Nucleic Acid Thermodynamics Derived from Mechanical Unzipping Experiments

Paolo Rissone [ID] **and Felix Ritort** *[ID]

Small Biosystems Lab, Condensed Matter Physics Department, University of Barcelona,
Carrer de Martí i Franqués 1, 08028 Barcelona, Spain; rissone.paolo@gmail.com
* Correspondence: fritort@gmail.com; Tel.: +34-934035869

Abstract: Force-spectroscopy techniques have led to significant progress in studying the physicochemical properties of biomolecules that are not accessible in bulk assays. The application of piconewton forces with laser optical tweezers to single nucleic acids has permitted the characterization of molecular thermodynamics and kinetics with unprecedented accuracy. Some examples are the hybridization reaction between complementary strands in DNA and the folding of secondary, tertiary, and other heterogeneous structures, such as intermediate and misfolded states in RNA. Here we review the results obtained in our lab on deriving the nearest-neighbor free energy parameters in DNA and RNA duplexes from mechanical unzipping experiments. Remarkable nonequilibrium effects are also observed, such as the large irreversibility of RNA unzipping and the formation of non-specific secondary structures in single-stranded DNA. These features originate from forming stem-loop structures along the single strands of the nucleic acid. The recently introduced barrier energy landscape model quantifies kinetic trapping effects due to stem-loops being applicable to both RNA and DNA. The barrier energy landscape model contains the essential features to explain the many behaviors observed in heterogeneous nucleic-acid folding.

Keywords: single-molecule biophysics; nucleic acid thermodynamics; statistical mechanics; fluctuation theorems; computational biophysics

1. Introduction

In molecular biophysics, the accurate knowledge of the thermodynamics of nucleic acids (NAs) and proteins is essential to obtain reliable predictions of protein folding [1–5], DNA/RNA hybridization [6], and their interactions with enzymes and ions [7]. Over the past decades, measurements of molecular free-energies, entropies, and enthalpies have been obtained by using bulk techniques such as UV absorbance, fluorescence, and calorimetry, among others [8]. DNA and RNA hybridization are key reactions in many biochemical processes, such as NA synthesis, RNA folding, and DNA amplification by PCR. The energy parameters used to model the hybridization reaction can be directly obtained from melting curves of oligos of varying sequence and length. Accurate knowledge of these parameters is crucial in many ways, such as optimizing heating–cooling protocols for PCR products, structural predictions in NAs folding and biosensor devices that use DNA hybridization for detection (e.g., surface-plasmon resonance), and sequencing technologies. Unified sets of energy parameters have been derived from DNA and RNA melting temperature data obtained by many laboratories worldwide [9,10]. They are currently used as reference values by all main prediction tools [11].

However, bulk methods yield results that are incoherent temporal averages over a large population of molecules that are in different states (Figure 1, left). The measured signal depends on the dominant species and reactions, limiting the capability of detecting fast events, rare non-native states, and reaction pathways [12]. For example, RNAs and proteins often become trapped in non-productive, misfolded structures [3,13–17]. Such

structures have been related to the development of many phenotype diseases such as Huntington's disease, fragile X-associated tremor ataxia syndrome, myotonic dystrophies, and spinocerebellar ataxias, among others [18–20]. By monitoring one molecule at a time, single-molecule techniques allow for the characterization of these structures, which go undetected in bulk assays. Notice that by averaging results over many events, single-molecule measurements approach the bulk limit to which results can be compared.

By overcoming the intrinsic limitations of bulk measurements, techniques such as single-molecule fluorescence [21], single-molecule translocation across nanopores [22], atomic-force microscopy (AFM), magnetic tweezers (MT), and laser optical tweezers (LOT) [23] offer a fresh view in biomolecular sciences (Figure 1, right panel). In particular, LOTs [24] were revealed to be a powerful tool to investigate the properties of NAs. In fact, LOTs permit the direct measurement of the mechanical work (and therefore of the free-energy) needed to unfold DNA and RNA hairpins, rendering force spectroscopy a valuable tool for NA thermodynamics [25,26]. Moreover, single-molecule manipulation sets a new bar for resolving complex molecular reactions [27], such as NAs' elastic response [28,29] and non-specific secondary structure formation [30], as well as their protein and NA folding [31,32].

During the past decade, single-molecule techniques have become standard for investigating NAs at the structural, biochemical, and thermodynamical levels. Here, we briefly review the thermodynamics of DNA and RNA folding as obtained by developments in our lab over the past ten years that combine LOTs with suitable data analysis methods. In this way, it has been possible to measure DNA and RNA base-pair free-energies from mechanical unzipping experiments with high accuracy (0.1 kcal/mol) in different experimental conditions [33–35]. Unzipping consists of pulling apart the 3′ and 5′ ends on one side of a helical NA structure stabilized by complementary (Watson–Crick) base-pair interactions. The reverse process, called rezipping, consists of the reformation or assembly of the helical structure (hybridization) by approaching the 3′ and 5′ ends. Even though both DNA and RNA form double-stranded helices, important differences are found in unzipping experiments. In particular, unzipping is a fully reversible process for DNA in a broad range of salt conditions and loading rates. Instead, RNA unzipping turns out to be strongly irreversible in the same experimental conditions. In this case, transient off-pathway misfolded structures appear during the unzipping–rezipping process, slowing down the hybridization reaction [36–38]. The characterization of these off-pathway structures that compete with the native stem is a challenging problem that requires the extraction of the free energies of kinetic (non-native) states from irreversible work measurements [25,39,40]. In unzipping experiments, DNA and RNA molecules unravel into a single NA chain (single-strand) of a given molecular extension. In contrast, in bulk experiments, no forces are involved, and the unfolded state is a relaxed random coil. To relate unzipping experiments with bulk measurements, it is necessary to correct free energy differences by the so-called stretching energy contribution. This term equals the work needed to stretch the NA chain from the random coil state (zero extension) to the elongated state (finite extension). Therefore, thermodynamic measurements with force spectroscopy require accurate knowledge of the ideal elastic response of the NA chains, which can also be measured by pulling the individual single-strands [29,30]. Here, we discuss the experimental results obtained in our lab and interpret them in the framework of the barrier energy landscape (BEL) model recently introduced by us to explain the strong irreversibility observed in RNA unzipping experiments [35]. We also extend the BEL model to predict of the force–extension curves reported in studies of the non-specific secondary structure in single-stranded DNA [30].

The paper is organized as follows: In Section 2, we describe the LOT setup and the experimental protocol to unzip NAs. In Section 3, we introduce the popular nearest-neighbor model used to characterize the hybridization reaction in DNA and RNA duplexes. In Section 4 we review the results obtained by unzipping long DNA and RNA hairpins with LOTs. In Section 5, we introduce the BEL model, which can be used to explain the strong irreversibility observed in RNA unzipping and the formation of non-specific secondary structures in DNA [30]. Finally, in Section 6, we discuss future perspectives.
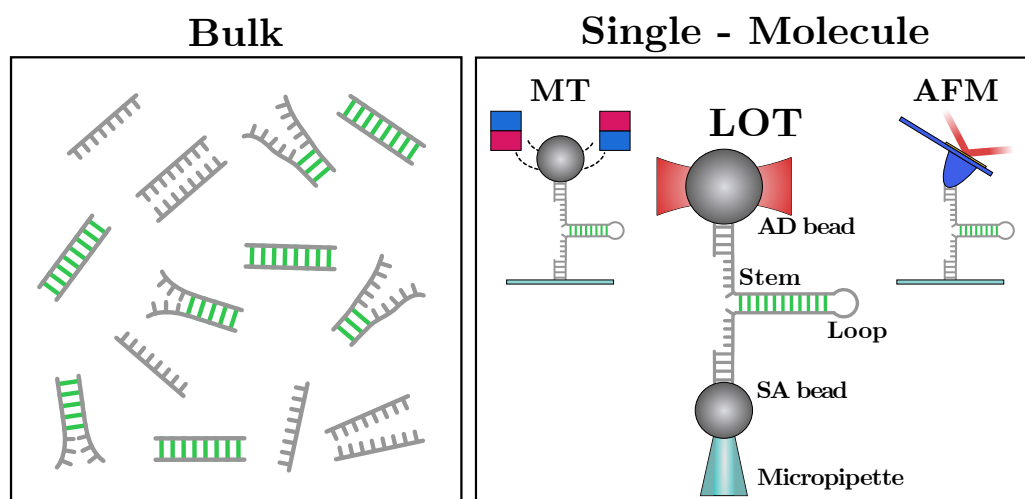
Figure 1. Information about biomolecules from bulk essays (**left panel**) is not a direct procedure and is obtained from the global behavior of a large number of molecules. On the contrary, single-molecule techniques (**right panel**) such as atomic-force microscopy (AFM), magnetic tweezers (MT), and laser optical tweezers (LOT), allow us to sample reactions one molecule at a time, characterizing the molecules at the microscopic level.

2. Mechanical Unzipping of Nucleic Acids (NAs)

To measure the NAs free-energy of formation at the single base-pair (bp) level, we unzip long DNA and RNA hairpins (a few thousand bases) consisting of a stem of fully complementary Watson–Crick base-pairs (Figure 2). The stem terminates in a tetra-loop GAAA. When fully unzipped, the hairpin converts to the single-stranded form. On the other side, the 3′ and 5′ ends of the hairpin are ligated to short (29 bp) double-stranded handles, one labeled with a digoxigenin tail (DIG) and the other with biotin (BIO). The two tags specifically bind to anti-DIG (AD)- and streptavidin-coated (SA) beads, respectively. The AD bead is optically trapped, while the SA one is held by air suction at the tip of a glass micro-pipette (Figure 1, right panel). In an unzipping experiment, the optical trap is moved with respect to the (fixed) micro-pipette at a constant speed. At the beginning of the protocol, the molecule is folded into its native double-stranded (ds) hairpin configuration (Figure 2, top-left). As the optical trap moves away from the pipette, the force applied to the hairpin increases until the intramolecular bonds at the beginning of the stem break open. As the unzipping progresses, groups of new bases open one after another in a sequential fashion. Unzipping is a stick–slip process consisting of the succession of an elastic deformation (stick) followed by the release of groups of bases that collectively unfold in a cooperative manner (slip), resulting in sudden force jumps. The unfolding protocol proceeds until the hairpin is fully unzipped and the single-strand (ss) form is fully stretched (Figure 2, bottom-left). At this point, the reverse process starts (rezipping), and the molecule refolds starting from the loop until the native ds hairpin has been reformed. Upon rezipping, groups of bases are cooperatively absorbed into the stem resulting in sudden increases in force. The force–distance curve (FDC) measured during unzipping and rezipping exhibits a saw-tooth pattern that depends on the sequence of the hairpin. In Figure 2 (right), we show the FDCs obtained by unzipping a 6.8 kbp DNA hairpin (top) in a 1 M NaCl solution and a 2 kbp RNA hairpin in a 500 mM NaCl solution (bottom). Notice that RNA FDCs exhibit many irreversible regions (insets) with large hysteresis and many long-lived intermediates. This does not happen in DNA, where unzipping experiments are carried out at equilibrium.
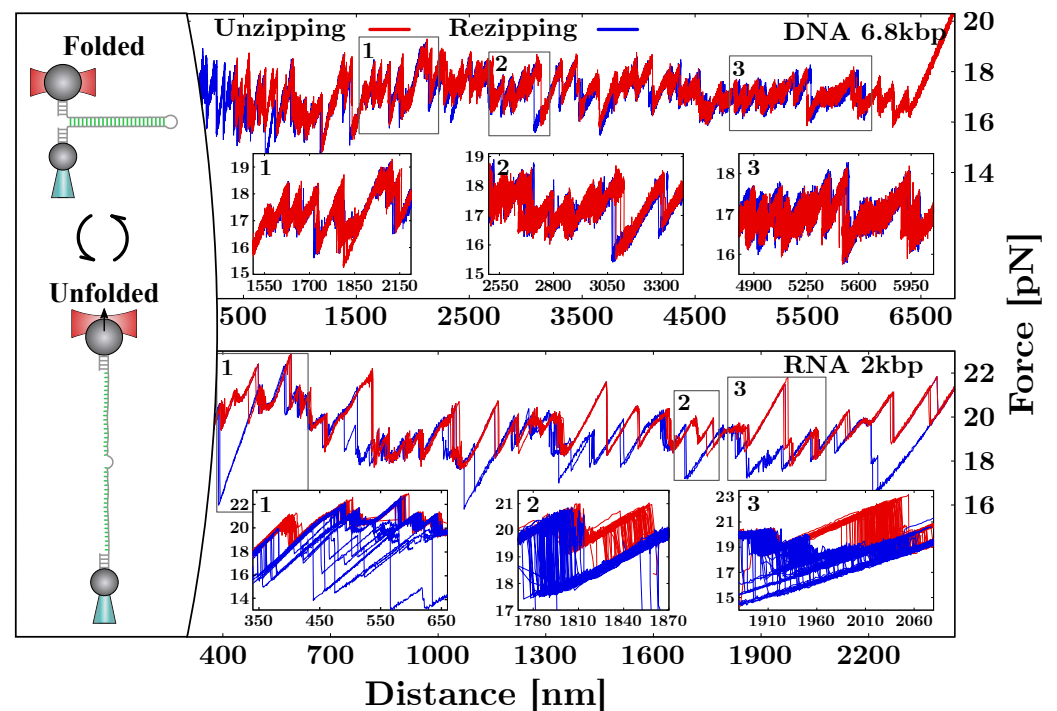
Figure 2. Unzipping experiment with LOTs. (**Left**) In a pulling experiment, the molecular construct is repeatedly unzipped and rezipped. By starting in the folded configuration (**top**), the trap–pipette distance is increased until the molecule is fully unfolded (**bottom**). Then, the process is reversed, and the hairpin refolds upon moving closer the optical trap to the micropipette. (**Right**) The typical FDCs obtained for a 6.8 kbp DNA hairpin (**top**) and a 2 kbp RNA hairpin (**bottom**) in a 1 M NaCl buffer and a 500 mM NaCl buffer, respectively. The insets magnify selected regions (gray squares) along the FDCs to point out the large irreversibility observed in RNA. In contrast, DNA unzipping is fully reversible.

3. Modeling the Unzipping Experiment

Single-molecule experiments can be used to measure the hybridization free-energy of NA, the interaction energy between hairpins and proteins [41–43], and ions [44]. They can also be used to monitor the folding pathway and detect intermediates and misfolded states [40]. As explained in the introduction, force spectroscopy measurements require the knowledge of the contributions of the experimental apparatus (optical trap) and the stretched molecular construct (handles and single-stranded NA). In fact, the standard free-energy of formation of a hairpin, ΔG_0, is defined as the free-energy difference between the native and the random coil state, the latter being the state (with zero extension) of the unfolded semi-flexible single-stranded NA when no external force is applied. However, in mechanical unzipping experiments, the unfolded molecule does not attain a random coil but a stretched semiflexible form, making the free energy difference $\Delta G_{tot} \neq \Delta G_0$. To obtain ΔG_0, we need to subtract the contribution by the experimental setup to the measured ΔG_{tot}. Therefore, a theoretical understanding of the experimental contributions to ΔG_{tot} is essential to derive reliable free energy values.

3.1. Bases Hybridization and Nearest-Neighbors Model

NAs are polymeric chains of monomers, the nucleotides that are linked together by covalent and non-covalent interactions [45] (see Figure 3A, left). The nucleotides are divided into purines, i.e., Adenine (A) and Guanine (G), and pyrimidines, i.e., Cytosine (C) and Thymine (T). For RNA, the nucleotide Uracil (U) substitutes Thymine. A nucleotide (see Figure 3A, right) is formed by one molecule of phosphoric acid (dark-yellow circle), one molecule of 2'-deoxyribose (light-yellow pentagon), and a nitrogenous base (which can be A,C,G,T/U). Two nucleotides concatenate through a bond between the phosphoric group

of the first nucleotide and the third carbon of the second nucleotide. The concatenation of multiple nucleotides forms a phosphate–deoxyribose backbone of linked bases featuring a phosphate group on one terminus (5′) and a hydroxyl group on the other one (3′). The structure of NAs is conventionally given as a sequence of bases in the direction 5′ → 3′.

The canonical interactions between nucleotides belonging to different strands are given by the Watson–Crick base-pairing rules [46] and account for purine–pyrimidine bonding: A links to T (U in the RNA case) and G to C with three and two hydrogen bonds, respectively. Although hydrogen bonding is responsible for the specificity of base interactions, much of the stability of the NA is due to base stacking [47]. In fact, nitrogenous planar bases are non-polar, and the hydrophobic stacking on top of each other maximizes the Van der Waals attraction. To account not only for the specific base-pairing but also for the stacking between adjacent base-pairs, the duplex energetics is described with the nearest-neighbor (NN) model [9,48–50].

In the NN model, the base-pairing energy of two complementary bases only depends on the base itself and on the first neighbor located in the same strand (in the 5′ → 3′ direction). Therefore, the total free energy of formation of a duplex, ΔG_0, is given by the sum over all the nearest-neighbor base-pair (NNBP) motifs occurring along the sequence:

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

where Δg_i is the free energy of motif i. Notice that the NNBP energies are negative, as they are defined as the free-energy loss upon hybridizing a base-pair, i.e., $\Delta g_i = g_i^H - g_i^O < 0$, where $g_i^H (g_i^O)$ is the free-energy of the hybridized (open) motif. There are 16 different motifs accounting for all possible combinations of adjacent NNBPs (see Figure 3B). This number is reduced from 16 to 10 by considering the degeneracy of the free-energies due to the Watson–Crick complementarity. It is possible to further reduce this number from 10 to 8 independent parameters by considering the circular symmetry of the NN model [51,52]. This symmetry yields additional self-consistent relations so that out of the 10 NNBP energies, 2 can be expressed as linear combinations of the remaining 8 [34,52,53]. Motifs TA/AT (UA/AU) and GC/CG are usually expressed as a function of the others:

$$\Delta g_{\text{TA(UA)}} = \Delta g_{\text{CG}} + \frac{1}{2}(\Delta g_{\text{AC}} + \Delta g_{\text{GA}} - \Delta g_{\text{AG}} - \Delta g_{\text{CA}})$$

$$\Delta g_{\text{GC}} = \Delta g_{\text{AT(AU)}} + \frac{1}{2}(\Delta g_{\text{GA}} + \Delta g_{\text{CA}} - \Delta g_{\text{AG}} - \Delta g_{\text{AC}}) \tag{2}$$

where Δg_{XY} indicates the free energy of motif 5′-XY-3′ hybridized with its complementary sequence. The accurate measure of the NNBP free-energies is key for the correct estimation of the total free energy of formation of the duplex native state. The 10 independent parameters have been extracted from melting experiments of short duplexes of varying sequences and lengths [10,54–56] for both DNA and RNA and are accessible in the Mfold server [11]. Single-molecule techniques allow for much more accurate free-energy measurements with respect to bulk experiments. However, to derive the 10 (8 if circular symmetry is considered) NNBP parameters from unzipping experiments, it is fundamental to have a theoretical model of the unzipping process to predict the experimental FDC.
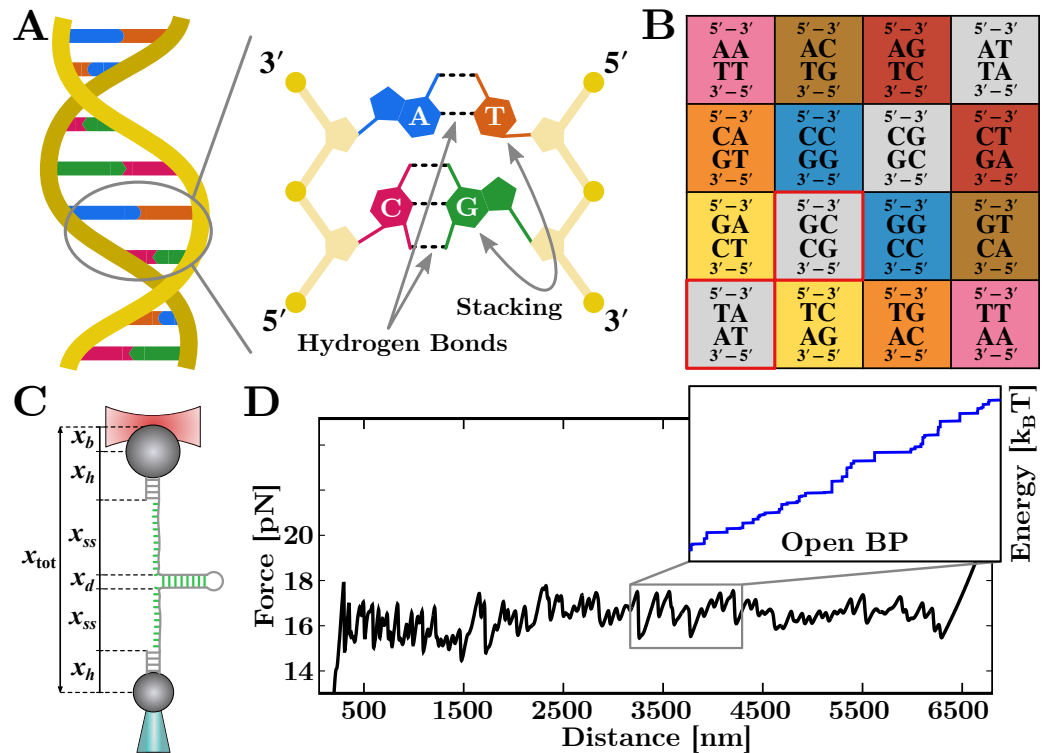
Figure 3. (**A**) Double-helix structure according the Watson–Crick base-pairing rules. (**B**) Matrix of the 16 NNBP motifs according to the NN model. Degenerate energies have the same cell color. Out of the 10 independent parameters, circular symmetry allows to express 2 NNBP energies (TA/AT and GC/CG—red-bordered cells) as a linear combination of the others (see Equation (2)). Notice that the same matrix is obtained for the RNA case by changing thymine (T) for uracil (U). (**C**) Different contributions of the molecular construct to the total extension. Each term (optical trap, dsDNA handles, ssDNA, and folded double-helix) has a different elastic response to the applied force. (**D**) Theoretical prediction of the FDC and of the minimum free-energy as a function of the open bp (inset).

3.2. Computation of the System Free-Energy

In unzipping experiments at controlled position, the trap–pipette distance, x_{tot}, is steadily increased (unfolding) or decreased (refolding) by moving the position of the optical trap. As the optical trap moves, the force f applied on the molecular construct changes and the number of open bases n varies accordingly. The total trap–pipette distance equals the sum of different contributions [57] (see Figure 3C), that is:

$$x_{tot}(f,n) = x_b(f) + 2x_h(f) + 2x_{ss}(f,n) + x_d(f) \tag{3}$$

where $x_b(f)$ is the displacement of the bead in relation to the center of the optical trap, $x_h(f)$ is the extension of the double-stranded handles, $x_{ss}(f,n)$ is the extension of the single-stranded (unfolded) molecule containing n bases, and $x_d(f)$ accounts for the extension of the diameter of the NA helix d (typically $d = 2$ nm for DNA and RNA hairpins [58]) projected along the force axis [59]. Notice that the latter term is zero when the hairpin is fully unzipped. The bead in the optical trap is modeled as a linear spring of stiffness k:

$$f(x_b) = kx_b. \tag{4}$$

The double-stranded handles and the single-stranded NA are modeled as elastic polymers. Typically, these terms are described by the inextensible or extensible Worm-Like Chain (WLC) model [60]. In the former case:

$$f(x) = \frac{k_B T}{4 l_p} \left[\left(1 - \frac{x}{n l_d} \right)^{-2} - 1 + 4 \frac{x}{n l_d} \right] \tag{5}$$

where k_B is the Boltzmann constant, T is the temperature, l_p is the persistence length, and l_d is the interphosphate distance. Finally, the extension upon orienting the double helix is modeled as a dipole of length equal to the helix diameter, d, that aligns along the force axis:

$$x_d(f) = d \left[\coth \left(\frac{fd}{k_B T} \right) - \frac{k_B T}{fd} \right]. \tag{6}$$

Notice that while $x_b(f)$ and $x_d(f)$ are directly obtained from Equations (4) and (6), respectively, $x_{ss}(f, n)$ requires inverting Equation (5) [57].

For a given x_{tot} and n, the total free-energy of the NA hairpin is given by these elastic terms plus the hybridization free-energy in Equation (1) so that:

$$\Delta G_{tot}(x_{tot}, n) = \Delta G_b(x_b) + 2\Delta G_h(x_h) + 2\Delta G_{ss}(x_{ss}, n) + \Delta G_d(x_d) + \Delta G_0(N - n) \tag{7}$$

where N is the total number of base-pairs in the sequence and the distance constraint Equation (3). The elastic free-energy contributions are obtained by computing the following integral:

$$\Delta G_{el}(x) = \int_0^x f(x') dx' \tag{8}$$

of Equations (4)–(6). Therefore, it is straightforward to obtain:

$$\Delta G_b = \frac{1}{2} k x_b^2 \tag{9}$$

and

$$\Delta G_{WLC}(x) = \frac{k_B T}{4 l_p} \left[n l_d \left(1 - \frac{x}{n l_d} \right)^{-1} - x + 2 \frac{x^2}{n l_d} \right] \tag{10}$$

from Equations (4) and (5), respectively. The computation of the free-energy needed to orient the dipole-like folded hairpin requires inverting Equation (6).

3.3. The Equilibrium FDC

Given the above model, it is possible to compute the equilibrium force of the system, f_{eq}, at each instant of the unzipping protocol. This ultimately allows us to obtain a theoretical prediction of the equilibrium FDC of the hairpin sequence. To do this, let us introduce the system partition function for a fixed x_{tot}, which is defined as the sum over all the possible states (all the possible values of n) so that:

$$Z(x_{tot}) = \sum_{n=0}^{N} \exp \left(-\frac{\Delta G_{tot}(x_{tot}, n)}{k_B T} \right) \tag{11}$$

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

$$f_{eq}(x_{tot}) = -k_B T \frac{\partial \ln Z(x_{tot})}{\partial x_{tot}}. \tag{12}$$

Computing Equation (11) requires solving the transcendental equation (3) (that can be performed numerically) with respect to f and then computing Equation (7) for all $n \in [0, N]$. The value n^* minimizing the equilibrium free-energy $\Delta G_{eq} = \Delta G_{tot}(x_{tot}, n^*(x_{tot}))$ gives

the most probable number of open base-pairs at a given x_{tot}. Eventually, the computation of the equilibrium force in Equation (12) gives a theoretical prediction for the unzipping curve of a given sequence (see Figure 3D). Notice that the total energy in Equation (7) is given by the balance between two independent contributions: the elastic terms and the hybridization energy. The latter term grows linearly with n (and does not depend by x_{tot}) while the elastic one is a non-linear function of (x_{tot}, n). This means that a variation of x_{tot} does not necessarily imply a variation of n^*. As a result, the equilibrium energy ΔG_{eq} is a rough function of x_{tot} exhibiting a sequence of steps (Figure 3D, inset) as n^* changes upon releasing groups of Δn^* bases. Each one of these jumps corresponds to a rip along the equilibrium FDC (Figure 3D).

4. Derivation of the NNBP Free-Energies

Unzipping data for long NA hairpins have been used to derive the formation free energies of the 10 NNBP motifs in DNA [33,34] and RNA [35]. To do this, we developed a Monte Carlo optimization algorithm that uses both the experimental data and the theoretical FDC prediction described in Section 3 [33]. By starting with an initial guess of the 10 independent parameters (or 8 if circular symmetry is applied), at each step of the optimization, a random increment in the energies is proposed and a prediction of the FDC is generated. The error made in approximating the experimental curve with the theoretical one, E, drives a Metropolis algorithm: a change in the energy parameters is accepted if the error difference with respect to the previous step is negative ($\Delta E < 0$). Otherwise ($\Delta E > 0$), the proposal is accepted if $\exp(\Delta E / T) < r$, with r being a random number uniformly distributed $r \in U(0, 1)$. The algorithm continues until convergence is achieved, i.e., until ΔE is smaller than a given threshold. Let us note that because of the high number of parameters, only experimental data from the unzipping of long molecules (a few kbp) allow for an accurate estimation of the NNBP energies. In fact, the algorithm relies on the saw-tooth pattern characteristic of the sequence to accept or reject an energies proposal: the longer the sequence is, the more accurate the values of the NNBP energies are.

In Figure 4, we show the results obtained for the eight independent NNBP free-energies of DNA (circles) and RNA (squares) at different concentrations of sodium (blue) and magnesium (red). In particular, the DNA energies have been measured in a salt range from 10 mM to 1 M for Na^+ and from 0.01 mM to 10 mM for Mg^{++}, while the RNA energies have been measured at 500 mM Na^+ and 10 mM Mg^{++}. The last two parameters, GC/CG and TA/AT (UA/AU for RNA), were obtained by applying the circular symmetry relations. It is apparent that the (negative) energy parameters are lower for RNA than for DNA. Notice that this difference is more marked for motifs containing at least one purine (A, G), i.e., where stacking interactions are stronger, explaining why RNA tends to fold into more compact structures than DNA does [61].

Moreover, the NNBP RNA energies have been used to test the validity of the 1/100 salt equivalence rule of thumb. The rule states that the RNA energy parameters at a given divalent salt concentration equal that at a 100-fold monovalent salt concentration, i.e., $\Delta g_i[\text{Div}^{++}] = \Delta g_i[\text{Mon}^+]_{\text{eq}}$, where $[\text{Mon}^+]_{\text{eq}} = 100 \times [\text{Div}^{++}]$. Our results proved that 10 mM Mg^{++} corresponds to 800 mM Na^+ (approximately a 1/80 equivalence), which is compatible with the 1/100 rule within the experimental errors [35]. Although this result has been tested in a single-salt condition, its validity extends to the dilute salt regime where cooperative salt effects are negligible ($[\text{Mg}^{++}] < 0.05$ M) and competition effects with sodium are weak ($R = \sqrt{[\text{Mg}^{++}]}/[\text{Na}^+] > 0.22 \text{M}^{-1/2}$) [62,63]. The 1/100 salt equivalence rule of thumb has been disputed on the basis of experimental data obtained in bulk experiments using atomic emission spectroscopy in buffer-equilibrated samples [64]. Although this technique is capable of determining the fraction of cations that are dissociated and bound to the RNA, it does not provide a direct measurement of the free energies.
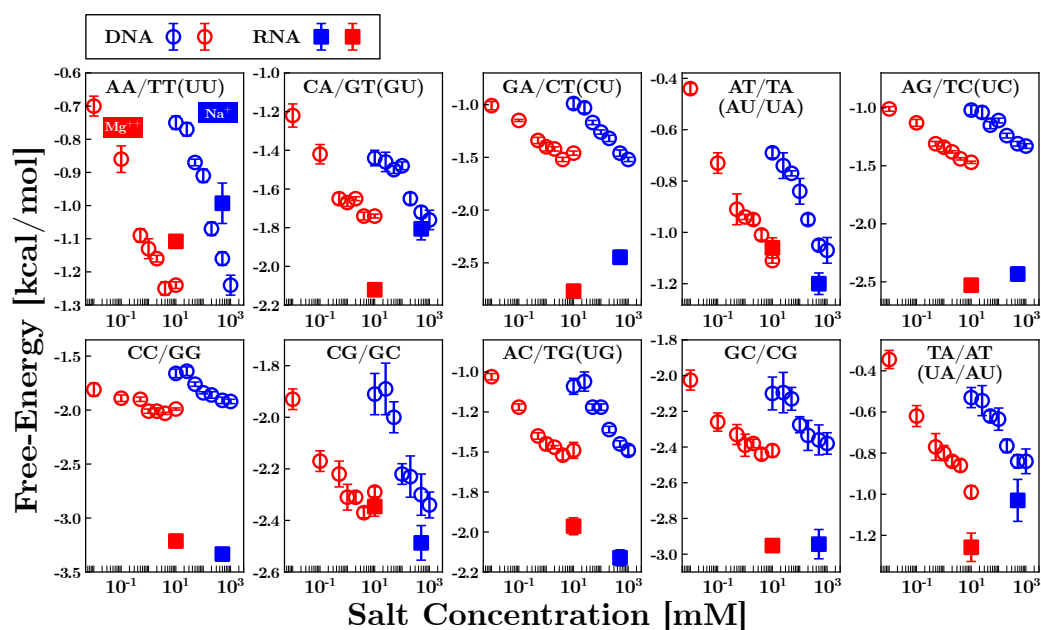
Figure 4. Experimentally measured NNBP free-energies of DNA (circles) and RNA (squares) at different salt concentrations of sodium (blue) and magnesium (red). The salt dependence of the NNBP energies in DNA has been studied in the ranges [10, 1000] mM NaCl and [0.01, 10] mM MgCl$_2$ [34]. In RNA, the energies have been measured at 500 mM NaCl and 10 mM MgCl$_2$ [35]. The energy of the last two motifs, GC/CG and TA/AT (UA/AU for RNA), has been computed by using the circular symmetry relations (see Equations (2)).

5. Out-of-Equilibrium Processes and Kinetics Effects

To characterize the irreversibility observed along the unzipping FDCs for RNA (see Figure 2), we introduced a many-valley barrier energy landscape (BEL) that accounts for the off-pathway competing stem-loop structures that can form along each single strand of the hairpin [35]. Upon unfolding (refolding) the hairpin, multiple stem-loops of varying lengths *L* can form along the single strands of the hairpin, i.e., in the vicinity of the hybridization junction separating the stem from the ssNA (Figure 5A, bottom right). Stem-loops slow down the hybridization of the native stem by stabilizing off-pathway conformations that kinetically trap the multiple apparent intermediates observed along the FDC (Figure 5B, grey circle snapshot). These complex structures are transiently stable; therefore, the stem-loops eventually unfold, and the native stem can form (Figure 5A).

We model the loop-BEL as follows (see Ref. [35] for details). Let us consider the set of all possible segments of *L* bases along a single-stranded sequence of *N* bases, $\mathcal{S}_L = \{[b_i, b_{i+L}]; 1 \leq i \leq N - L\}$, where b_i and b_{i+L} are the initial and the final bases of the segment, respectively. For a given *L*-segment, $[b_i, b_{i+L}]$, there are several competing folds, most of them stabilized by a few Watson–Crick base-pairs plus one or more loops of varying sizes (mostly from three to eight bases). By using the DINAmelt web application [65,66], we found the minimum free energy fold for a given segment. This gives the set of lowest free energy folds, $\{\epsilon_{L,i}^0\}$ for all *L*-segments in $\mathcal{S}_L$ at standard conditions. To construct the loop-BEL, one should consider all possible excitations (i.e., higher energy states) formed by multiple stem-loops folding along the two ssDNA strands at both sides of the junction. In principle, any number of stem-loops can form at arbitrary positions along the two strands. As counting all possible excitations is a daunting task, in the simplest approach, one can simplify the treatment by considering two stem-loops (one per strand) that are located at arbitrary and independent positions. However, stem-loops that are located far away from the junction, even if energetically favorable, cannot interfere with the unzipping–rezipping of the hairpin, a reaction taking place precisely at the junction. For a given stem-loop at a distance *j* from the junction and a given force *f*, we introduce an

energy penalty equal to the mechanical work at constant f needed to bring that stem-loop from distance $j + L$ to the junction, $\int_0^f x_{j+L}(f')df'$ (Figure 5A, bottom left).

Given L, we define the loop-BEL at force f and junction position n as:

$$\Delta G_L(n, f) = -k_B T \log \sum_{j_1, j_2=0}^{N-n} \exp\left(-\frac{\Delta g_L^{(1)}(j_1, f) + \Delta g_L^{(2)}(j_2, f)}{k_B T}\right), \tag{13}$$

where $\Delta g_L^{(1,2)}(j, f)$ is the free-energy of forming a single stem-loop in strand $(1, 2)$ of length L plus the work at force f to bring it from position j to the junction located at position n. $\Delta g_L^{(1,2)}(j, f)$ is then given by:

$$\Delta g_L^{(1,2)}(j, f) = \epsilon_{L,j}^{0(1,2)} + \int_0^f x_{L+j}(f')df'. \tag{14}$$

Here, $\epsilon_{L,j}^{0(1,2)}$ is the (negative) free energy of formation at zero force of the stem-loop $[b_j, b_{j+L}]$ in strand $(1, 2)$. The integral term accounts for the free energy cost at force f to bring the base of the stem-loop located at the farthest end, $j + L$ bases away from the junction, towards the junction. As previously said, this term penalizes stem-loops that are formed far away from the junction because they cannot kinetically trap the native RNA hairpin. The elastic response term $x_{j+L}(f)$ has been modeled with the WLC in Equation (5).

We found that the minima of the loop-BEL (Figure 5B) directly correlate with the amount of hysteresis observed along the FDC [35]. This correlation is maximum for stem-loops of length $L \sim 20$ bases, indicating that L-segments of this size are the most likely to fold. Moreover, the analysis of the hairpin sequence showed that the regions of large irreversibility contain a high fraction of stacked (A, U) bases that are prone to Watson–Crick pairing within the same strand. On the contrary, regions of low hysteresis are characterized by lower stacking and intra-strand pairing interactions. The formation of stem-loops structures may contribute to explaining the broad phenomenology of heterogeneous RNA folding, from misfolding and multiplicity of native states to the formation of complex tertiary structures. In particular, the BEL model can be extended to the formation of non-specific secondary structures observed in pulling experiments of ssDNA and ssRNA. Although this phenomenon has not been investigated in RNA, it has been extensively studied in DNA [28,30,67–69]: upon stretching, the ssDNA's elastic response deviates from the expected ideal behavior of a polymeric chain (described for example with the WLC) forming a shoulder below $f \sim 10$–12 pN (see Figure 5C, bottom).

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

$$\Delta G_L(f) = -k_B T \log \sum_{k=0}^{K} \exp\left(-\frac{\Delta g_L(k, f)}{k_B T}\right), \tag{15}$$

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

$$\Delta g_L(k, f) = E_L(k) + \left[(N - kL)\Delta G_{ss}^1(f) + k\Delta G_d(f)\right]. \tag{16}$$

The term $E_L(k)$ accounts for the most energetically stable configuration of k stem-loops randomly positioned along the sequence. The term $(N - kL)\Delta G_{ss}^1(f)$ is the energy gain upon stretching the free $N - kL$ bases at force f corrected by the (smaller) energy contribution, $k\Delta G_d(f)$, of orienting k stem-loops along the force axis. Notice that $(N - kL)\Delta G_{ss}^1(f)$ is an extensive quantity, equal to the number of monomers, $N - kL$, times the energy cost

to stretch a single monomer, $\Delta G_{ss}^1(f) = -\int_0^f x_{ss}^1(f')df'$, where $x_{ss}^1(f)$ has been modeled according to the WLC in Equation (5). The same consideration holds for the dipole contribution of k stem-loops, $\Delta G_d(f) = -\int_0^f x_d(f')df'$, where $x_d(f)$ is modeled according to Equation (6).

An exact computation of $E_L(k)$ in Equation (16) requires considering non-overlapping stem-loops: if a stem-loop of length L forms at position n along the sequence, the next stem-loop can only form outside the interval $[n - L : n + L]$. This is an unaffordable mathematical task that we simplified by considering overlapping stem-loops in a mean-field approximation. In this approximation, $E_L(k)$ is taken as the typical total energy of k stem-loops randomly chosen over the ensemble of C_k different realizations, without imposing any constraints on these loops (i.e., they can be overlapping or non-overlapping). In contrast, the stretching contribution $(N - kL)\Delta G_{ss}^1(f) + k\Delta G_d(f)$ is taken independent of the k stem-loops realization. Therefore, we have:

$$E_L(k) \approx \min_{C_k}\left\{ \sum_k \epsilon_{L,k}^0 \right\} - k_{\mathrm{B}}T\log(C_k),\tag{17}$$

where, for the typical energy of k stem-loops ($\epsilon_{L,k}^0 < 0, \forall k, L$), we took the most stable configuration (i.e., the one of lowest energy) within the ensemble C_k. For large N, the total number of configurations in C_k is enormous, so we restricted the sampling to a few hundreds of configurations (typically 500). The second term in the rhs of Equation (17) is an entropic contribution stabilizing the formation of stem-loops. The total number of configurations is given by the binomial coefficient, $C_k = \binom{K}{k}$, or the number of ways k objects (stem-loops) can be arranged into $K = \lfloor N/L \rfloor$ different positions.

From Equations (15) and (16), we can compute the average ssDNA extension for a given L, which is defined as:

$$x_{ss,L}(f) = -\frac{\partial \Delta G_L(f)}{\partial f} = \frac{1}{Z_L}\sum_{k=0}^{K} x_L(k,f)\exp\left(-\frac{\Delta g_L(k,f)}{k_{\mathrm{B}}T}\right),\tag{18}$$

where Z_L is the system's partition function for a given L (c.f. Equation (15)):

$$Z_L = \exp\left(-\frac{\Delta G_L(f)}{k_{\mathrm{B}}T}\right) = \sum_{k=0}^{K} \exp\left(-\frac{\Delta g_L(k,f)}{k_{\mathrm{B}}T}\right)\tag{19}$$

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

$$x_L(k,f) = (N - kL)x_{ss}^1(f) + kx_d(f).\tag{20}$$

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

$$\Delta G(f) = -k_{\mathrm{B}}T\log\sum_L \exp\left(-\frac{\Delta G_L(f)}{k_{\mathrm{B}}T}\right).\tag{21}$$

and

$$x_{ss}(f) = \frac{1}{Z}\sum_L x_{ss,L}(f)\exp\left(-\frac{\Delta G_L(f)}{k_{\mathrm{B}}T}\right),\tag{22}$$

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

In Figure 5C we show the ssDNA extension predicted by the BEL model for a random DNA sequence of $N = 2027$ bases at 10 mM (top panel) and 1 M (middle panel) NaCl salt concentrations. These results are compared with experimental data from pulling experiments of a long DNA hairpin in the same salt conditions (bottom panel). The expected ssDNA elastic response for $L \in [10, 100]$ computed with Equation (18) (solid

lines) is shown along with $x_{ss}(f)$ in Equation (22) (dashed black line). It is apparent that the BEL model reproduces the deviation from the ideal WLC model (dashed gray line) experimentally observed below ~10 pN. This behavior results from the competition between the stem-loops of different sizes: the lower the force, the larger the contribution to Equation (22) by larger stem-loops. Therefore, as the ssDNA approaches the random coil state ($f = 0$ pN), the energetic gain to stretch large L-segments in Equation (16) tends to zero, while $E_L(k)$ remains constant, favoring the formation of large stem-loop structures. Remarkably, as the force approaches the dsDNA unzipping force ($f \approx 15$ pN), stem-loops of length $L \sim 30$–40 bases become the most likely folds. This number is not far from what has been reported for RNA, where $L \sim 20$ (see above and Ref. [35]). At higher forces, the elastic response collapses to the WLC as experimentally observed. Let us notice that at both 10 mM and 1 M NaCl, the predicted extension differs from the experimental data (solid dots in Figure 5C, bottom) when $f < 6$ pN. This is particularly evident at 10 mM NaCl as our model indicates that stem-loops still form at low force while no secondary structure is observed in the pulling trajectories. This and other potential inconsistencies come from the crude approximations made in Equation (17). In fact, this approximation only holds when $k \ll K$, i.e., when the typical distance between consecutive stem-loops is much larger than L, so that the overlapping is negligible. Despite the simplicity of the mean-field approximation, the BEL model is useful to study the complex behaviors observed in NA. A more rigorous analytical treatment may lead to a deeper understanding of heterogeneous folding in NA.
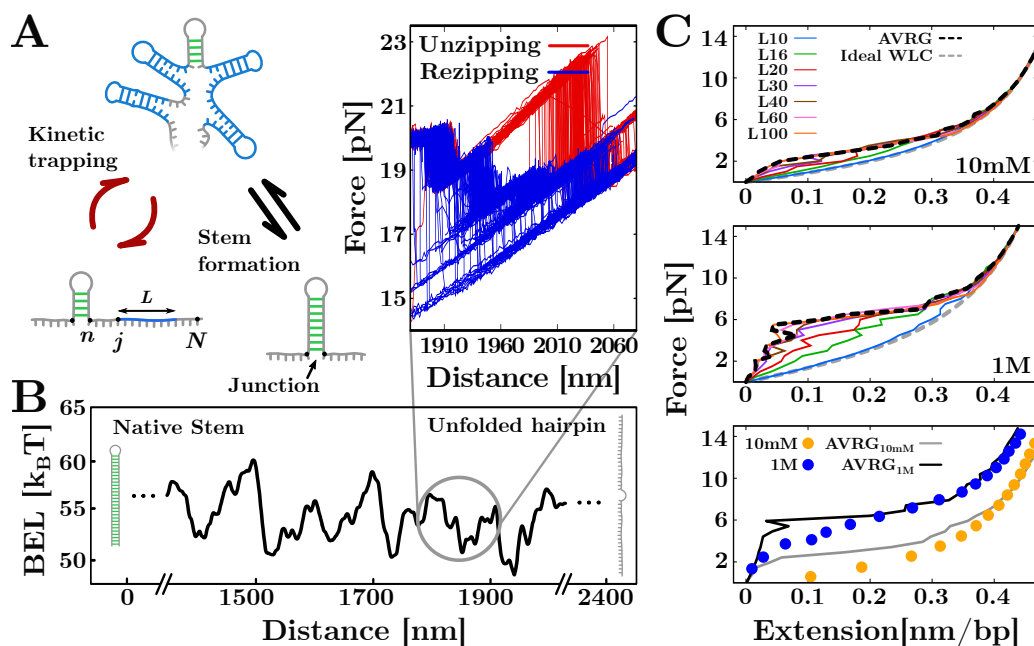

Figure 5. Stem-loops' formation along the single-stranded sequence. (**A**) Kinetic trapping during unfolding and refolding processes allows for the formation of stem-loops of different lengths folding along the unpaired strands of the molecule. As the state of highest stability is always the native state, the stem eventually forms. (**B**) Loop-BEL profile as a function of the trap–pipette distance. The zoom shows the FDC region corresponding to the circled loop-BEL. The minima of the landscape correlates with the measured hysteresis. (**C**) Extension of ssDNA predicted by the BEL model for different L-segments (solid lines) and over all L (dashed black lines) at 10 mM and 10 M NaCl (top and middle panels). The comparison between these predictions (solid lines) and experiments (solid dots) shows that the model reproduces the observed formation of non-specific secondary structure (bottom panel). The deviation form this behavior at low force ($f < 6$ pN) is due to the approximation made in computing the model.

6. Discussion

We discussed recent advances in single-molecule pulling experiments with laser optical tweezers (LOT) to investigate NAs' thermodynamics. Many of the results that we have presented cannot be obtained with bulk methods, which, by averaging over populations of molecules, do not permit sampling rare NAs folds such as misfolded configurations and intermediates. In addition to accurate experimental measurements, it is also key to have a reliable thermodynamic model for the unzipping reaction. In Section 3, we described the nearest-neighbor (NN) model that reproduces the experimental force–distance curves (FDCs). To date, the NN model provides the best theoretical description of the unzipping experiments. Conversely, unzipping experiments provide an elegant verification of the NN model. Our experimental-theoretical approach ultimately allows for predicting the unzipping FDC and the extraction of the nearest-neighbor base-pair (NNBP) energy parameters with 0.1 kcal/mol accuracy.

In Section 4, we summarized the NNBP energy parameters for DNA [34] and RNA [35]. Remarkably, the results are in agreement with the literature based on bulk assays [9,10]. The difference observed between the DNA and RNA energy parameters, where energies are lower for RNA, explains the higher stability of dsRNA duplexes in melting and unzipping experiments. It also explains the slower kinetics in RNA folding, where heterogeneous structures are often observed. This is confirmed by the multiple intermediates visible in the RNA FDCs that are not observed in DNA (Figure 2. Even though the irreversibility in RNA have been observed before in pulling experiments of short duplexes [70], a characterization of this general phenomenon in the unzipping of long RNA hairpins has been fundamental to measure the RNA NNBP free-energies. We do not exclude the possibility that similar intermediates also occur in DNA folding; however, this would require different experimental conditions, such as lower temperatures [5,71].

In Section 5, we have shown that the formation of stem-loops along the single strands explains the kinetic phenomena observed in unzipping experiments. To model this mechanism, we used the free-energy landscape formalism, useful to understand protein and NA folding into complex tertiary structures [5,72]. On the one hand, the barrier energy landscape (BEL) model introduced in Ref. [35] and defined in Equation (13) correlates with the hysteresis measured along the RNA FDCs. This shows that stem-loops are off-pathway structures that kinetically trap the RNA hairpin, slowing down the formation of the stem. On the other hand, the BEL model allows for predicting the formation of non-specific secondary structure in ssDNA pulling experiments [30]. This feature appears to be caused by the competition between multiple stem-loop folds of different sizes that make the extension to deviate from the ideal WLC behavior. Stem-loops formation appears as a general mechanism driving the folding and refolding processes of RNA hairpins as well as the elastic response of ssDNA sequences, and it may help understanding the broad phenomenology showed by NAs. However, the predictions obtained by the present model are limited by crude approximations introduced to simplify the enormous complexity of a complete stem-loops modeling. Firstly, in the computation of the stem-loops free-energy (Equations (16) and (17)), we disregarded overlapping effects between consecutive L-segments. This implies that partially overlapping L-segments can simultaneously fold into stem-loops. Moreover, we neglected the cooperativity effects, which favor the nucleation of contiguous stem-loops in an avalanche fashion reminiscent of phase transitions. Ultimately, a full BEL modeling would require allowing for the simultaneous formation of stem-loops of different lengths L, whereas the current model accounts for this effect through a mean-field approximation (Equations (21) and (22)). Although a comprehensive modeling of stem-loops formation is lacking, the phenomenon might explain many features of heterogeneous NA folding. The development of a more accurate description of the BEL model accounting for the complex phenomenology discussed above is left for future work.

References

1. Bryngelson, J.D.; Onuchic, J.N.; Socci, N.D.; Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **1995**, *21*, 167–195. [CrossRef] [PubMed]
2. Aabert, B.; Johnson, A.L.J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 4th ed.; Garland Science: New York, NY, USA, 2002.
3. Englander, S.W.; Mayne, L.; Krishna, M.M. Protein folding and misfolding: Mechanism and principles. *Q. Rev. Biophys.* **2007**, *40*, 1–41. [CrossRef] [PubMed]
4. Englander, S.W.; Mayne, L. The nature of protein folding pathways. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15873–15880. [CrossRef] [PubMed]
5. Rico-Pasto, M.; Zaltron, A.; Davis, S.J.; Frutos, S.; Ritort, F. Molten globule–like transition state of protein barnase measured with calorimetric force spectroscopy. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2112382119. [CrossRef] [PubMed]
6. Felsenfeld, G.; Miles, H.T. The physical and chemical properties of nucleic acids. *Annu. Rev. Biochem.* **1967**, *36*, 407–448. [CrossRef]
7. Shamsi, M.H.; Kraatz, H.B. Interactions of metal ions with DNA and some applications. *Inorg. Organomet. Polym. Mater.* **2013**, *23*, 4–23. [CrossRef]
8. Cantor, C.R.; Schimmel, P.R. *Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function*; Macmillan: New York, NY, USA, 1980.
9. SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1460–1465. [CrossRef]
10. Xia, T.; SantaLucia, J., Jr.; Burkard, M.E.; Kierzek, R.; Schroeder, S.J.; Jiao, X.; Cox, C.; Turner, D.H. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base-pairs. *Biochemistry* **1998**, *37*, 14719–14735. [CrossRef]
11. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415. [CrossRef]
12. Neupane, K.; Foster, D.A.; Dee, D.R.; Yu, H.; Wang, F.; Woodside, M.T. Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* **2016**, *352*, 239–242. [CrossRef]
13. Treiber, D.K.; Williamson, J.R. Exposing the kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* **1999**, *9*, 339–345. [CrossRef]
14. Johnson, S.M.; Wiseman, R.L.; Sekijima, Y.; Green, N.S.; Adamski-Werner, S.L.; Kelly, J.W. Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: A focus on the transthyretin amyloidoses. *Acc. Chem. Res.* **2005**, *38*, 911–921. [CrossRef] [PubMed]
15. Manosas, M.; Junier, I.; Ritort, F. Force-induced misfolding in RNA. *Phys. Rev. E* **2008**, *78*, 061925. [CrossRef] [PubMed]
16. Russell, R. RNA misfolding and the action of chaperones. *Front. Biosci.* **2008**, *13*, 1–20. [CrossRef] [PubMed]
17. Rodgers, M.L.; Woodson, S.A. A roadmap for rRNA folding and assembly during transcription. *Trends Biochem. Sci.* **2021**, *46*, 889–901. [CrossRef]
18. Jain, A.; Vale, R.D. RNA phase transitions in repeat expansion disorders. *Nature* **2017**, *546*, 243–247. [CrossRef]
19. Błaszczyk, L.; Rypniewski, W.; Kiliszek, A. Structures of RNA repeats associated with neurological diseases. *Wiley Interdiscip. Rev. RNA* **2017**, *8*, e1412. [CrossRef]
20. Zhao, X.; Usdin, K. (Dys) function follows form: Nucleic acid structure, repeat expansion, and disease pathology in FMR1 disorders. *Int. J. Mol. Sci.* **2021**, *22*, 9167. [CrossRef]
21. Shashkova, S.; Leake, M.C. Single-molecule fluorescence microscopy review: Shedding new light on old problems. *Biosci. Rep.* **2017**, *37*, BSR20170031. [CrossRef]
22. Meller, A.; Nivon, L.; Branton, D. Voltage-driven DNA translocations through a nanopore. *Phys. Rev. Lett.* **2001**, *86*, 3435. [CrossRef]

23. Neuman, K.C.; Nagy, A. Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* **2008**, *5*, 491. [CrossRef] [PubMed]

24. Ashkin, A.; Mourou, G.; Strickland, D. The 2018 Nobel Prize in Physics: A gripping and extremely exciting tale of light. *Curr. Sci.* **2018**, *115*, 1844.

25. Junier, I.; Mossa, A.; Manosas, M.; Ritort, F. Recovery of free energy branches in single molecule experiments. *Phys. Rev. Lett.* **2009**, *102*, 070602. [CrossRef] [PubMed]

26. Tinoco, I.; Li, P.T.; Bustamante, C. Determination of thermodynamics and kinetics of RNA reactions by force. *Q. Rev. Biophys.* **2006**, *39*, 325–360. [CrossRef] [PubMed]

27. Bustamante, C.; Liphardt, J.; Ritort, F. The nonequilibrium thermodynamics of small systems. *Phys. Today* **2005**, *58*, 43. [CrossRef]

28. Bosco, A.; Camunas-Soler, J.; Ritort, F. Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Res.* **2014**, *42*, 2064–2074. [CrossRef]

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

30. Viader-Godoy, X.; Pulido, C.; Ibarra, B.; Manosas, M.; Ritort, F. Cooperativity-dependent folding of single-stranded DNA. *Phys. Rev. X* **2021**, *11*, 031037. [CrossRef]

31. Best, R.B.; Paci, E.; Hummer, G.; Dudko, O.K. Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules. *J. Phys. Chem. B* **2008**, *112*, 5968–5976. [CrossRef]

32. Dudko, O.K.; Hummer, G.; Szabo, A. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15755–15760. [CrossRef]

33. Huguet, J.M.; Bizarro, C.V.; Forns, N.; Smith, S.B.; Bustamante, C.; Ritort, F. Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15431–15436. [CrossRef] [PubMed]

34. Huguet, J.M.; Ribezzi-Crivellari, M.; Bizarro, C.V.; Ritort, F. Derivation of nearest-neighbor DNA parameters in magnesium from single molecule experiments. *Nucleic Acids Res.* **2017**, *45*, 12921–12931. [CrossRef] [PubMed]

35. Rissone, P.; Bizarro, C.V.; Ritort, F. Stem–loop formation drives RNA folding in mechanical unzipping experiments. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2025575119. [CrossRef] [PubMed]

36. Liphardt, J.; Onoa, B.; Smith, S.B.; Tinoco, I., Jr.; Bustamante, C. Reversible unfolding of single RNA molecules by mechanical force. *Science* **2001**, *292*, 733–737. [CrossRef]

37. Chen, S.J.; Dill, K.A. RNA folding energy landscapes. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 646–651. [CrossRef]

38. Woodson, S.A. Compact intermediates in RNA folding. *Annu. Rev. Biophys.* **2010**, *39*, 61–77. [CrossRef]

39. Liphardt, J.; Dumont, S.; Smith, S.B.; Tinoco, I., Jr.; Bustamante, C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* **2002**, *296*, 1832–1835. [CrossRef]

40. Alemany, A.; Mossa, A.; Junier, I.; Ritort, F. Experimental free-energy measurements of kinetic molecular states using fluctuation theorems. *Nat. Phys.* **2012**, *8*, 688. [CrossRef]

41. Koch, S.J.; Shundrovsky, A.; Jantzen, B.C.; Wang, M.D. Probing protein-DNA interactions by unzipping a single DNA double helix. *Biophys. J.* **2002**, *83*, 1098–1105. [CrossRef]

42. Camunas-Soler, J.; Alemany, A.; Ritort, F. Experimental measurement of binding energy, selectivity, and allostery using fluctuation theorems. *Science* **2017**, *355*, 412–415. [CrossRef]

43. Suren, T.; Rutz, D.; Mößmer, P.; Merkel, U.; Buchner, J.; Rief, M. Single-molecule force spectroscopy reveals folding steps associated with hormone binding and activation of the glucocorticoid receptor. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 11688–11693. [CrossRef] [PubMed]

44. Bizarro, C.V.; Alemany, A.; Ritort, F. Non-specific binding of Na+ and Mg2+ to RNA determined by force spectroscopy methods. *Nucleic Acids Res.* **2012**, *40*, 6922–6935. [CrossRef]

45. Calladine, C.R.; Drew, H. *Understanding DNA: The Molecule and How It Works*; Academic Press: Cambridge, MA, USA, 1997.

46. Saenger, W. *Principles of Nucleic Acid Structure*; Springer: Berlin/Heidelberg, Germany, 1984.

47. Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M.D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006**, *34*, 564–574. [CrossRef] [PubMed]

48. DeVoe, H.; Tinoco, I., Jr. The stability of helical polynucleotides: Base contributions. *J. Mol. Biol.* **1962**, *4*, 500–517. [CrossRef]

49. Crothers, D.M.; Zimm, B.H. Theory of the melting transition of synthetic polynucleotides: Evaluation of the stacking free energy. *J. Mol. Biol.* **1964**, *9*, 1–9. [CrossRef]

50. Breslauer, K.J.; Frank, R.; Blöcker, H.; Marky, L.A. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 3746–3750. [CrossRef]

51. Goldstein, R.F.; Benight, A.S. How many numbers are required to specify sequence-dependent properties of polynucleotides? *Biopolymers* **1992**, *32*, 1679–1693. [CrossRef]

52. Licinio, P.; Guerra, J.C.O. Irreducible representation for nucleotide sequence physical properties and self-consistency of nearest-neighbor dimer sets. *Biophys. J.* **2007**, *92*, 2000–2006. [CrossRef]

53. Gray, D.M.; Tinoco, I., Jr. A new approach to the study of sequence-dependent properties of polynucleotides. *Biopolymers* **1970**, *9*, 223–244. [CrossRef]

54. Mathews, D.H.; Sabina, J.; Zuker, M.; Turner, D.H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **1999**, *288*, 911–940. [CrossRef]

55. Walter, A.E.; Turner, D.H.; Kim, J.; Lyttle, M.H.; Müller, P.; Mathews, D.H.; Zuker, M. Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9218–9222. [CrossRef] [PubMed]

56. Freier, S.M.; Kierzek, R.; Jaeger, J.A.; Sugimoto, N.; Caruthers, M.H.; Neilson, T.; Turner, D.H. Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 9373–9377. [CrossRef] [PubMed]

57. Severino, A.; Monge, A.M.; Rissone, P.; Ritort, F. Efficient methods for determining folding free energies in single-molecule pulling experiments. *J. Stat. Mech. Theory Exp.* **2019**, *2019*, 124001. [CrossRef]

58. Woodside, M.T.; Anthony, P.C.; Behnke-Parks, W.M.; Larizadeh, K.; Herschlag, D.; Block, S.M. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* **2006**, *314*, 1001–1004. [CrossRef]

59. Forns, N.; de Lorenzo, S.; Manosas, M.; Hayashi, K.; Huguet, J.M.; Ritort, F. Improving signal/noise resolution in single-molecule experiments using molecular constructs with short handles. *Biophys. J.* **2011**, *100*, 1765–1774. [CrossRef]

60. Bustamante, C.; Marko, J.; Siggia, E.; Smith, S. Entropic elasticity of λ-phage DNA. *Proc. Nati. Acad. Sci. USA* **1991**, *88*, 10009. [CrossRef]

61. Lipfert, J.; Skinner, G.M.; Keegstra, J.M.; Hensgens, T.; Jager, T.; Dulin, D.; Köber, M.; Yu, Z.; Donkers, S.P.; Chou, F.C.; et al. Double-stranded RNA under force and torque: Similarities to and striking differences from double-stranded DNA. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15408–15413. [CrossRef]

62. Tan, Z.J.; Chen, S.J. RNA helix stability in mixed Na^+/Mg^{2+} solution. *Biophys. J.* **2007**, *92*, 3615–3632. [CrossRef]

63. Owczarzy, R.; Moreira, B.G.; You, Y.; Behlke, M.A.; Walder, J.A. Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* **2008**, *47*, 5336–5353. [CrossRef]

64. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding nucleic acid–ion interactions. *Annu. Rev. Biochem.* **2014**, *83*, 813–841. [CrossRef]

65. Markham, N.R.; Zuker, M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* **2005**, *33*, W577–W581. [CrossRef] [PubMed]

66. Markham, N.; Zuker, M.; Keith, J. UNAFold: Software for nucleic acid folding and hybridization. *Bioinformatics* **2008**, *2*, 3–31.

67. Montanari, A.; Mézard, M. Hairpin formation and elongation of biomolecules. *Phys. Rev. Lett.* **2001**, *86*, 2178. [CrossRef] [PubMed]

68. Dessinges, M.N.; Maier, B.; Zhang, Y.; Peliti, M.; Bensimon, D.; Croquette, V. Stretching single stranded DNA, a model polyelectrolyte. *Phys. Rev. Lett.* **2002**, *89*, 248102. [CrossRef]

69. Viader-Godoy, X.; Manosas, M.; Ritort, F. Sugar-Pucker Force-Induced Transition in Single-Stranded DNA. *Int. J. Mol. Sci.* **2021**, *22*, 4745. [CrossRef]

70. Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S.B.; Tinoco, I.; Bustamante, C. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* **2005**, *437*, 231–234. [CrossRef]

71. De Lorenzo, S.; Ribezzi-Crivellari, M.; Arias-Gonzalez, J.R.; Smith, S.B.; Ritort, F. A temperature-jump optical trap for single-molecule manipulation. *Biophys. J.* **2015**, *108*, 2854–2864. [CrossRef]

72. Gupta, A.N.; Vincent, A.; Neupane, K.; Yu, H.; Wang, F.; Woodside, M.T. Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nat. Phys.* **2011**, *7*, 631–634. [CrossRef]