modelled typical pulling experiments performed with OTs by applying a common model for the FEL using the description of a DP. Taking into account all possible intermediates in the unfolding kinetics, we found the tendency of a two-state behaviour. With the help of KMC simulations, unfolding trajectories in form of force-distance curves were generated, serving as surrogate data for experimental findings. Our simulated work data was comparable to the experimental data of five DNA hairpin molecules with different length of the end-loop [128] for small values of the attempt frequency ν, when we simulated the same number of curves as were recorded in the experiments. This behaviour was confirmed by a simulation of $N = 10^6$ trajectories.

It is an open question why relatively low ν values are required to fit the simulated results to the experimental ones. Maybe phenomenological Bell-Evans kinetic models, as we applied in our KMC simulation, can not capture all details of the underlying process and thus fail to describe these experiments properly [74]. In addition, a different type of coarse-grained model for DNA could be used [114, 126, 127]. Molecular dynamics simulations might provide a complementary access to this topic [135, 134, 147, 148, 53].

A systematic research on kinetic parameters on the level of single molecules has not been executed yet [94]. For instance, the attempt frequency has so far been only partially studied (see e. g. [25, 42]). Using single-molecule force clamp spectroscopy, this important property has been recently extracted from SMEs for the first time [94] and complemented by [132]. From a theoretical point of view, entropic factors [103] might play an important role for the value of the attempt frequency, which are not thermally activated.

In addition, the simulated results were verified by extracting FEDs by means of the Jarzynski equality. Including many technical subtleties in the simulation, the nearest neighbour model seems to be appropriate to generate surrogate data and we obtain good agreement with the already known values of the free energy of formation for the Gaussian case. When the work histograms may no longer be considered Gaussian, an experimentally and even computationally unfeasible high number of curves is required in order to estimate the FEDs well in case of unidirectional processes (see sec. 1.2). Even when using the free energy recovery method of [128],

it is hard to estimate the fit parameters reasonably in these highly dissipative cases where the reverse process is not accessable. Our knowledge about the left tail of the work probability distribution based on a high number of simulated surrogate data allows us to complement the recipe of [128] in order to improve the extracted FEDs when few data is available. Even for less than 200 curves, such a combined analysis method, using the precisely estimated fitting values obtained by dint of a high number of simulated curves, yields deviations lying usually within $1\,k_BT$ and thus much below the typical free energy of formation of one bp ($2.5\,k_BT$).

The simulation is based on already known values of the interaction energy used in the nearest neighbour model. They could be improved iteratively by means of the experimental data. Like that, it could be possible to extract improved loop terms out of measured work data, e. g. for end-loops or interior loops (cp. sec. 4), for which less accurate values than for the bp interactions are available [70]. For example, one could investigate one stem attached to different end-loops, similar to the approach of [128]. Analogously, our simulation might serve as a tool to produce arbitrarily large data sets which can be used to test and improve common analysis methods in other contexts where it is a challenging task to aquire a high number of experimental data.

Bibliography

[1] Anna Alemany. Free energy landscapes in DNA hairpins - An approach to the concept of molecular fragility. Biophysics Master Project, Universitat de Barcelona, January 2009.

[2] Anna Alemany, Alessandro Mossa, Ivan Junier, and Felix Ritort. Experimental free-energy measurements of kinetic molecular states using fluctuation theorems. Nat. Phys., 8:688–694, 2012.

[3] Anna Alemany, Marco Ribezzi-Crivellari, and Felix Ritort. Recent progress in fluctuation theorems and free energy recovery. AIP Conf. Proc., 1332(1):96–110, 2011. We also used the version on arXiv: 1101.3174.

[4] Andrea Alessandrini and Paolo Facci. AFM: a versatile tool in biophysics. Meas. Sci. Technol., 16:R65 – R92, 2005.

[5] Boian S. Alexandrov, Vladimir Gelev, Yevgeniya Monisova, Ludmil B. Alexandrov, Alan R. Bishop, Kim O. Rasmussen, and Anny Usheva. A nonlinear dynamic model of DNA with a sequence-dependent stacking term. Nucleic Acids Res., 37(7):2405, 2009.

[6] Alessandro Ammenti, Fabio Cecconi, Umberto M. B. Marconi, and Angelo Vulpiani. A statistical model for translocation of structured polypeptide chains through nanopores. J. Phys. Chem. B, 113(30):10348–10356, 2009.

[7] Valentina Baldazzi, Simona Cocco, Enzo Marinari, and Remi Monasson. Inference of DNA sequences from mechanical unzipping: an ideal-case study. Phys. Rev. Lett., 96(12):128102, 2006.

[8] Graeme I. Bell. Models for the specific adhesion of cells to cells. Science, 200(4342):618–627, 1978.

[9] Charles H. Bennett. Efficient estimation of free-energy differences from Monte Carlo data. J. Comp. Phys., 22:245–268, 1976.

[10] Cristiano V. Bizarro, Anna Alemany, and Felix Ritort. Non-specific binding of Na(+) and Mg(2+) to RNA determined by force spectroscopy methods. Nucleic Acids Res., 40(14):6922–6935, 2012.

[11] Valentin Blickle, Thomas Speck, Laurent Helden, Udo Seifert, and Clemens Bechinger. Thermodynamics of a colloidal particle in a time-dependent non-harmonic potential. Phys. Rev. Lett., 96(7):070603, 2006.

[12] Steven M. Block, Charles L. Asbury, Joshua W. Shaevitz, and Matthew J. Lang. Probing the kinesin reaction cycle with a 2D optical force clamp. Proc. Natl. Acad. Sci., 100(5):2351–2356, 2003.

[13] Ulrich Bockelmann, Philippe Thomen, B. Essevaz-Roulet, Virgile Viasnoff, and Francois Heslot. Unzipping DNA with optical tweezers: high sequence sensitivity and force flips. Biophys. J., 82(3):1537–1553, 2002.

[14] Alessandro Bosco, Joan Camunas-Soler, and Felix Ritort. Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. Nucleic Acids Res., 42(3):2064–2074, 2014.

[15] Oliver Braun, Andreas Hanke, and Udo Seifert. Probing molecular free energy landscapes by periodic loading. Phys. Rev. Lett., 93(15):158105, 2004.

[16] Jose Javier Brey and Jose Antonio Prados. Residual properties of a two-level system. Phys. Rev. B, 43(10):8350–8361, 1991.

[17] Carlos Bustamante, Wei Cheng, and Yara X. Mejia. Revisiting the central dogma one molecule at a time. Cell, 144(4):480–497, 2011.

[18] Carlos Bustamante, Jan Liphardt, and Felix Ritort. The nonequilibrium thermodynamics of small systems. Phys. Today, 58:43–48, 2005. We used the long version on arXiv:cond-mat/0511629.

[19] Carlos Bustamante, John F. Marko, Eric D. Siggia, and Steven B. Smith. Entropic elasticity of lambda-phage DNA. Science, 265(5178):1599–1600, 1994.

[20] Carlos Bustamante, Steven B. Smith, Jan Liphardt, and Doug Smith. Single-molecule studies of DNA mechanics. Curr. Opin. Struct. Biol., 10(3):279–285, 2000.

[21] Song Cao and Shi-Jie Chen. A new computational approach for mechanical folding kinetics of RNA hairpins. Biophys. J., 96:4024–4034, 2009.

[22] Nick J. Carter and Robert A. Cross. Kinesin's moonwalk. Curr. Opin. Cell Biol., 18(1):61–67, 2006.

[23] Wuen-Shiu Chen, Wei-Hung Chen, Zephan Chen, Ashton A. Gooding, Kuan-Jiuh Lin, and Ching-Hwa Kiang. Direct observation of multiple pathways of single-stranded DNA stretching. Phys. Rev. Lett., 105(21):218104, 2010.

[24] Chi-Han Chiou, Yu-Yen Huang, Meng-Han Chiang, Huei-Huang Lee, and Gwo-Bin Lee. New magnetic tweezers for investigation of the mechanical properties of single DNA molecules. Nanotechnology, 17:1217–1224, 2006.

[25] Simona Cocco, John F. Marko, and Remi Monasson. Slow nucleic acid unzipping kinetics from sequence-defined barriers. Eur. Phys. J. E, 10(2):153–161, 2003.

[26] Simona Cocco, Remi Monasson, and John F. Marko. Force and kinetic barriers to unzipping of the DNA double helix. Proc. Natl. Acad. Sci., 98(15):8608–8613, 2001.

[27] Simona Cocco, Remi Monasson, and John F. Marko. Force and kinetic barriers to initiation of DNA unzipping. Phys. Rev. E, 65:041907, 2002.

[28] Delphine Collin, Felix Ritort, Christopher Jarzynski, Steven B. Smith, Ignacio Tinoco Jr., and Carlos Bustamante. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. Nature, 437(7056):231–234, 2005.

[29] Richard S. Conroy and Claudia Danilowicz. Unravelling DNA. Contemp. Phys., 45(4):277–302, 2004.

[30] Francis Crick. Central dogma of molecular biology. Nature, 227(5258):561–563, 1970.

[31] Gavin E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys. Rev. E, 60(3):2721–2726, 1999.

[32] Donald M. Crothers and Bruno H. Zimm. Theory of the melting transition of synthetic polynucleotides: evaluation of the stacking free energy. J. Mol. Biol., 9:1–9, 1964.

[33] Claudia Danilowicz, Vincent W. Coljee, Cedric Bouzigues, David K. Lubensky, David R. Nelson, and Mara Prentiss. DNA unzipped under a constant force exhibits multiple metastable intermediates. Proc. Natl. Acad. Sci., 100(4):1694–1699, 2003.

[34] Xavier Daura, Roman Affentranger, and Alan E. Mark. On the relative merits of equilibrium and non-equilibrium simulations for the estimation of free-energy differences. ChemPhysChem, 11(17):3734–3743, 2010.

[35] Thierry Dauxois, Michel Peyrard, and Alan R. Bishop. Entropy-driven DNA denaturation. Phys. Rev. E, 47(1):R44, 1993.

[36] Michel de Messieres, Barbara Brawn-Cinani, and Arthur La Porta. Measuring the folding landscape of a harmonically constrained biopolymer. Biophys. J., 100:2736–2744, 2011.

[37] Ashok A. Deniz, Samrat Mukhopadhyay, and Edward A. Lemke. Single-molecule biophysics: at the interface of biology, physics and chemistry. J. R. Soc. Interface, 5(18):15–45, 2008.

[38] Marie-Noelle Dessinges, Berenike Maier, Y. Zhang, M. Peliti, David Bensimon, and Vincent Croquette. Stretching single stranded DNA, a model polyelectrolyte. Phys. Rev. Lett., 89(24):248102, 2002.

[39] Howard DeVoe and Ignacio Tinoco Jr. The stability of helical polynucleotides: base contributions. J. Mol. Biol., 4:500–517, 1962.

[40] Abhishek Dhar. Work distribution functions in polymer stretching experiments. Phys. Rev. E., 71:036126, 2005.

[41] Fangyuan Ding, Maria Manosas, Michelle M. Spiering, Stephen J. Benkovic, David Bensimon, Jean-Francois Allemand, and Vincent Croquette. Single-molecule mechanical identification and sequencing. Nat. Methods, 9(4):367–372, 2012.

[42] Olga K. Dudko, Gerhard Hummer, and Attila Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. Phys. Rev. Lett., 96(10):108101, 2006.

[43] Sophie Dumont, Wei Cheng, Victor Serebrov, Rudolf K. Beran, Ignacio Tinoco Jr., Anna Marie Pyle, and Carlos Bustamante. RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. Nature, 439(7072):105–108, 2006.

[44] Mario Einax and Philipp Maass. Work distributions for ising chains in a time-dependent magnetic field. Phys. Rev. E, 80:020101, 2009.

[45] Andreas Engel. Asymptotics of work distributions in nonequilibrium systems. Phys. Rev. E, 80:021120, 2009.

[46] Sandra Engel. Folding and unfolding of a triple-branch DNA hairpin molecule with four conformational states by mechanical stretching. Diplomarbeit, Technische Universitaet Ilmenau, January 2010.

[47] Massimiliano Esposito and Christian Van den Broeck. Three detailed fluctuation theorems. Phys. Rev. Lett., 104(9):090601, 2010.

[48] Furqan M. Fazal and Steven M. Block. Optical tweezers study life under tension. Nat. Photonics, 5:318–321, 2011.

[49] Jay K. Fisher, Jeremy Cribb, Kalpit V. Desai, Leandra Vicci, Benjamin Wilde, Kurtis Keller, Russell M. Taylor, Julian Haase, Kerry Bloom, E. Timothy O'Brien, and Richard Superfine. Thin-foil magnetic force system for high-numerical-aperture microscopy. Rev. Sci. Instrum., 77(2):023702, 2006.

[50] Jay K. Fisher, Leandra Vicci, Jeremy Cribb, E. Timothy O'Brien, Russell M. Taylor, and Richard Superfine. Magnetic force micromanipulation systems for the biological sciences. Nano, 1(3):191–205, 2006.

[51] Nuria Forns, Sara de Lorenzo, Maria Manosas, Kumiko Hayashi, Josep Maria Huguet, and Felix Ritort. Improving signal/noise resolution in single-molecule experiments using molecular constructs with short handles. Biophys. J., 100(7):1765–1774, 2011.

[52] Ronald F. Fox. Using nonequilibrium measurements to determine macromolecule free-energy differences. Proc. Natl. Acad. Sci., 100(22):12537–12538, 2003.

[53] Ignacio Franco, Mark A. Ratner, George C. Schatz. "Single molecule pulling: phenomenology, and interpretation". In: Nano and cell mechanics: Fundamentals and frontiers. Edited by H. D. Espinosa and G. Bao. Wiley, chap. 14: 359–388, 2013.

[54] Ulrich Gerland, Ralf Bundschuh, and Terence Hwa. Mechanically probing the folding pathway of single RNA molecules. Biophys. J., 84(5):2831–2840, 2003.

[55] Jeff Gore, Felix Ritort, and Carlos Bustamante. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. Proc. Natl. Acad. Sci., 100(22):12564–12569, 2003.

[56] Charlie Gosse and Vincent Croquette. Magnetic tweezers: micromanipulation and force measurement at the molecular level. Biophys. J., 82(6):3314–3329, 2002.

[57] Lisa Green, Chul-Hyun Kim, Carlos Bustamante, and Ignacio Tinoco Jr. Characterization of the mechanical unfolding of RNA pseudoknots. J. Mol. Biol., 375(2):511–528, 2008.

[58] William J. Greenleaf, Michael T. Woodside, Elio A. Abbondanzieri, and Steven M. Block. Passive all-optical force clamp for high-resolution laser trapping. Phys. Rev. Lett., 95(20):208102, 2005.

[59] William J. Greenleaf, Michael T. Woodside, and Steven M. Block. High-resolution, single-molecule measurements of biomolecular motion. Annu. Rev. Biophys. Biomol. Struct., 36:171–190, 2007.

[60] Dan Grilley, Vinod Misra, Gokhan Caliskan, and David E. Draper. Importance of partially unfolded conformations for Mg(2+)-induced folding of RNA tertiary structure: structural models and free energies of Mg(2+) interactions. Biochemistry, 46(36):10266–10278, 2007.

[61] Peter Gross, Geraldine Farge, Erwin J. G. Peterman, and Gijs J. L. Wuite. Combining optical tweezers, single-molecule fluorescence microscopy, and microfluidics for studies of DNA-protein interactions. Methods Enzymol., 475:427–453, 2010.

[62] Peter Gross, Niels Laurens, Lene B. Oddershede, Ulrich Bockelmann, Erwin J. G. Peterman, and Gijs J. L.Wuite. Quantifying how DNA stretches, melts and changes twist under tension. Nat. Phys., 7:731–736, 2011.

[63] Taekjip Ha. Single-molecule fluorescence methods for the study of nucleic acids. Curr. Opin. Struct. Biol., 11(3):287–292, 2001.

[64] Nolan C. Harris, Yang Song, and Ching-Hwa Kiang. Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality. Phys. Rev. Lett., 99(6):068101, 2007.

[65] Iddo Heller, Tjalle P. Hoekstra, Graeme A. King, Erwin J. G. Peterman, and Gijs J. L. Wuite. Optical tweezers analysis of DNA-protein complexes. Chem. Rev., 114(6):3087–3119, 2014.

[66] Viktor Holubec, Petr Chvosta, Mario Einax, and Philipp Maass. Attempt time Monte Carlo: An alternative for simulation of stochastic jump processes with time-dependent transition rates. Europhy. Lett., 93:40003, 2011.

[67] Viktor Holubec, Petr Chvosta, and Philipp Maass. Dynamics and energetics for a molecular zipper model under external driving. J. Stat. Mech., P11009, 2012.

[68] Silvia Hormeno and J. Ricardo Arias-Gonzalez. Exploring mechanochemical processes in the cell with optical tweezers. Biol. Cell, 98(12):679–695, 2006.

[69] Thorsten Hugel and Markus Seitz. The study of molecular interactions by AFM force spectroscopy. Macromol. Rapid Commun., 22(13):989–1016, 2001.

[70] Josep Maria Huguet, Cristiano V. Bizarro, Nuria Forns, Steven B. Smith, Carlos Bustamante, and Felix Ritort. Single-molecule derivation of salt dependent base-pair free energies in DNA. Proc. Natl. Acad. Sci., 107(35):15431–15436, 2010.

[71] Josep Maria Huguet, Nuria Forns, and Felix Ritort. Statistical properties of metastable intermediates in DNA unzipping. Phys. Rev. Lett., 103(24):248106, 2009.

[72] Josep Maria Huguet i Casades. Statistical and thermodynamic properties of DNA unzipping experiments with optical tweezers. PhD thesis, Universitat de Barcelona, October 2010.

[73] Gerhard Hummer and Attila Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. Proc. Natl. Acad. Sci., 98(7):3658–3661, 2001.

[74] Gerhard Hummer and Attila Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. Biophys. J., 85(1):5–15, 2003.

[75] Gerhard Hummer and Attila Szabo. Free energy surfaces from single-molecule force spectroscopy. Acc. Chem. Res., 38(7):504–513, 2005.

[76] Alberto Imparato and Luca Peliti. Kinetic barriers in RNA unzipping. Eur. Phys. J. B, 39:357–363, 2004.

[77] Alberto Imparato and Luca Peliti. Work probability distribution in single-molecule experiments. Europhys. Lett., 69(4):643, 2005.

[78] Alberto Imparato and Luca Peliti. Work-probability distribution in systems driven out of equilibrium. Phys. Rev. E, 72(4 Pt 2):046114, 2005.

[79] Andreas Janshoff, Marcus Neitzert, York Oberdoerfer, and Harald Fuchs. Force spectroscopy of molecular systems - Single molecule spectroscopy of polymers and biomolecules. Angew. Chem., 39(18):3212–3237, 2000.

[80] Christopher Jarzynski. Nonequilibrium equality for free energy differences. Phys. Rev. Lett., 78:2690–2693, 1997.

[81] Christopher Jarzynski. Rare events and the convergence of exponentially averaged work values. Phys. Rev. E, 73:046105, 2006.

[82] Christopher Jarzynski. Nonequilibrium work relations: foundations and applications. Eur. Phys. J. B, 64:331–340, 2008.

[83] Ivan Junier, Alessandro Mossa, Maria Manosas, and Felix Ritort. Recovery of free energy branches in single molecule experiments. Phys. Rev. Lett., 102(7):070602, 2009.

[84] Jiho Kim, Soeren Doose, Hannes Neuweiler, and Markus Sauer. The initial step of DNA hairpin folding: a kinetic analysis using fluorescence correlation spectroscopy. Nucleic Acids Res., 34(9):2516–2527, 2006.

[85] Steven J. Koch, Alla Shundrovsky, Benjamin C. Jantzen, and Michelle D. Wang. Probing protein-DNA interactions by unzipping a single DNA double helix. Biophys. J., 83(2):1098–1105, 2002.

[86] Rupert Krautbauer, Matthias Rief, and Hermann Gaub. Unzipping DNA oligomers. Nano Lett., 3(4):493–496, 2003.

[87] Sanjay Kumar and Mai Suan Li. Biomolecules under mechanical force. Phys. Rep., 486:1–74, 2010.

[88] Anupam Kundu, Sanjib Sabhapandit, and Abhishek Dhar. Application of importance sampling to the computation of large deviations in nonequilibrium processes. Phys. Rev. E, 83(3 Pt 1):031119, 2011.

[89] Markita P. Landry, Patrick M. McCall, Zhi Qi, and Yann R. Chemla. Characterization of photoactivated singlet oxygen damage in single-molecule optical trap experiments. Biophys. J., 97(8):2128–2136, 2009.

[90] Matthew J. Lang, Charles L. Asbury, Joshua W. Shaevitz, and Steven M. Block. An automated two-dimensional optical force clamp for single molecule studies. Biophys. J., 83(1):491–501, 2002.

[91] Jean-Francois Leger, Jerome Robert, Laurent Bourdieu, Didier Chatenay, and John F. Marko. Reca binding to a single double-stranded DNA molecule: a possible role of DNA conformational fluctuations. Proc. Natl. Acad. Sci., 95(21):12295–12299, 1998.

[92] Desirae Leipply and David E. Draper. Effects of Mg(2+) on the free energy landscape for folding a purine riboswitch RNA. Biochemistry, 50(14):2790–2799, 2011.

[93] Pan T. X. Li, Delphine Collin, Steven B. Smith, Carlos Bustamante, and Ignacio Tinoco Jr. Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. Biophys. J., 90(1):250–260, 2006.

[94] Jian Liang and Julio M. Fernandez. Kinetic measurements on single-molecule disulfide bond cleavage. J. Am. Chem. Soc., 133(10):3528, 2011.

[95] Kaj Ulrik Linderstrøm-Lang. Proteins and Enzymes, volume 6. Lane Medical Lectures. Stanford University Publications, University Series, Medical Sciences, Stanford University Press, 1952.

[96] Jan Liphardt, Sophie Dumont, Steven B. Smith, Ignacio Tinoco Jr., and Carlos Bustamante. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. Science, 296(5574):1832–1835, 2002.

[97] Jan Liphardt, Bibiana Onoa, Steven B. Smith, Ignacio Tinoco Jr., and Carlos Bustamante. Reversible unfolding of single RNA molecules by mechanical force. Science, 292(5517):733–737, 2001.

[98] Fei Liu and Zhong-Can Ou-Yang. Monte Carlo simulation for single RNA unfolding by force. Biophys. J., 88(1):76–84, 2005.

[99] David K. Lubensky and David R. Nelson. Single molecule statistics and the polynucleotide unzipping transition. Phys. Rev. E, 65(3 Pt 1):031917, 2002.

[100] Berenike Maier, David Bensimon, and Vincent Croquette. Replication by a single DNA polymerase of a stretched single-stranded DNA. Proc. Natl. Acad. Sci., 97(22):12002–12007, 2000.

[101] Maria Manosas. Mechanical folding/unfolding of RNA molecules: Experimental facts and theoretical models. PhD thesis, Universitat de Barcelona, February 2007.

[102] Maria Manosas, Delphine Collin, and Felix Ritort. Force-dependent fragility in RNA hairpins. Phys. Rev. Lett., 96(21):218301, 2006.

[103] Maria Manosas, Ivan Junier, and Felix Ritort. Force-induced misfolding in RNA. Phys. Rev. E, 78(6 Pt 1):061925, 2008.

[104] Maria Manosas, Alessandro Mossa, Nuria Forns, Josep Maria Huguet, and Felix Ritort. Dynamic force spectroscopy of DNA hairpins. II. Irreversibility and dissipation. J. Stat. Mech., P02061, 2009.

[105] Maria Manosas and Felix Ritort. Thermodynamic and kinetic aspects of RNA pulling experiments. Biophys. J., 88(5):3224–3242, 2005.

[106] Maria Manosas, Jin-Der Wen, Pan T. X. Li, Steven B. Smith, Carlos Bustamante, Ignacio Tinoco Jr., and Felix Ritort. Force unfolding kinetics of RNA using optical tweezers. II. Modeling experiments. Biophys. J., 92(9):3010–3021, 2007.

[107] Paul Maragakis, Martin Spichty, and Martin Karplus. A differential fluctuation theorem. J. Phys. Chem. B, 112(19):6168–6174, 2008.

[108] John F. Marko and Eric D. Siggia. Stretching DNA. Macromolecules, 28:8759–8770, 1995.

[109] Simone Melchionna, Maria Fyta, Efthimios Kaxiras, and Sauro Succi. Exploring DNA translocation through a nanopore via a multiscale lattice-Boltzmann molecular-dynamics methodology. Int. J. Mod. Phys. C, 18:685, 2007.

[110] Amit Meller, Lucas Nivon, Eric Brandin, Jene Golovchenko, and Daniel Branton. Rapid nanopore discrimination between single polynucleotide molecules. Proc. Natl. Acad. Sci., 97(3):1079–1084, 2000.

[111] Brian Mickey and Jonathon Howard. Rigidity of microtubules is increased by stabilizing agents. J. Cell. Biol., 130(4):909–917, 1995.

[112] Jeffrey R. Moffitt, Yann R. Chemla, David Izhaky, and Carlos Bustamante. Differential detection of dual traps improves the spatial resolution of optical tweezers. Proc. Natl. Acad. Sci., 103(24):9006–9011, 2006.

[113] Jeffrey R. Moffitt, Yann R. Chemla, Steven B. Smith, and Carlos Bustamante. Recent advances in optical tweezers. Annu. Rev. Biochem., 77:205–228, 2008.

[114] Alex Morriss-Andrews, Joerg Rottler, and Steven S. Plotkin. A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality. J. Chem. Phys., 132(3):035105, 2010.

[115] Alessandro Mossa, Sara de Lorenzo, Josep Maria Huguet, and Felix Ritort. Measurement of work in single-molecule pulling experiments. J. Chem. Phys., 130(23):234116, 2009.

[116] Alessandro Mossa, Josep Maria Huguet, and Felix Ritort. Investigating the thermodynamics of small biosystems with optical tweezers. Physica E, 42(3):666–671, 2010.

[117] Alessandro Mossa, Maria Manosas, Nuria Forns, Josep Maria Huguet, and Felix Ritort. Dynamic force spectroscopy of DNA hairpins. I. Force kinetics and free energy landscapes. J. Stat. Mech., P02060, 2009.

[118] Richard A. Neher and Ulrich Gerland. Dynamics of force-induced DNA slippage. Phys. Rev. Lett., 93(19):198102, 2004.

[119] Richard A. Neher and Ulrich Gerland. DNA as a programmable viscoelastic nanoelement. Biophys. J., 89(6):3846–3855, 2005.

[120] Keir C. Neuman, Timothee Lionnet, and Jean-Francois Allemand. Single-molecule micromanipulation techniques. Annu. Rev. Mater. Res., 37:33–67, 2007.

[121] Keir C. Neuman and Attila Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat. Methods, 5(6):491–505, 2008.

[122] Daniel Nickelsen and Andreas Engel. Asymptotics of work distributions: the pre-exponential factor. Eur. Phys. J. B, 82(3-4):207–218, 2011.

[123] Paolo Nicolini, Piero Procacci, and Riccardo Chelli. Hummer and Szabo-like potential of mean force estimator for bidirectional nonequilibrium pulling experiments/simulations. J. Phys. Chem. B, 114(29):9546–9554, 2010.

[124] Piercen M. Oliver, Jin Seon Park, and Dmitri Vezenov. Quantitative high-resolution sensing of DNA hybridization using magnetic tweezers with evanescent illumination. Nanoscale, 3(2):581–591, 2011.

[125] Bibiana Onoa, Sophie Dumont, Jan Liphardt, Steven B. Smith, Ignacio Tinoco Jr., and Carlos Bustamante. Identifying kinetic barriers to mechanical unfolding of the T. thermophila ribozyme. Science, 299(5614):1892–1895, 2003.

[126] Thomas E. Ouldridge, Ard A. Louis, and Jonathan P. K. Doye. DNA nanotweezers studied with a coarse-grained model of DNA. Phys. Rev. Lett., 104(17):178101, 2010.

[127] Thomas E. Ouldridge, Ard A. Louis, and Jonathan P. K. Doye. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. J. Chem. Phys., 134(8):085101, 2011.

[128] Matteo Palassini and Felix Ritort. Improving free-energy estimates from unidirectional work measurements: theory and experiment. Phys. Rev. Lett., 107(6):060601, 2011.

[129] Luca Peliti. On the work-Hamiltonian connection in manipulated systems. J. Stat. Mech., P05002, 2008.

[130] Michel Peyrard and Alan R. Bishop. Statistical mechanics of a nonlinear model for DNA denaturation. Phys. Rev. Lett., 62(23):2755, 1989.

[131] Dietmar Poerschke. Model calculations on the kinetics of oligonucleotide double helix coil transitions. Evidence for a fast chain sliding reaction. Biophys. Chem., 2(2):83–96, 1974.

[132] Ionel Popa, Julio M. Fernandez, and Sergi Garcia-Manyes. Direct quantification of the attempt frequency determining the mechanical unfolding of ubiquitin protein. J. Biol. Chem., 286(36):31072–31079, 2011.

[133] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. Numerical recipes in Fortran 90, volume 2. Cambridge University Press, 2nd edition, 1997.

[134] Piero Procacci and Simone Marsili. Energy dissipation asymmetry in the non equilibrium folding/unfolding of the single molecule alanine decapeptide. Chem. Phys., 375:8–15, 2010.

[135] Piero Procacci, Simone Marsili, Alessandro Barducci, Giorgio F. Signorini, and Riccardo Chelli. Crooks equation for steered molecular dynamics using a Nose-Hoover thermostat. J. Chem. Phys., 125(16):164101, 2006.

[136] Marco Ribezzi-Crivellari, Josep Maria Huguet, and Felix Ritort. Counter-propagating dual-trap optical tweezers based on linear momentum conservation. Rev. Sci. Instrum., 84:043104, 2013.

[137] Felix Ritort. Single-molecule experiments in biological physics: Methods and applications. J. Phys.: Condens. Matter, 18:R531–R583, 2006.

[138] Felix Ritort. The nonequilibrium thermodynamics of small systems. C. R. Physique, 8:528–539, 2007.

[139] Felix Ritort. Nonequilibrium fluctuations in small systems: From physics to biology. Adv. Chem. Phys., 137:31, 2008.

[140] Felix Ritort, Carlos Bustamante, and Ignacio Tinoco Jr. A two-state kinetic model for the unfolding of single molecules by mechanical force. Proc. Natl. Acad. Sci., 99(21):13544–13548, 2002.

[141] Felix Ritort, Shirley Mihardja, Steven B. Smith, and Carlos Bustamante. Condensation transition in DNA-polyaminoamide dendrimer fibers studied using optical tweezers. Phys. Rev. Lett., 96(11):118301, 2006.

[142] Claudio Rivetti, Chip Walker, and Carlos Bustamante. Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. J. Mol. Biol., 280(1):41–59, 1998.

[143] John SantaLucia Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci., 95:1460–1465, 1998.

[144] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. Rep. Prog. Phys., 75(12):126001, 2012.

[145] Michael R. Shirts, Eric Bair, Giles Hooker, and Vijay S. Pande. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. Phys. Rev. Lett., 91(14):140601, 2003.

[146] Robert M. Simmons, Jeffrey T. Finer, Steven Chu, and James A. Spudich. Quantitative measurements of force and displacement using an optical trap. Biophys. J., 70(4):1813–1822, 1996.

[147] Jens Smiatek, Chun Chen, Dongsheng Liu, and Andreas Heuer. Stable conformations of a single stranded deprotonated DNA i-motif. J. Phys. Chem. B, 115(46):13788–13795, 2011.

[148] Jens Smiatek, Daniel Janssen-Mueller, Rudolf Friedrich, and Andreas Heuer. Systematic detection of hidden complexities in the unfolding mechanism of a cytosine-rich DNA strand. Physica A, 394:136–144, 2014.

[149] Douglas E. Smith, Sander J. Tans, Steven B. Smith, Shelley Grimes, Dwight L. Anderson, and Carlos Bustamante. The bacteriophage straight phi29 portal motor can package DNA against a large internal force. Nature, 413(6857):748–752, 2001.

[150] Steven B. Smith, Yujia Cui, and Carlos Bustamante. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. Science, 271(5250):795–799, 1996.

[151] Steven B. Smith, Yujia Cui, and Carlos Bustamante. Optical-trap force transducer that operates by direct measurement of light momentum. Methods in Enzymology, 361:134–162, 2003.

[152] Evěn Šubrt and Petr Chvosta. Exact analysis of work fluctuations in two-level systems. J. Stat. Mech., 2007(09):P09019, 2007.

[153] Srijeeta Talukder, Pinaki Chaudhury, Ralf Metzler, and Suman K Banik. Determining the DNA stability parameters for the breathing dynamics of heterogeneous DNA by stochastic optimization. J. Chem. Phys., 135(16):165103, 2011.

[154] Ignacio Tinoco Jr. Force as a useful variable in reactions: unfolding RNA. Annu. Rev. Biophys. Biomol. Struct., 33:363–385, 2004.

[155] Joost van Mameren, Peter Gross, Geraldine Farge, Pleuni Hooijman, Mauro Modesti, Maria Falkenberg, Gijs J. L. Wuite, and Erwin J. G. Peterman. Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. Proc. Natl. Acad. Sci., 106(43):18231–18236, 2009.

[156] Nikolaos K. Voulgarakis, Amaia Redondo, Alan R. Bishop, and Kim O. Rasmussen. Probing the mechanical unzipping of DNA. Phys. Rev. Lett., 96(24):248101, 2006.

[157] Nikolaos K. Voulgarakis, Amaia Redondo, Alan R. Bishop, and Kim O. Rasmussen. Sequencing DNA by dynamic force spectroscopy: limitations and prospects. Nano Lett., 6(7):1483–1486, 2006.

[158] Bernd Wagner, Rainer Tharmann, Ilka Haase, Markus Fischer, and Andreas R. Bausch. Cytoskeletal polymer networks: the molecular structure of cross-linkers determines macroscopic properties. Proc. Natl. Acad. Sci., 103(38):13974–13978, 2006.

[159] Michelle D. Wang, Hong Yin, Robert Landick, Jeff Gelles, and Steven M. Block. Stretching DNA with optical tweezers. Biophys. J., 72(3):1335–1346, 1997.

[160] James D. Watson and Francis H. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature, 171(4356):737–738, 1953.

[161] Jin-Der Wen, Laura Lancaster, Courtney Hodges, Ana-Carolina Zeri, Shige H. Yoshimura, Harry F. Noller, Carlos Bustamante, and Ignacio Tinoco Jr. Following translation by single ribosomes one codon at a time. Nature, 452(7187):598–603, 2008.

[162] Jin-Der Wen, Maria Manosas, Pan T. X. Li, Steven B. Smith, Carlos Bustamante, Felix Ritort, and Ignacio Tinoco Jr. Force unfolding kinetics of RNA

using optical tweezers. I. Effects of experimental variables on measured results. Biophys. J., 92(9):2996–3009, 2007.

[163] Robert H. Wood, Wolfgang C. F. Muehlbauer, and Peter T. Thompson. Systematic errors in free energy perturbation calculations due to a finite sample of configuration space: sample-size hysteresis. J. Phys. Chem., 95:6670–6675, 1991.

[164] Michael T. Woodside, Peter C. Anthony, William M. Behnke-Parks, Kevan Larizadeh, Daniel Herschlag, and Steven M. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. Science, 314(5801):1001–1004, 2006.

[165] Michael T. Woodside, William M. Behnke-Parks, Kevan Larizadeh, Kevin Travers, Daniel Herschlag, and Steven M. Block. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. Proc. Natl. Acad. Sci., 103(16):6190–6195, 2006.

[166] Michael T. Woodside, Cuauhtemoc Garcia-Garcia, and Steven M. Block. Folding and unfolding single RNA molecules under tension. Curr. Opin. Chem. Biol., 12(6):640–646, 2008.

[167] Gijs J. Wuite, Steven B. Smith, Mark Young, David Keller, and Carlos Bustamante. Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. Nature, 404(6773):103–106, 2000.

[168] Toshio Yanagida, Michiyuki Nakase, Katsumi Nishiyama, and Fumio Oosawa. Direct observation of motion of single F-actin filaments in the presence of myosin. Nature, 307(5946):58–60, 1984.

[169] Ryohei Yasuda, Hiroyuki Noji, Masasuke Yoshida, Kazuhiko Kinosita, and Hiroyasu Itoh. Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase. Nature, 410(6831):898–904, 2001.

[170] Hong Yin, Michelle D. Wang, Karel Svoboda, Robert Landick, Steven M. Block, and Jeff Gelles. Transcription against an applied force. Science, 270(5242):1653–1657, 1995.

[171] Yan Zeng and Giovanni Zocchi. Mismatches and bubbles in DNA. Biophys. J., 90(12):4522, 2006.

[172] Xiaowei Zhuang and Matthias Rief. Single-molecule folding. Curr. Opin. Struct. Biol., 13(1):88–97, 2003.

[173] Eric N. Zimanyi and Robert J. Silbey. The work-Hamiltonian connection and the usefulness of the Jarzynski equality for free energy calculations. J. Chem. Phys., 130(17):171102, 2009.

[174] Daniel M. Zuckerman and Thomas B. Woolf. Theory of a systematic computational error in free energy differences. Phys. Rev. Lett., 89(18):180602, 2002.

[175] Michael Zuker. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res., 31(13):3406–3415., 2003. See also http://mfold.rna.albany.edu/.

List of publications

The results of the present work have been published in:

- Chapter 3: Sandra Engel, Anna Alemany, Nuria Forns, Philipp Maass, and Felix Ritort. Folding and unfolding of a triple-branch DNA molecule with four conformational states. Phil. Mag., 91(13):2049–2065, 2011.
- Chapter 4: Sandra Nostheide, Viktor Holubec, Petr Chvosta, and Philipp Maass. Unfolding kinetics of periodic DNA hairpins. J. Phys.: Condens. Matter, 26:205102, 2014.
- Chapter 5: Sandra Nostheide and Philipp Maass. Unfolding kinetics of DNA hairpins using a distance protocol. To be published.

List of Figures

List of Tables

List of abbreviations

AD bead ..... antidigoxigenin-coated bead
AFM ..... atomic force microscope
bp, bps ..... base pair, base pairs
ch. ..... chapter
DP ..... distance protocol
DFC ..... distance-force curve
DNA ..... deoxyribonucleic acid
ds ..... double-stranded
EFC ..... extension-force curve
eq., eqs. ..... equation, equations
FP ..... force protocol
FDC ..... force-distance curve
FEC ..... force-extension curve
FED ..... free energy difference
FEL ..... free energy landscape
fig., figs. ..... figure, figures
FJC ..... freely jointed chain
KMC ..... kinetic Monte Carlo
OTs ..... optical tweezers
MTs ..... magnetic tweezers
PBD ..... Peyrard-Bishop-Dauxois
RNA ..... ribonucleic acid
SA bead ..... streptavidin-coated bead
sec. ..... section
SMEs ..... single-molecule experiments
ss ..... single-stranded
tab. ..... table
WLC ..... worm-like chain
WPDs ..... work probability distributions

Acknowledgement

"Zwei Dinge sind zu unserer Arbeit nötig:
Unermüdliche Ausdauer und die Bereitschaft,
etwas, in das man viel Zeit und Arbeit gesteckt hat,
wieder wegzuwerfen."

Albert Einstein (1879-1955), German physicist

This work was solely achievable with the assistance of a great number of people whom I am deeply indebted to.

First of all, I want to thank my supervisor Professor Philipp Maaß for his guidance and for his many inspiring ideas which were pushing me beyond my limits. Special thanks go to Dr. Tayebeh Jadidi, Hendrik Pils, Dr. Susanne Hahne, Dr. Omar Bakhti and the rest of my research group at the University of Osnabrück. I would like to express my sincere gratitudes to Professor Jochen Gemmer for examining this dissertation. Besides, I am deeply grateful for the various financial support of the University of Osnabrück.

Furthermore, I greatly appreciate that I could work together with Professor Felix Ritort who made me discover the utterly interesting and marvellous world of biomolecular studies. My warm thanks go to the whole group of the Small Biosystems Lab at the University of Barcelona for their manifold support, above all to Anna Alemany and Dr. Josep Maria Huguet.

Another group has played an important role for my research activities, namely Professor Petr Chvosta and Dr. Viktor Holubec at the Charles University of Prague. Thank you all for your hospitality and fruitful discussions.

Last but not least, I want to thank my awesome friends and family, especially my dear husband and my parents, for their incomparable love and support.