

Cooperativity-Dependent Folding of Single-Stranded DNA

X. Viader-Godoy^{1,‡}, C. R. Pulido,² B. Ibarra^{1,†}, M. Manosas^{1,†} and F. Ritort^{1,*}

¹*Small Biosystems Lab, Departament de Física de la Matèria Condensada, Facultat de Física, Universitat de Barcelona, Carrer de Martí i Franquès, 1, 08028 Barcelona, Spain*

²*Instituto Madrileño de Estudios Avanzados en Nanociencia, IMDEA Nanociencia, 28049 Madrid, Spain*

 (Received 12 December 2020; revised 18 May 2021; accepted 16 June 2021; published 17 August 2021)

The folding of biological macromolecules is a fundamental process of which we lack a full comprehension. Mostly studied in proteins and RNA, single-stranded DNA (ssDNA) also folds, at physiological salt conditions, by forming nonspecific secondary structures that are difficult to characterize with biophysical techniques. Here, we present a helix-coil model for secondary-structure formation, where ssDNA bases are organized in two different types of domains (compact and free). The model contains two parameters: the energy gain per base in a compact domain, ϵ , and the cooperativity related to the interfacial energy between different domains, γ . We test the ability of the model to quantify the formation of secondary structure in ssDNA molecules mechanically stretched with optical tweezers. The model reproduces the experimental force-extension curves in ssDNA of different molecular lengths and varying sodium and magnesium concentrations. Salt-correction effects for the energy of compact domains and the interfacial energy are found to be compatible with those of DNA hybridization. The model also predicts the folding free energy and the average size of domains at zero force, finding good agreement with secondary-structure predictions by MFold. We envision the model could be further extended to investigate native folding in RNA and proteins.

DOI: [10.1103/PhysRevX.11.031037](https://doi.org/10.1103/PhysRevX.11.031037)

Subject Areas: Biological Physics, Soft Matter,
Statistical Physics

I. INTRODUCTION

Single-stranded nucleic acids participate in a myriad of processes [1,2]. Bases in single-stranded DNA (ssDNA) and ssRNA tend to form base pairs of different geometries and stabilities, from the most stable Watson-Crick to the weaker Hoogsteen or wobble base pairs [3]. Base pairs are stabilized by hydrogen bonds and stacking interactions [4,5] that induce high-order tertiary structures [6,7]. Biologically active ssRNAs fold into compact domains, stabilized by base-pairing interactions, connected by flexible unpaired regions facilitating RNA conformational changes [8]. Although DNA is found in a double-stranded form in the cell genome, ssDNA is generated as an essential intermediary during every aspect of DNA metabolism, replication [9–11], transcription [12], and repair [13–15]. The tendency

of ssDNA to form secondary structures interferes with and modulates the action of DNA processing enzymes.

Understanding how molecular interactions drive molecular folding is a fundamental problem in biology, from RNA [16] to proteins [17] but also in naturally occurring [18,19] and synthetic [20] DNA structures. After decades of research, conflicting views on how nucleic acids and proteins fold remain [19,21–25]. Besides the key biological role of ssDNA, understanding how ssDNA folds might shed light on the general mechanisms behind molecular folding in nucleic acids and proteins due to the ubiquitous presence of hydrogen-bonded and stacking interactions.

Secondary-structure formation in nucleic acids can be investigated with pulling experiments with optical and magnetic tweezers [26], which allow stretching individual molecules by applying mechanical forces in the piconewton range and measuring their force-extension curves (FECs) [27–29]. These experiments show that ssDNA behaves as a semiflexible polymer, with FECs described using ideal elastic polymer models [such as the wormlike chain (WLC) model [30]]. However, at low forces ($\lesssim 15$ pN), FECs deviate from the ideal elastic behavior, showing consistent compaction in the form of a force shoulder [31,32]. This compaction indicates that ssDNA folds via secondary-structure formation quantified by the reduction in the molecular extension.

We introduce a helix-coil model for ssDNA folding. In the model, bases along the ssDNA chain belong to two

*Corresponding author.
fritort@gmail.com

†Corresponding author.
mmanosas@gmail.com

‡Present address: Department of Physics and Astronomy, University of Padova, via Marzolo 8, 35131 Padova, Italy.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

types of domains, compact (C) and free (F). C domains contain bases forming secondary structure, whereas F domains contain unfolded free bases. Originally proposed by Zimm-Bragg and Lifson for protein folding [33,34], the helix-coil model has been extended to DNA hybridization [35–39] and RNA folding [40]. Surprisingly, though, and to the best of our knowledge, the helix-coil model has never been used before for nonspecific secondary-structure formation in DNA and RNA.

We perform pulling experiments with optical tweezers on individual ssDNA molecules over a wide range of experimental conditions, varying the molecular length and ionic strength, to test our model. The model reproduces the FECs of individual ssDNA molecules over 2–3 decades of molecular lengths and sodium and magnesium salt concentrations. In comparison with more complex models [31,41,42], the one presented here is described in terms of two parameters: the energy per base forming C domains (ϵ) and the interfacial energy between adjacent domains (cooperativity γ). Previous studies address the effect of salt on secondary-structure formation in ssDNA [28,32], yet the effect of molecular length has never been addressed. The helix-coil model permits us to determine the finite-size corrections to the free energy of formation of folded ssDNA and the average size of domains at zero force. Despite its simplicity, the model is compatible with MFold secondary-structure predictions, showing its power for predicting cooperativity-dependent folding of ssDNA.

II. MATERIALS AND METHODS

A. Optical tweezers setup

Experiments are carried out using a miniaturized dual-beam optical tweezers setup described in Refs. [43,44]. Briefly, two tightly focused counterpropagating laser beams ($P = 200$ mW, $\lambda = 845$ nm) create a single optical trap. A DNA molecule is tethered between two micron-sized polystyrene beads, one captured in the optical trap and the other held on top of a glass micropipette [Fig. 1(a)]. dsDNA handles are labeled with biotin or digoxigenin (see Sec. S1 in Supplemental Material [45] for details) to bind selectively to streptavidin- ($2.1 \mu\text{m}$ Kisker Biotech) and antidigoxigenin-coated beads ($3.0\text{--}3.4 \mu\text{m}$ Kisker Biotech). The experiments are performed in a microfluidics chamber. The force exerted on the optically trapped bead is determined by directly measuring the change of light momentum using position-sensitive detectors (PSDs). The position of the optical trap is determined by diverting approximately 8% of each laser beam to a secondary PSD. The instrument has a resolution of 0.1 pN and 1 nm at a 1 kHz acquisition rate.

B. DNA substrates

ssDNA easily adsorbs nonspecifically on surfaces, which makes it difficult to manipulate and to control the number

of ssDNA nucleotides that are effectively stretched. In order to avoid this problem, we use DNA hairpins that are unfolded mechanically to generate ssDNA (see Sec. II D for details). In addition, the position at which mechanical unwinding of the hairpin starts can be used as a fiducial marker for alignment of the force-extension curves (see below). We use eight hairpins, with different sequences and lengths spanning from approximately 120 b to approximately 14 kb, named H_N , N being the number of ssDNA bases released by the hairpin unfolding. We perform two different preparations for the DNA substrates (see Fig. S1 in Supplemental Material [45] for details). The shorter hairpins (120 and 204 b) are synthesized by annealing and ligating a different set of oligonucleotides (see Sec. S2 in Supplemental Material [45] for their sequences). The longer hairpins (from 700 to 13680 b) are synthesized using a long DNA fragment obtained either by a polymerase chain reaction amplification or by a digestion of the linearized λ -phage DNA. All hairpins are annealed to 29 bp dsDNA handles that are used as spacers. Labeling of the handles is achieved by a digoxigenin or biotin tailing using a terminal transferase (see Sec. S1 in Supplemental Material [45] for details).

C. Buffers

Experiments with ssDNA of different lengths (Figs. 1 and 3) are performed at 10 mM Tris-HCl pH 7.5, 0.01% NaN_3 , and 10 mM MgCl_2 . For varying NaCl and MgCl_2 concentration, previous data from our lab [32] are used: (i) sodium buffer (10 mM Tris-HCl pH 7.5, ethylenediamine tetra-acetic acid 1 mM, 0.01% NaN_3) and NaCl concentrations (10, 25, 50, 100, 250, 500, and 1000 mM) and (ii) magnesium buffer (10 mM Tris-HCl pH 7.5, 0.01% NaN_3) and MgCl_2 concentrations (0.5, 1, 2, 4, and 10 mM). To determine the ideal elastic response of ssDNA, we also perform tests in glyoxal at 1M concentration (see the Appendix B). All data are obtained at 25°C.

D. ssDNA generation and determination of its molecular extension

Several methods have been developed for obtaining ssDNA [46–49]. In this work, ssDNA substrate is generated from a DNA hairpin using the blocking oligonucleotide method described in Ref. [32]. Initially, the DNA hairpin is tethered between the two polystyrene beads. An oligonucleotide of approximately 25–30 b, complementary to the loop region of the hairpin, is flowed into the chamber (see Table S2 in Supplemental Material [45] for the sequences). Above 15 pN, the hairpin is fully unzipped [Fig. 1(b)], allowing the annealing of the complementary oligonucleotide to the exposed hairpin-loop region. Reannealing of the hairpin at forces lower than 15 pN is prevented by the annealed oligonucleotide, which works as a kinetic block. Then, the excess of oligonucleotide is washed out of the chamber to prevent any possible interference with the secondary structure [32]. Force-distance curves are

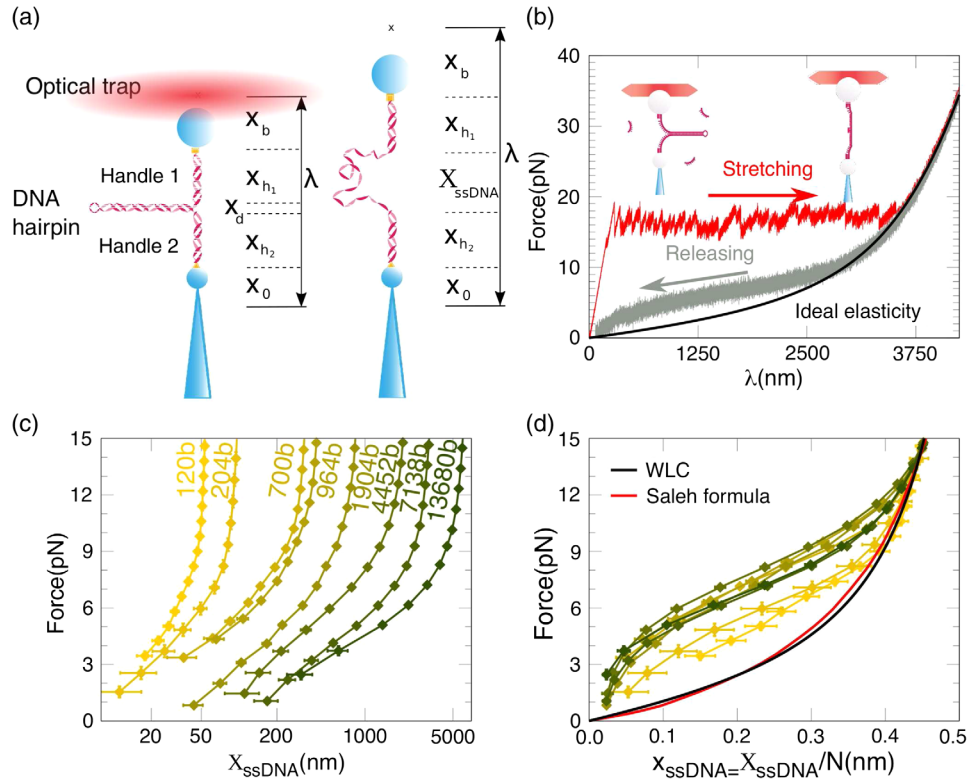


FIG. 1. Experimental setup and ssDNA force-extension curves. (a) Sketch of the experimental setup for generating ssDNA. A DNA hairpin, bearing a functionalized dsDNA handle at each strand, is tethered between two micron-sized beads; by moving the optical trap, the molecule is stretched, and the unfolding of the hairpin is observed at $f \sim 15$ pN. (b) Stretching (red) and releasing (gray) force-distance curves of the hairpin H_{7138} in the presence of an oligonucleotide complementary to the loop region. Binding of the oligonucleotide blocks the re-zipping of the hairpin, leading to the generation of ssDNA (Sec. II D) and secondary-structure formation (gray curve). The black line shows the elastic WLC model (Sec. II E). (c) ssDNA FECs obtained for the different hairpins (lengths ranging from 120 to approximately 14000 b, from yellow to dark green) by converting trap distance (λ) into molecular extension (X_{ssDNA}) (Sec. II D). The x axis is represented in log scale. (d) FECs in (c) normalized by the number of bases (N). The color code is as in (c). The secondary structure emerges at forces below approximately 10 pN, as a shoulder deviating from the ideal elasticity, as predicted by the WLC [Eq. (2)] (black line) and Saleh's formula [Eq. (3)] (red line). For $N \gtrsim 500$, FECs collapse into a single curve. The force shoulder for the 204 and 120 b molecules is systematically lower, showing finite-length effects. Error bars are obtained by averaging 5–10 releasing cycles of 5–10 different tethers for each hairpin. Experiments are performed at 25°C in 10 mM MgCl₂.

recorded by repeatedly moving the optical trap up and down at approximately 100 nm/s between a maximum and a minimum force (approximately 1 and approximately 40 pN) and measuring the force and the distance λ between the trap center and the tip of the micropipette [Fig. 1(b)]. About 5–10 stretching (force increasing) and releasing (force decreasing) curves are recorded for several tethers (approximately 5–10 in each condition). At low forces (approximately 1 pN), the ssDNA attaches nonspecifically to the bead surface. However, when low forces are avoided (forces above approximately 1 pN), no hysteresis is observed between stretching and releasing curves, showing that pulling curves are reversible.

The molecular extension of ssDNA $X_{\text{ssDNA}}(f)$ is related to the extension measured experimentally, λ [Fig. 1(a)], as follows:

$$X_{\text{ssDNA}}(f) = \lambda(f) - x_h(f) - x_b(f) - x_0, \quad (1)$$

where $x_h(f) = x_{h1}(f) + x_{h2}(f)$ is the extension of the two 29 bp handles (characterized in Ref. [50]), x_0 is an arbitrary shift of the trap position (see Appendix A 1 for details), and $x_b(f)$ is the displacement of the bead from the center of the optical trap. The latter is given by $x_b = f/k_b$ with the trap stiffness k_b measured with a power spectrum analysis (see Sec. S3 in Supplemental Material [45] for details). The force plotted versus the extension X_{ssDNA} defines the FEC.

The binding of the oligonucleotide to the loop region introduces a correction to Eq. (1) due to the hybridized dsDNA segment. We verify that this correction is negligible for molecules of 200 b or larger (see Sec. S4 in Supplemental Material [45] for details). However, for the shortest 120 b hairpin, the correction would be significant. Therefore, we use an alternative approach, the so-called two-branches method [51]. This method implies the use of a longer loop (20 b) to slow down the loop formation process, allowing measuring the ssDNA extension between

approximately 3–4 and approximately 15 pN (see Appendix A 2). The two-branches method compares differences in extension between the folded and unfolded branches, without considering corrections due to the optically trapped bead and the handles. Instead, in the blocking oligonucleotide method, bead and handle extensions must be subtracted, leading to higher uncertainties.

The FECs of the ssDNA for the eight different molecules studied are shown in Fig. 1(c). The FECs are scaled by dividing the measured extension over the number of nucleotides of each sequence, $x_{\text{ssDNA}} = X_{\text{ssDNA}}/N$, as shown in Fig. 1(d).

E. Elastic models

1. Wormlike chain

Different polymer models have been proposed to describe the ideal elasticity (i.e., without secondary structure) of nucleic acids [27,52,53]. A widely used model is the wormlike chain (WLC) model with the following interpolation formula between the low- and high-force regimes [52]:

$$f(x) = \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{Nl} \right)^{-2} - \frac{1}{4} + \frac{x}{Nl} \right], \quad (2)$$

where x is the polymer extension at a given applied force f , l is the contour length per base, p is the persistence length, and N is the number of monomers of the ideal chain (with the total contour length given by $L = N \cdot l$). Here, we use Eq. (2) to reproduce the ideal ssDNA elastic behavior at 10 mM MgCl₂ (Fig. 3), with elastic parameters taken from Ref. [32] ($p = 0.75$ nm, $l = 0.69$ nm).

Previous to fitting the data to the helix-coil model, the measured ssDNA extensions are corrected by $\lesssim 5\%$ to match the WLC model at $f = 15$ pN (see Table S4 in Supplemental Material [45] for the used factors).

2. Saleh formula

Previous studies show that, at low forces, the measured ssDNA FECs deviate from the WLC predictions [54,55]. These deviations are mainly due to excluded volume effects, and they become specially relevant at low salt concentrations. Saleh and collaborators propose different formulas for the regimes of low and high forces in sodium conditions [54–56]: At low forces, the molecular extension shows a power law dependence with force due to the excluded volume effects; at high forces, the elastic response is described by a WLC model with a salt-dependent internal electrostatic tension; finally, a logarithmic dependence interpolates the two force regimes (intermediate force regime). We combine these results in a single elastic response function covering the three force regimes (low, intermediate, and high). The fitting function is denoted as the Saleh-interpolating formula and is given by [54–56]

$$x(f) = \begin{cases} l_1 \left(\frac{f}{f_1} \right)^\gamma, & \text{if } f \leq f_1; \\ l_1 [\gamma \log(\frac{f}{f_1}) + 1], & \text{if } f_1 < f < f_2; \\ l_2 \left(\sqrt{\frac{k_B T}{4l_p(f+f_{el})}} \right), & \text{if } f_2 \leq f \end{cases} \quad (3)$$

with $f_1 = a(1000c)^\gamma$, $a = 0.06$ pN, $\gamma = 0.54$, $l_2 = 0.7$ nm, $l_1 = l_2 / \{2[\gamma \log(f_2/f_1) + 1]\}$ (continuity at f_2), and $f_2 = (k_B T/l_p) - f_{el}$. c is the salt concentration (in M). The electrostatic tension f_{el} reads

$$f_{el} = \frac{k_B T l_B}{b^2} \left[\kappa b \frac{e^{-\kappa b}}{1 - e^{-\kappa b}} - \log(1 - e^{-\kappa b}) \right], \quad (4)$$

with an average charge length $b = 1.14$ nm [54]. l_B and κ are the Bjerrum length and the inverse of the Debye length, respectively.

Note that the dependencies described in Eq. (3) have been tested only in sodium conditions. However, Eq. (3) can be extrapolated to magnesium by using the 100:1 salt rule for the equivalence of the screening ion effect of monovalent and divalent ions [32,57]. In Appendix B, we show results of the ssDNA elasticity in the presence of glyoxal (which removes secondary structure). At high salts, the WLC and the Saleh formula are compatible with the experimental data; however, there are differences at low salts, where the Saleh formula reproduces the data better. Consequently, we use the Saleh formula [Eq. (3)] to describe the ideal ssDNA elasticity at different salt conditions (Fig. 4). Comparison of the results for the fitting parameters of the helix-coil model, ϵ and γ , using the WLC and the Saleh formula are shown in Fig. S4 in Supplemental Material [45].

The measured ssDNA extensions are corrected by $\lesssim 5\%$ to match the Saleh formula model at $f = 15$ pN (see Tables S5 and S6 in Supplemental Material [45] for the used factors).

F. Helix-coil model for secondary structure

The FECs of individual ssDNA molecules deviate from the ideal polymer behavior due to the formation of secondary structures at forces below approximately 15 pN [Figs. 1(b)–1(d)]. Here, we develop a helix-coil cooperative model that describes the formation of secondary structures and explains the experimental FECs within the full force range under study (0–40 pN).

We envision the ssDNA as forming a sequence of structured compact domains interspersed by chains of free monomers (Fig. 2). These two domain types are denoted by C (compact) or F (free chain). The ssDNA is then modeled as a chain of N monomers (bases) that can be in two states $\sigma_i = \pm 1$: If $\sigma_i = -1$, the monomer is part of a C domain; otherwise ($\sigma_i = +1$), it belongs to an F domain. F -type domains pulled by a force f follow the WLC model or the Saleh formula (Sec. II E). C -type domains are stabilized by

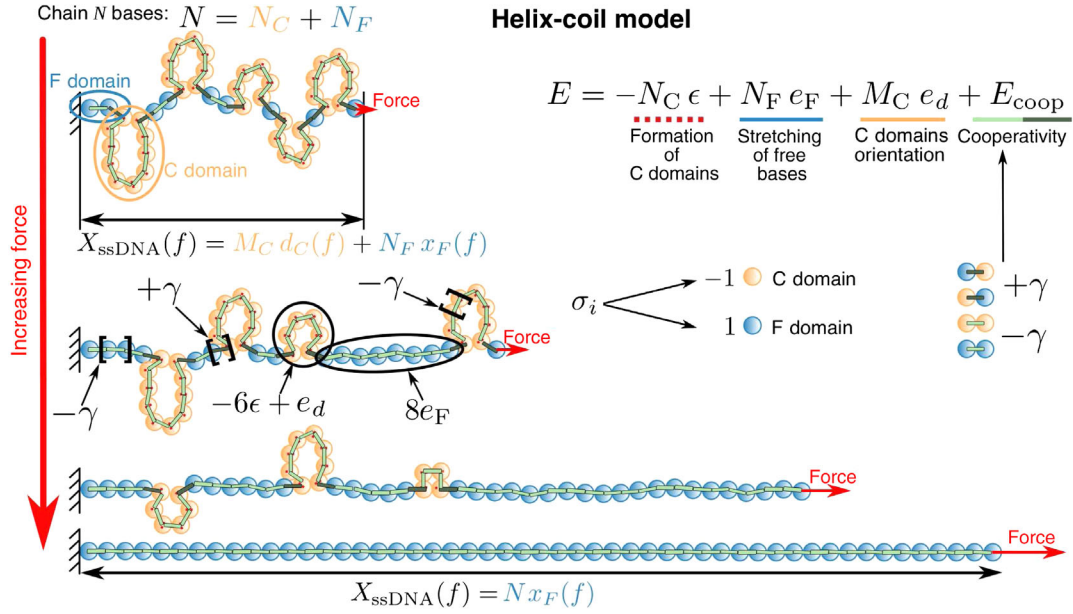


FIG. 2. Helix-coil model for secondary-structure formation. Each monomer can be either in the free state ($\sigma_i = 1$ in F domain, blue) or in the compact state ($\sigma_i = -1$ in C domain, orange). Each domain type has an elastic contribution to the energy: $e_F = -\int_0^f x_F df$, per monomer in F domain (N_F being the total monomers in F domains), and $e_d = -\int_0^f d_C df$, per C domain (M_C being the total number of C domains). A stretch of consecutive monomers in $\sigma_i = -1$ constitutes a C domain with an energy per base ϵ . The model also includes a cooperativity term (E_{coop}): Equal state adjacent monomers are energetically favored ($-\gamma$, light green) with respect to adjacent monomers in a domain wall separating C from F domains ($+\gamma$, dark green). At sufficiently low forces, the majority of the monomers are organized in C domains. As force increases, C domains unfold, gradually forming larger F domains. At high forces (approximately 15 pN), a fully stretched unfolded state is reached, where all the monomers are in the F domain.

an average energy per monomer ϵ . The model includes cooperativity through an interfacial energy parameter γ that penalizes (rewards) adjacent monomers in different (equal) domain types. Let $N_F = \sum_i^N \delta_{\sigma_i, 1}$ and $N_C = \sum_i^N \delta_{\sigma_i, -1}$ be the total number of monomers of F type and C type, respectively, fulfilling $N = N_F + N_C$. The model Hamiltonian is given by

$$\mathcal{H} = -\epsilon N_C(f) - M_C(f) \int_0^f d_C(f') df' - N_F(f) \int_0^f x_F(f') df' - \gamma \sum_i \sigma_i \sigma_{i+1}. \quad (5)$$

The first two terms account for the total energy of the compact domains pulled at a force f ; $M_C(f)$ is the number of C domains, each contributing with an extension $d_C(f)$; the third term is the stretching energy of F -type monomers with the extension per monomer $x_F(f)$, obtained by Eq. (2) or (3) [58]; and the last term stands for the interfacial energy between adjacent monomers. A schematic representation of this model is shown in Fig. 2. At a given force, N_F monomers with $\sigma = +1$ contribute by $x_F(f)$ to the total extension. The rest of the monomers, N_C , are distributed into M_C compact domains, each one contributing by $d_C(f)$ to the full extension. Hence, the total molecular extension is given by $X_{\text{ssDNA}}(f) = M_C(f) d_C(f) + N_F(f) x_F(f)$.

The Hamiltonian in Eq. (5) is analytically solvable (see Sec. II F), leading to the extension per monomer, X_{ssDNA}/N :

$$x_{\text{ssDNA}}(f) = \phi_F(f) x_F(f) + \frac{M_C(f)}{N} d_C(f), \quad (6)$$

where the fraction of free monomers, $\phi_F(f) = N_F(f)/N$, is given by [59]

$$\phi_F(f) = \frac{1}{2} \left(1 + \frac{\sinh[\beta A(f)]}{\sqrt{e^{-4\beta B(f)} + \sinh^2[\beta A(f)]}} \right), \quad (7)$$

with $\beta = 1/k_B T$. A and B quantities are defined, respectively, by

$$A(f) = -\frac{\epsilon}{2} + \frac{1}{2} \int_0^f x_F(f') df'; \quad (8)$$

$$B(f) = -\frac{1}{4} \int_0^f d_C(f') df' + \gamma. \quad (9)$$

Equation (7) can be used to obtain the fraction of bases in C domains, $\phi_C(f) = N_C(f)/N$, as

$$\phi_C(f) = 1 - \phi_F(f). \quad (10)$$

The number of C domains is given by

$$M_C(f) = \frac{N}{2} \frac{\sinh[\beta A(f)]}{\cosh[\beta A(f)] + \sqrt{e^{-4\beta B(f)} + \sinh^2[\beta A(f)]}}. \quad (11)$$

The average number of monomers per C domain reads as

$$n_C(f) = \frac{N\phi_C(f)}{M_C(f)}. \quad (12)$$

In Eq. (6), $x_F(f)$ is given by the WLC model [Eq. (2)] or the Saleh formula [Eq. (3)], with the parameters described in Sec. II E. For the extension $d_C(f)$, we consider each domain as a rigid object of length δ that aligns along the stretching direction (Appendix C). For example, a hairpinlike C domain initiating with a stem would have a length equal to the diameter of a B-DNA double helix, $\delta \sim 2$ nm [60]. We check that this parameter does not affect the predictions of the model significantly (Appendix C). Therefore, we neglect the domain orientation term ($\delta = 0$) in the analysis and fit our data to the simpler expression

$$x_{\text{ssDNA}}(f) = \phi_F(f)x_F(f). \quad (13)$$

We fit the model [Eq. (13)] to the experimental FECs with ϵ and γ as fitting parameters between $2 \leq f \leq 15$ pN. Lower forces ($f < 2$ pN) are not considered in the fit to avoid any influence of incorrect tethering (especially in molecules with $N \lesssim 1000$ b). Importantly, the model reproduces the experimental FECs throughout the whole force range $0 < f \leq 15$ pN in the different conditions tested (different lengths and salts).

III. RESULTS

A. Force-extension curves

For ssDNA measurements, a DNA hairpin is fully unzipped by applying a force above 15 pN at the 3' and 5' extremities of the molecule, exposing the loop region. Unzipping of the hairpin is performed in the presence of an oligonucleotide complementary to the loop (Sec. II D). Figure 1(b) shows the pulling curve (red) corresponding to the unzipping of the H_{7138} hairpin with the characteristic sawtooth pattern [47]. This curve is followed by the relaxation curve (gray), after the oligonucleotide hybridizes to the loop region, corresponding to the ssDNA FEC. It differs from the ideal elastic curve (WLC model shown in black) at low forces, where ssDNA compacts and the extension shortens, showing the characteristic shoulder of secondary-structure formation [32,61]. It is well known that DNA sequence (i.e., GC content) and monovalent and divalent cations (which screen the negatively charged DNA phosphate backbone) affect secondary-structure formation [28,32,62]. However, the effect of ssDNA length has never

been studied. The nature of the compact structure is an open question due to its highly disordered and dynamic features. A dependence of the FEC with the ssDNA length might be the signature of long-range interactions along the chain, giving information about the nature of the secondary structure.

B. Effects of molecular length on secondary structure

We measure the FECs of eight different ssDNA molecules [Fig. 1(c)] with lengths spanning from 120 to 13680 b in 10 mM MgCl_2 (Sec. II C) and similar GC content, approximately 50% [63]. FECs are shown in Fig. 1(d) with the extension rescaled by the number of ss bases of each molecule. Molecules with $N \gtrsim 500$ b approximately collapse into a single curve, deviating from the ideal elastic behavior at forces $f \lesssim 10$ pN and showing a remarkable shoulder. Small differences are due to different GC content (from 44% to 53%) which correlates with the height of the force shoulder observed in the FECs (see below and Appendix E). In contrast, for $N \lesssim 500$ b, FECs do not collapse into the same curve and the height of the shoulder decreases with N , indicating finite-length effects. This result might be due to the depletion of secondary-structure arrangements in the regime $N \lesssim 500$ b.

As shown in Fig. 3(a), the FECs for the studied molecules (yellow to dark green circles for increasing lengths) are well reproduced by Eq. (13) in the proposed model (green curves). The best-fitting values for the average energy per base of C domains, ϵ , and the cooperativity parameter γ as a function of the molecular length are shown in Figs. 3(b) and 3(c). ϵ is found to follow a phenomenological $1/N$ correction:

$$\epsilon(N) = \epsilon_0 - \frac{b}{N}, \quad (14)$$

with $\epsilon_0 = 0.18(1)$ kcal/mol and $b = 18(4)$ kcal/mol [Fig. 3(b)]. γ appears to be constant [$\gamma = 0.61(2)$ kcal/mol] for all ssDNA lengths [Fig. 3(c)]. Equation (14) quantifies corrections to the stabilization energy of C domains due to the free ends of the polymer. In generic statistical physics models, corrections to the equilibrium free energy are determined by the surface to volume ratios, being on the order of $1/N$ for the one-dimensional helix-coil model.

For shorter molecules ($N \lesssim 500$ b), ϵ starts deviating from the saturation value ϵ_0 . We stress that the FECs in Fig. 3(a) cannot be fitted to the model with $\gamma = 0$, as shown in Fig. S5 in Supplemental Material [45]. Therefore, cooperativity is key to reproduce the experimental results, suggesting the suitability of the helix-coil model.

C. Salt dependence of secondary structure

In order to investigate how monovalent and divalent ions affect ssDNA folding, we fit the helix-coil model to previously published data by us of a ssDNA molecule

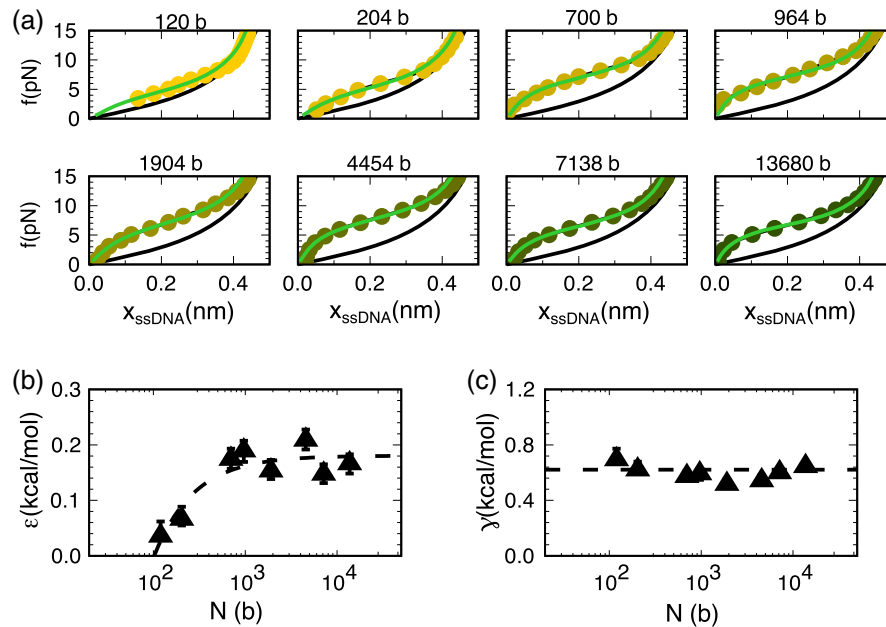


FIG. 3. ssDNA FECs for varying molecular length. (a) Experimental FECs (circles) and the corresponding fits of the helix-coil model [Eq. (13)] (green lines) for ssDNA molecules with lengths ranging from approximately 120 to approximately 14000 b [color code as in Figs. 1(c) and 1(d)]. The WLC model [Eq. (2)] is shown as a black line [Saleh’s formula giving the same results at high salt; see Fig. 1(d)]. Fits give the energy parameters ϵ and γ for each molecular length. (b),(c) Average formation energy per base in C domain, ϵ (b), and cooperativity parameter, γ (c), obtained from the fits shown in (a) as a function of the ssDNA molecular length (triangles). Dashed lines are fits to Eq. (14) [$\epsilon_0 = 0.18(1)$ kcal/mol and $b = 18(4)$ kcal/mol] and to a constant value [$\gamma = 0.61(2)$ kcal/mol]. Experiments are performed at 25°C in 10 mM MgCl_2 . Error bars are the statistical errors.

(H_{13680} , approximately 50% GC content) under different NaCl and MgCl_2 concentrations [32]. FECs for H_{13680} are shown in Figs. 4(a) (NaCl) and 4(b) (MgCl_2). At high salts ($\gtrsim 100$ mM NaCl, $\gtrsim 0.5$ mM MgCl_2), the data (circles) deviate from the ideal elastic behavior (black curves). The height of the shoulder in the FEC reaches its maximum at the highest salt concentrations ($[\text{NaCl}] = 1000$ mM and $[\text{MgCl}_2] = 10$ mM). At low salts ($\lesssim 50$ mM NaCl), the shoulder disappears, and the FECs approach the ideal elastic behavior [32,64–67].

The helix-coil model successfully reproduces the FECs in sodium and magnesium [Figs. 4(a) and 4(b)]. For high salts (above 100 mM NaCl and 1 mM MgCl_2), the WLC model Eq. (2) and the Saleh formula Eq. (3) give compatible results. Otherwise, at low salts, the Saleh formula reproduces better the ideal elastic response of ssDNA (Appendix B), so we use it throughout the following figures. A comparison of the two models (WLC versus Saleh) for the energy parameters ϵ and γ and the FECs is shown in Figs. S4 and S6 in Supplemental Material [45]. Salt-induced stabilization of secondary structure is captured by the model as an increase in ϵ with NaCl (magenta) and MgCl_2 (blue) concentrations. Note that, for $\epsilon = 0$ and zero force, F and C domains are equally probable, meaning that half of the bases are of C and F type. ϵ values can be compared with the Watson-Crick base-pair energies of the nearest-neighbor (NN) model. Therefore, we define $\epsilon_{\text{bp}} = 2\epsilon$ as the equivalent of the NN base-pair energy for

DNA hybridization. Similarly to the energies in the NN model [47,68], we find that ϵ_{bp} depends logarithmically on salt concentration [Fig. 4(c)]:

$$\epsilon_{\text{bp}}(c) = \epsilon_{\text{bp}}^0 + m_\epsilon \log(c), \quad (15)$$

with c being the salt concentration (in M units), ϵ_{bp}^0 the value at 1M, and m_ϵ the salt-correction parameter for secondary structure. Equation (15) is a phenomenological expression that follows from applying standard physical chemistry theories of thermodynamic activity to diluted ionic solutions. Fits of ϵ_{bp} to Eq. (15) are shown in Fig. 4(c) as dashed lines. The results from the fits are $\epsilon_{\text{bp}}^0 = 0.34(3)$ kcal/mol and $m_\epsilon = 0.11(2)$ kcal/mol in sodium and $\epsilon_{\text{bp}}^0 = 0.75(9)$ kcal/mol and $m_\epsilon = 0.09(2)$ kcal/mol in magnesium. The values for ϵ_{bp}^0 are approximately 5 times smaller than the average NN base-pair energy (approximately 1.6 kcal/mol) and almost 3 times smaller than the most unstable NN bp (AT/TA ~ 0.85 kcal/mol at 1M NaCl and 10 mM MgCl_2) [47,68]. Such a lower stabilization is expected for random sequences that lack full complementarity regions. In fact, C domains might consist of aggregates of base pairs but also large loops and mismatches, which decrease base-pairing stability [69,70]. In agreement with this possibility, unzipping forces in dsDNA are approximately 15 pN [Fig. 1(b)] (see also Ref. [71]), much larger than the force at the shoulder in the FEC

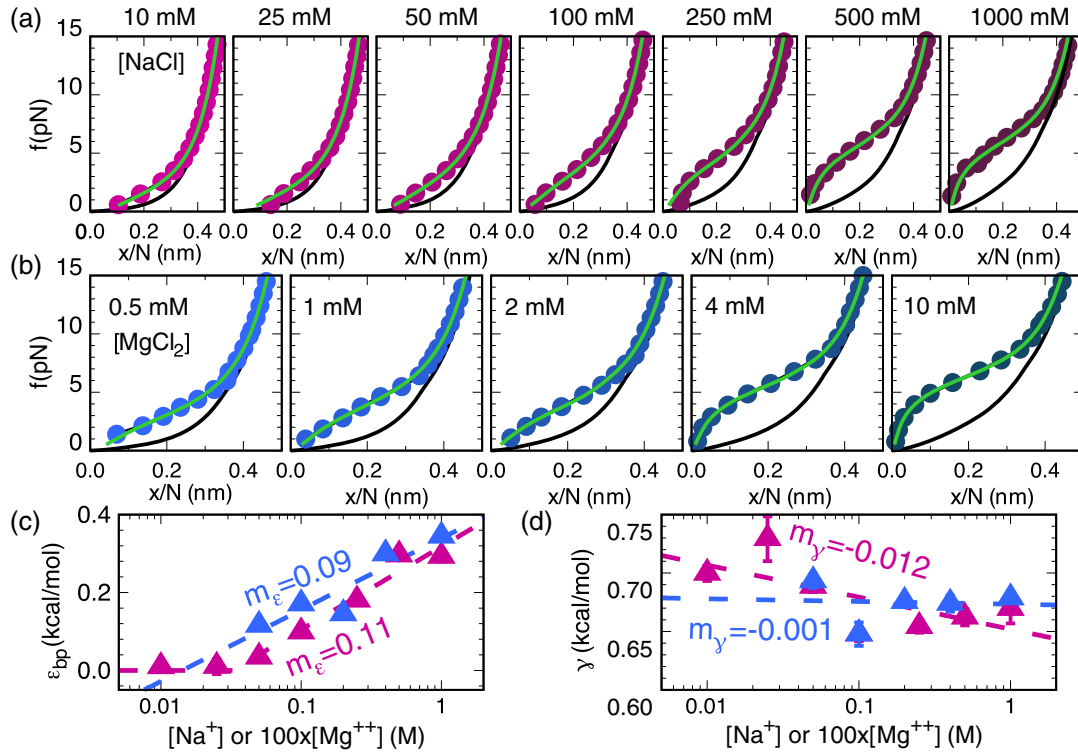


FIG. 4. ssDNA FECs for varying salt conditions. (a),(b) Fits of the helix-coil model [Eq. (13)] (green lines) to the FECs of 13680 b ssDNA at increasing NaCl concentration [magenta circles, (a)] and MgCl_2 concentration [blue circles, (b)]. The Saleh formula [Eq. (3)] is shown as a black line. Fits give the energy parameters ϵ and γ for each salt condition. (c),(d) Dependence of ϵ_{bp} ($\epsilon_{\text{bp}} = 2\epsilon$) and γ with salt concentration (magenta and blue triangles for sodium and magnesium, respectively). Data in (c) above 50 mM are fitted to Eq. (15) (dashed lines), giving $\epsilon_{\text{bp}}^0 = 0.34(2)$ kcal/mol and $m_\epsilon = 0.11(2)$ kcal/mol for NaCl (magenta) and $\epsilon_{\text{bp}}^0 = 0.75(9)$ kcal/mol and $m_\epsilon = 0.09(2)$ kcal/mol for MgCl_2 (blue). The values for ϵ below $[\text{NaCl}] \sim 50$ mM appear to saturate to 0. γ also follows a logarithmic dependence with salt concentration [Eq. (16)], with $\gamma_0 = 0.64(1)$ kcal/mol and $m_\gamma = -0.012(5)$ kcal/mol for NaCl and $\gamma_0 = 0.67(2)$ kcal/mol and $m_\gamma = -0.001(4)$ for MgCl_2 . Error bars are the statistical errors.

(approximately 5–10 pN). Interestingly, the slope $m_\epsilon = 0.11(2)$ kcal/mol [Eq. (15)] in sodium is compatible with the homogeneous salt-correction parameter of the unified oligonucleotide dataset [72] and with that derived from unzipping experiments [47] ($m = 0.114$ kcal/mol and $m = 0.104$ kcal/mol, respectively). For the MgCl_2 case, $m_\epsilon = 0.09(2)$ kcal/mol is about twice the MFold value [73] ($m = 0.055$ kcal/mol) but closer to the unzipping value [74] [$m = 0.07(2)$ kcal/mol]. Note that MFold assumes that the salt correction for sodium is exactly twice that of magnesium ($m_\epsilon^{\text{NaCl}} = 2 \times m_\epsilon^{\text{MgCl}_2}$).

Finally, for the cooperativity term γ , we also assume the phenomenological logarithmic salt dependence of Eq. (14) [Fig. 4(d)]:

$$\gamma = \gamma_0 + m_\gamma \log(c), \quad (16)$$

with c being the salt concentration (in M units), γ_0 the value at $1M$, and m_ϵ the salt-correction parameter. We get $\gamma_0 = 0.64(2)$ kcal/mol and $m_\gamma = -0.012(5)$ kcal/mol for NaCl and $\gamma_0 = 0.67(2)$ kcal/mol and $m_\gamma = -0.001(4)$ kcal/mol

for MgCl_2 . Although γ weakly depends on c ($m_\gamma < 0$) for sodium, the dependence is negligible for magnesium. This result indicates that salt screening effects for the cooperativity between adjacent bases are weak, raising the question whether it depends at all. The salt dependence of γ [Fig. 4(d)] is the same if we apply the 100th rule of thumb between sodium and magnesium concentrations, as we do for ϵ_{bp} [Fig. 4(c)]. The values of γ at the reference salt condition, γ_0 , are compatible with the average γ obtained with varying the molecular length at the same condition [$\gamma = 0.61(2)$ kcal/mol; see Sec. III B]. Notice that the switching of a base (from F to C or the opposite) involves an interfacial energy cost or gain of $\pm 2\gamma$ [Eq. (5)]. Note the opposite sign of the salt corrections for ϵ and γ . In both cases, salt screens the electrostatic interactions by weakening phosphate repulsion. This screening leads to an increased base-pairing stabilization and a (much weaker) decreased cooperativity.

A compendium of all results in this section is shown in Table I, and the fitting parameters obtained with the WLC model are shown in Table S7 in Supplemental Material [45].

TABLE I. Fitting parameters obtained for each studied condition.

Condition	Relation	Values (kcal/mol)
Length (10 mM MgCl ₂)	$\epsilon = \epsilon_0 + b/N$ $\gamma = \text{const}$	$\epsilon_0 = 0.18(1)$, $b = 18(4)$ $\gamma = 0.61(2)$
Salt	$\epsilon_{\text{bp}}(c) = \epsilon_{\text{bp}}^0 + m_\epsilon \log(c)$	$\epsilon_{\text{bp}}^0 = 0.34(3)$, $m_\epsilon = 0.11(2)$
	$\gamma(c) = \gamma_0 + m_\gamma \log(c)$	$\gamma_0 = 0.64(2)$, $m_\gamma = -0.012(5)$
	$c = [\text{NaCl}]$	$\epsilon_{\text{bp}}^0 = 0.75(9)$, $m_\epsilon = 0.09(2)$
	$c = [\text{MgCl}_2]$	$\gamma_0 = 0.67(2)$, $m_\gamma = -0.001(4)$

D. C- and F-domain organization

We use the proposed helix-coil cooperative model (Sec. II F) to investigate how bases are distributed between *F* and *C* domains. To this end, we compute the fraction of bases belonging to *C* domains at each force *f*, $\phi_C(f)$ [Eq. (10)], and the average number of bases per *C* domain, $n_C(f)$ [Eq. (12)].

Figure 5(a) shows $\phi_C(f)$ as a function of the force at different ssDNA lengths (top), NaCl (middle), and MgCl₂ (bottom) concentrations obtained from fitting the model to

the experimental data. Insets in Fig. 5(a) show ϕ_C for each condition at two different forces, $f = 0$ and 5 pN. $\phi_C(f)$ presents its maximum value at 0 pN, $\phi_C(0) = 0.5$ –0.9, depending on salt and length. $\phi_C(f)$ decreases until a threshold force is reached, f^* . Above f^* , the secondary structure does not form [$\phi_C(f^*) \sim 0$], and the experimental FEC matches the ideal elastic model prediction Eqs. (2) and (3). f^* depends on molecular length and salt concentration: For long molecules ($N \geq 500$ b) and high salts ($[\text{NaCl}] = 1000$ mM, $[\text{MgCl}_2] = 10$ mM), $f^* \sim 10$ pN,

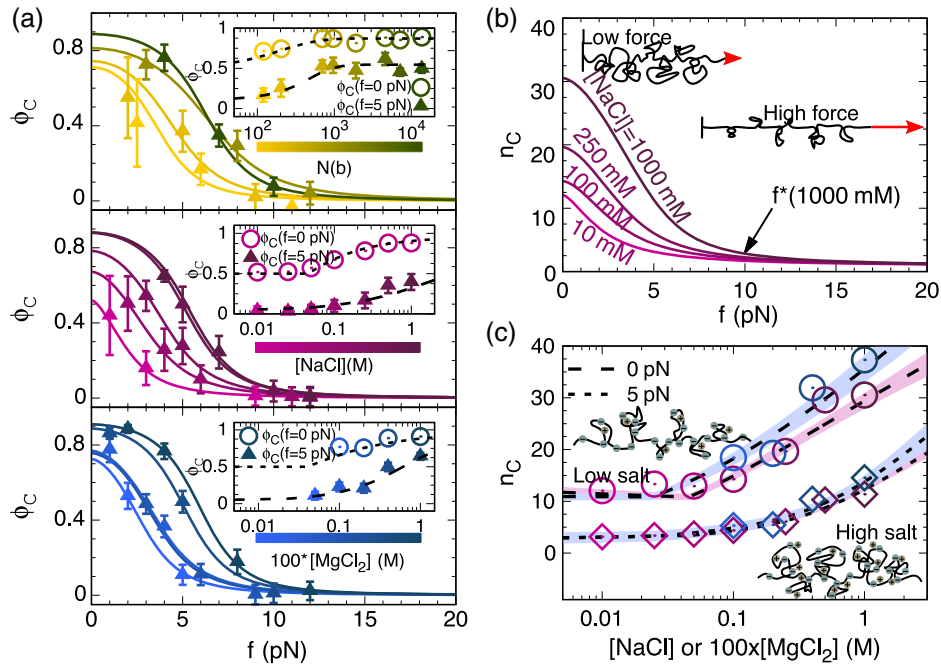


FIG. 5. C- and F-domain organization. (a) Fraction of bases in *C* domains, ϕ_C , as a function of the force, obtained from Eq. (10) using the best-fitting values for ϵ and γ [Figs. 3(b) and (c) and 4(c) and 4(d)] for (top) four different molecular lengths (120, 204, 1904, and 13680 b, from yellow to dark green), (middle) five NaCl concentrations (10, 100, 250, 500, and 1000 mM, from light to dark magenta), and (bottom) five MgCl₂ concentrations (0.5, 1, 2, 4, and 10 mM, from light to dark blue). The insets show the dependence of the fraction ϕ_C with respect to the molecular length (top) and salt concentration (middle and bottom) at two different forces, 0 and 5 pN. Dashed lines are the theoretical prediction by Eqs. (7) and (10) using salt- and length-dependent ϵ and γ parameters as given in Table I. (b) Average number of bases or size per *C* domain, n_C [Eq. (12), with salt-dependent ϵ and γ parameters given in Table I], as a function of the force for 10, 100, 250, and 1000 mM NaCl. The schematic depictions show a highly compacted form of the ssDNA at low forces that is stretched at higher forces, decreasing the size of the *C* domains. The arrow indicates the threshold force f^* . (c) n_C dependence on salt concentration (magenta for NaCl and blue for MgCl₂) at 0 and 5 pN. Symbols correspond to Eq. (12) using the best-fitting values for ϵ and γ [Figs. 4(c) and 4(d)]. Dashed lines are the theoretical prediction by Eq. (12) with salt-dependent ϵ and γ parameters given in Table I. Shaded areas show the uncertainty region of n_C , from the errors in ϵ and γ . The schematic depictions show the negatively charged ssDNA (black line) screened by the cations (gray dots) in the solvent stabilizing *C* domains.

whereas for either short molecules or low salts ($[\text{NaCl}] = 10\text{--}50$ mM, $[\text{MgCl}_2] = 0.5$ mM), $f^* \sim 5\text{--}7$ pN. Note that $\phi_C(0)$ is non-negligible, even at the lowest NaCl concentrations [10–50 mM NaCl, dashed lines in the insets in Fig. 5(a), middle]. In spite of this fact, a shoulder in the FEC is not observed [Fig. 4(a)], because ϵ is too small.

Figure 5(b) shows $n_C(f)$ versus force obtained by fitting the model to the experimental data, and a similar trend to $\phi_C(f)$ is observed: It is maximum at zero force, $n_C(0)$, and decreases as force increases, until the threshold force is reached, f^* , for which $n_C \rightarrow 1$ (the minimum value allowed by the model). The values of f^* for $\phi_C(f)$ and $n_C(f)$ are the same, indicating that secondary structure is disrupted above f^* . As shown in Fig. 5(c), n_C increases with sodium and magnesium for all forces. The values for n_C range from approximately 10 b (low salt) to approximately 30 b (high salt). Overall, the model predicts that the applied force destabilizes the secondary structure, by reducing both the size and number of C domains [sketch in Fig. 5(b)].

IV. COMPARISON OF THE MODEL WITH MFold PREDICTIONS

MFold predicts secondary structures of nucleic acids from their primary sequence. In order to validate our model, we compare its predictions at zero force with those obtained from MFold for H_{120} , H_{204} , H_{700} , H_{964} , H_{1904} , and H_{4454} [75]. For a given ssDNA sequence, MFold computes folded structures of free energy ΔG (with respect to the random coil) in a given range above the absolute minimum, defining the energy spectrum of that sequence. We analyze each folded structure in the spectrum [Fig. 6(a)] and assign nucleotides as being of F or C type depending on whether they belong to a folded hairpinlike structure (C domain) or they are part of a free linker (F domain). This analysis permits us to calculate ϕ_C and n_C , by counting the number of bases forming C domains (ϕ_C) and the number of bases per C domain (n_C).

To compare our model predictions with MFold, we derive an analytical expression for the free energy of the folded ssDNA at zero force from the helix-coil model (Appendix G):

$$\frac{\Delta G}{N} = -\frac{\epsilon}{2} - \frac{1}{\beta} \log \left[\cosh \left(\frac{\beta\epsilon}{2} \right) + \sqrt{e^{-4\beta\gamma} + \sinh^2 \left(\frac{\beta\epsilon}{2} \right)} \right]. \quad (17)$$

Equation (17) nicely matches the MFold prediction with varying N [Fig. 6(b)] and varying sodium concentration [Fig. 6(c)]. Note that there is no fitting procedure; we just use the values ϵ and γ from the FEC fits (purple triangles) and the fitted phenomenological formulas [dashed lines in Figs. 3(c) and (d) and Figs. 4(c) and 4(d)]. The MFold

prediction in Fig. 6(c) corresponds to a 1004 b random sequence with the same GC content as in H_{13680} , since the MFold analysis cannot be performed with very long sequences (larger than approximately 5000 b [72]). The same results are obtained for different sequence lengths whenever the GC content is conserved. It is important to stress the strong dependence of the free energy ΔG with the GC content (see Fig. S7 in Supplemental Material [45]), which must be taken into account for a proper comparison. The match between the MFold prediction and Eq. (17) works better when using the Saleh formula [Eq. (3)] as compared to the WLC [Eq. (2)] for the ideal ssDNA elasticity (see Fig. S7 in Supplemental Material [45]). The equivalent of Fig. 6(c) in magnesium is not predictive, because MFold uses a salt correction that is known to be inaccurate [74].

In addition, for short ssDNA molecules ($N < 1000$ b), the helix-coil model and MFold also predict similar values for $\phi_C(0)$ and $n_C(0)$ [Figs. 6(d) and 6(e)]. However, for longer ssDNA molecules, discrepancies emerge. While MFold predicts that $\phi_C(0) \rightarrow 1$ [Fig. 6(d), black squares] and $n_C(0)$ rapidly increases with N [Fig. 6(e), black squares], our model predicts that approximately 10% of all bases are free [Fig. 6(d), dashed line] and $n_C(0)$ saturates at approximately 40 b [Fig. 6(e), dashed line]. These discrepancies might be attributed to the simplicity of the helix-coil model, which does not consider sequence heterogeneity (energy parameters ϵ and γ are taken as uniform). In contrast, MFold incorporates sequence effects favoring the hybridization of nearly complementary regions, even if far apart, a feature absent in our model. This feature leads to large C -domain sizes in MFold [e.g., Fig. 6(a), right structure]. To test this assumption, we compute low-lying energy structures with MFold by constraining the maximum distance between base pairs. For 150 b distance, the values for $\phi_C(0)$ and $n_C(0)$ predicted by MFold for long molecules ($N > 1000$ b) agree with our model [red circles in Figs. 6(d) and 6(e)], supporting this interpretation. As shown in Figs. 6(f) and 6(g), the helix-coil model also reproduces the MFold predictions for $\phi_C(0)$ and $n_C(0)$ as a function of the salt concentration for short enough molecules [MFold predictions are obtained for the same 1004 b sequence as in Fig. 6(c)].

It is remarkable that, despite its simplicity, the helix-coil model compares so well with MFold. MFold is a predictive tool that uses the NN model with energy parameters derived from large datasets obtained from DNA melting experiments. The excellent agreement between our model and MFold reflects the fact that both build upon the energetics of hybridization. The relevant motifs in MFold (base pairs, loops, single-stranded segments, etc.) are also the ultimate building blocks of the compact disordered structures. Interestingly, the combined effects of these blocks are effectively described by only two parameters (ϵ and γ) in the helix-coil model.

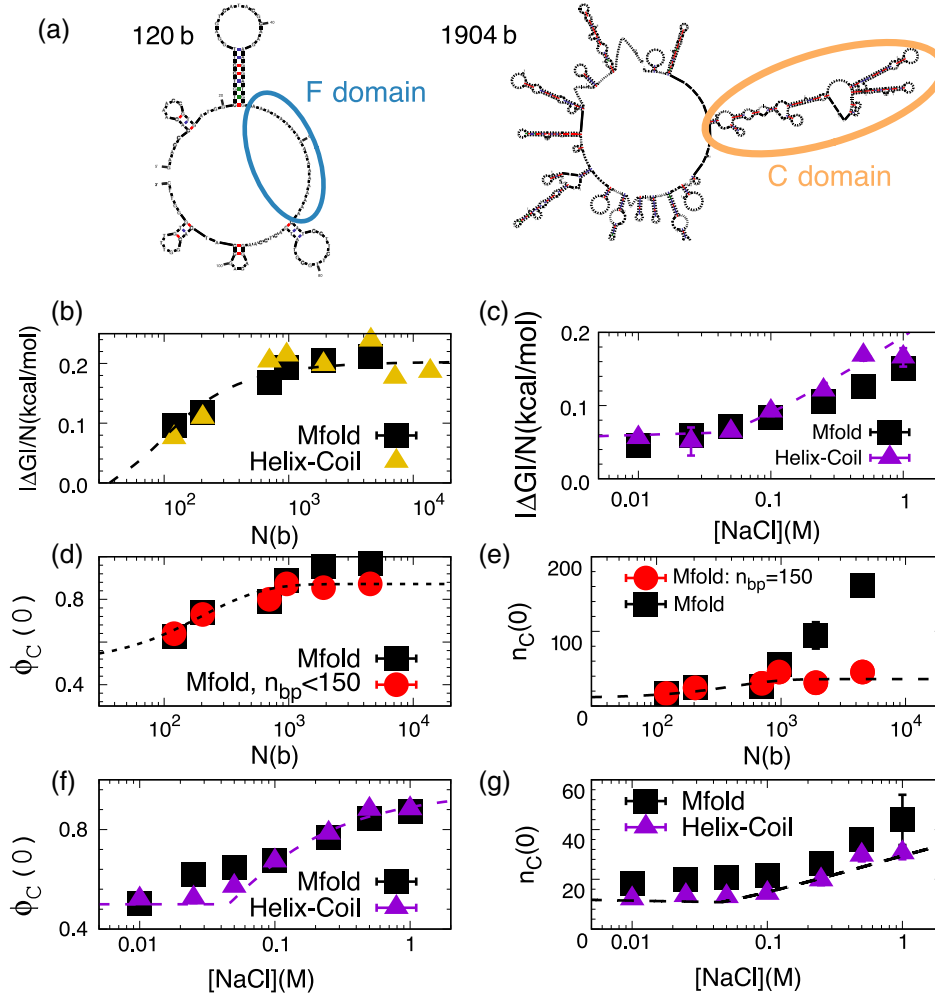


FIG. 6. Comparison between the helix-coil and mFold models for ssDNA secondary-structure formation. (a) The first excited secondary structures above the native hairpin as given by mFold for a short (120 b) and a long molecule (1904 b) at 10 mM MgCl_2 . (b) Absolute value of the free energies per base at zero force, $|\Delta G|/N$, given by mFold (black squares) and predicted by the helix-coil model [Eq. (17)] (dashed line) as a function of molecular length N . (c) The same as in (b) but a function of the sodium concentration. The structures calculated for mFold are from a 500 bp hairpin (H_{1004}) with a random sequence with the same GC content as in H_{13680} . Purple triangles correspond to the values obtained from the fits to the experimental FECs. (d),(e) Fraction of bases in C domains (d) and average number of bases n_C per C domain (e) given by mFold (red circles and black squares) and the helix-coil model (dashed line). Red circles are obtained from mFold by constraining the maximum distance between base pairing to 150 b. Errors are smaller than the size of the symbols. (f),(g) The same as in (d),(e) but as a function of the sodium concentration. As in (c), the structures of mFold are obtained for H_{1004} . The agreement between mFold and the helix-coil model predictions is, in general, very good.

V. DISCUSSION AND CONCLUSION

We have developed a cooperativity-dependent folding model that reproduces experimental FECs of ssDNA over different molecular lengths and salt conditions. The model assumes that ssDNA folds forming compact domains (C) interspersed by free bases (F). While free bases are described by the ideal elastic model, C domains are assumed to have a negligible extension. Compared to more complex theoretical models [31,76], the model presented here includes only two parameters: ϵ , the average energy per base in a C domain, and γ , the interfacial energy between adjacent domains (cooperativity term). With only

two parameters, the model reproduces the experimental FECs of ssDNA spanning two decades of contour length ($100\text{b} \lesssim N \lesssim 14\text{ kb}$) and three decades of sodium and magnesium concentrations ($10 \leq [\text{NaCl}] \leq 1000\text{ mM}$ and $0.5 \leq [\text{MgCl}_2] \leq 10\text{ mM}$). Remarkably, $\gamma = 0$ is unable to fit the data, suggesting that cooperativity is essential for ssDNA folding, as shown in Fig. S5 in Supplemental Material [45]. The simplest interpretation of the cooperativity term is the nature of the stacking interaction itself. In nucleic acids, two distinct forces stabilize a duplex: hydrogen bonding and stacking. Hydrogen bonding acts transversely to the phosphate backbone, bringing bases close to

each other. In contrast, stacking acts longitudinally along the backbone, which is the natural direction of the single-stranded polymer. It is this geometric alignment between the stacking force and the polymer direction that generates cooperativity. This alignment explains why ssDNA and ssRNA show a cooperative plateau in the FECs [49,76] and many RNAs form tertiary structures with coaxial stacking between different helices (making RNA folding a cooperative process [77,78]).

To further check the applicability of the model, we carried out pulling experiments to predict the FECs on another ssDNA (approximately 8 kb, approximately 50% GC content) obtained by unpeeling an 8 kbp dsDNA [46,48] in a different optical tweezers instrument [27]. In Appendix F, we show the predicted FEC from the model using the ϵ and γ parameters derived in our study [Figs. 3(b) and 3(c)] without any fitting procedure. The good match with the experimental data further validates our model's predictive power and the derived energy parameters.

In a previous work [32], we found a phenomenological formula for the FEC at different salt conditions. This formula included two parameters, a critical force (f^\ddagger) and a force-width parameter (δ), akin to ϵ and γ in the helix-coil model. The phenomenological formula gives comparable fits to the helix-coil model; however, the former has a limited scope at the level of physical interpretation, especially at low forces where their behaviors differ (as shown in Sec. S10 in Supplemental Material [45]).

A remarkable feature of the ssDNA FECs is the absence of force rips, which is due to the low values of ϵ ($\lesssim 0.3$ kcal/mol) compared to $k_B T$ (approximately 0.6 kcal/mol at $T = 298$ K). For a random sequence, Brownian forces smear out any force rips appearing during the formation of compact domains. This smearing gives rise to the smooth and reversible FECs observed in the experiments. For $\epsilon \gtrsim k_B T$, we should observe FECs with specific sequence-dependent force rips as is the case in unzipping experiments [Fig. 1(b)], where base-pair energies are approximately $2-4k_B T$. Still, sequence effects are observed in the height of the shoulder of the FEC that increases with GC content. Concomitantly, the value of ϵ increases, too (Appendix D).

Regarding salt dependence effects, we observe that ϵ and γ have logarithmic salt corrections. The salt-correction parameter for ϵ in sodium [$m_\epsilon = 0.11(2)$ kcal/mol] is compatible with that of dsDNA hybridization [$m = 0.11(1)$ kcal/mol]. This compatibility indicates that monovalent cations play a similar role in secondary-structure formation and dsDNA hybridization. The stabilization effect of magnesium appears to be slightly larger for ssDNA folding [$m_\epsilon = 0.09(2)$ kcal/mol] as compared to duplex hybridization [$m = 0.06(2)$ kcal/mol]. This result demonstrates the distinct role of divalent cations in bringing together distant nucleotides to fold ssDNA. Such a long-range effect

should be less prominent for duplex formation, where the stem always grows in the proximity of the hybridization junction.

Apparently, the interfacial energy γ decreases weakly with the ionic strength $|m_\gamma| \ll m_\epsilon$, raising the question whether the logarithmic dependence of Eq. (16) is the right phenomenological description. It would be very interesting to extend thermodynamic activity theories to describe salt corrections to the cooperativity term. Using the less-accurate WLC model to fit the data, a systematic salt dependence in both sodium and magnesium that is compatible with Eq. (16) is obtained (see Fig. S6 in Supplemental Material [45] for details). This result would be an indication that salt similarly screens the interaction between adjacent bases in secondary structure and dsDNA hybridization.

The model predicts the free energy of the folded structure and its nature at zero force [$\phi_C(0)$ and $n_C(0)$]. Free energies are in agreement with MFold predictions for all investigated lengths, while $\phi_C(0)$ and $n_C(0)$ are not well captured for long molecules ($N \gtrsim 1000$ b) due to the underestimation of large C domains (approximately 100 b) by our model. Interestingly, the model also predicts that ssDNA folds at zero force even for very short lengths $N \sim 10-100$ b at physiological conditions (10 mM MgCl₂) [i.e., $\phi_C(0) \sim 0.5$ in Figs. 5(a) (top) and 6(d)]. This prediction is of biological relevance, because short ssDNA fragments (15–30 b) play important roles during DNA metabolic processes.

The agreement between the helix-coil model and MFold is remarkable. This agreement suggests that MFold includes cooperativity as an essential ingredient for secondary-structure prediction. Indeed, cooperativity is implicit in the energy of the stem-loop structures formed by stacking adjacent base pairs in the NN model and other secondary motifs. MFold does not account for base-pair mismatches and tertiary motifs, in general. This limitation of MFold excludes the possibility of predicting disordered structures made of a large number of small folded regions of low thermodynamic stability but entropically favorable because of their large number. On the other hand, our model does not include heterogeneity in the sequence while MFold does, predicting large secondary domains. These two features make the most important differences between the helix-coil model and MFold.

Thermal molecular folding (at zero force) cannot be directly measured with force spectroscopy. However, techniques such as hydrogen exchange combined with NMR and mass spectrometry might grant access to naturally folded domains and folding kinetics [79,80]. The kind of structures expected in ssDNA should differ from those reported in native RNAs and proteins [19,25]. Still, the cooperativity mechanisms might be similar and ultimately related to the fundamental nature of electrostatic interactions [81,82].

It might be interesting to apply the current model to investigate ssRNA folding, where base stacking interactions are stronger. Future work should also extend this study to incorporate sequence disorder while keeping the simplicity of a few energy parameters. This extension might be useful for models of folding in native RNAs and proteins showing cooperativity-dependent sequential folding.

ACKNOWLEDGMENTS

X. V.-G., M. M., and F. R. acknowledge support from European Unions Horizon 2020 Grant No. 687089, Spanish Research Council Grants No. FIS2016-80458-P and No. PID2019-111148 GB-I00. F. R. also acknowledges support from ICREA Academia Prizes 2013 and 2018. M. M. also acknowledges support from the Spanish Ramon y Cajal program of MICINN. C. R. P. and B. I. acknowledge support from Spanish Research Council (PGC2018-099341-B-I00) and Comunidad de Madrid (PEJD-2016/IND-2451).

APPENDIX A: OBTAINING SSDNA EXTENSION

1. Trap position shift determination

Molecular extension is obtained as described in Eq. (1), where x_0 is a shift of the trap position λ . x_0 results from imposing $\lambda = 0$, at zero force by extrapolating the force-distance curve (FDC) in the initial slope (i.e., before the

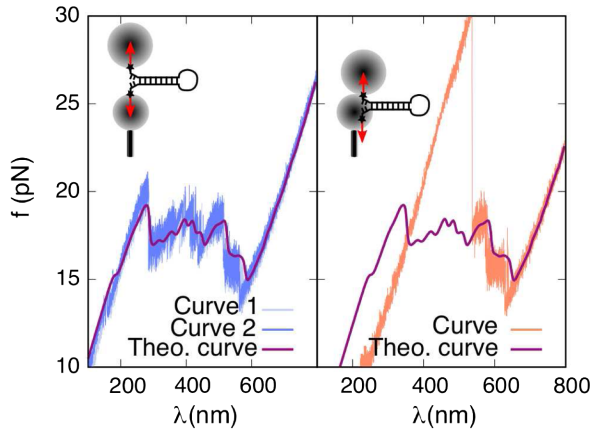


FIG. 7. Checking tether alignment. Experimental force-distance unzipping curves obtained for an aligned tether (left, blue) and a misaligned tether (right, orange). The agreement between the theoretical curve (purple) and the experimental one (blue) is a signature of correct tether alignment (left). Misaligned tethers usually lead to the lack of some of the initial characteristic unzipping rips (right). A schematic depiction of aligned and misaligned tethers are shown on top. For obtaining the theoretical curves, we use the NN energies for the base-pair energies [47], the elastic parameters from Ref. [32], and a bead stiffness of $k_b \sim 0.07$ pN/nm.

first rip) to zero force (left scheme in Fig. 7). For this procedure to work, tethers must be aligned along the pulling axis. To verify the tether alignment, we compare the pattern of the experimental unzipping curve with the theoretically predicted FDC in equilibrium [47], as shown in Fig 7. If both curves match, the alignment is correct.

2. Two-branches method

The two-branches method is used for obtaining the ssDNA FEC of the shortest molecule (H_{120}). As shown in Fig. 8, the molecule is unzipped and reziped with pulling velocities of approximately 100 nm/s. Two branches are defined depending on whether the hairpin is folded (F) or unfolded (U). The optical trap positions in the F (λ_F) and U (λ_U) branches read, respectively, as

$$\lambda_F(f) = x_h(f) + x_b(f) + x_d(f) + \lambda_{0,F}, \quad (\text{A1})$$

$$\lambda_U(f) = x_h(f) + x_b(f) + x_{\text{ssDNA}}(f) + \lambda_{0,U}, \quad (\text{A2})$$

where x_d is the extension of the folded hairpin and $\lambda_{0,F}$ and $\lambda_{0,U}$ are drift corrections.

$x_{\text{ssDNA}}(f)$ can be obtained along the force range where the two branches coexist (approximately 4–17 pN) from Eqs. (A1) and (A2):

$$x_{\text{ssDNA}}(f) = \lambda_F(f) - \lambda_U(f) + x_d(f) + \lambda_0(t), \quad (\text{A3})$$

where $\lambda_0(t)$ stands for the drift correction (see Sec. S11 in Supplemental Material [45] for details).

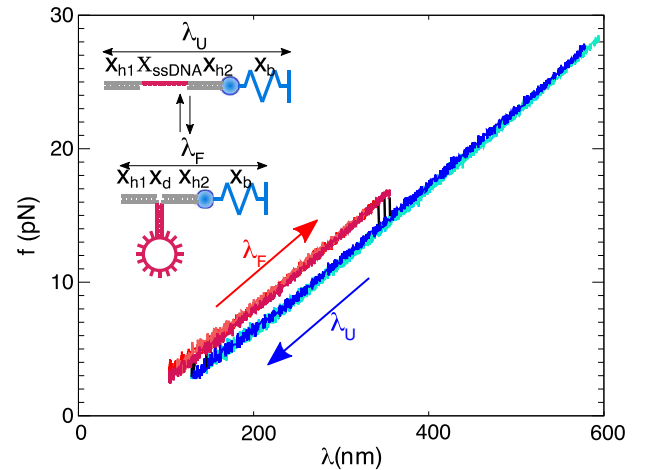


FIG. 8. Two-branches method. Force-distance curves for the H_{120} , with several stretching and releasing cycles. Experimental configurations corresponding to F and U branches are schematically shown in the top. The folded (λ_F) and unfolded (λ_U) branches are shown in red and blue, respectively, with unfolding and refolding events shown in black.

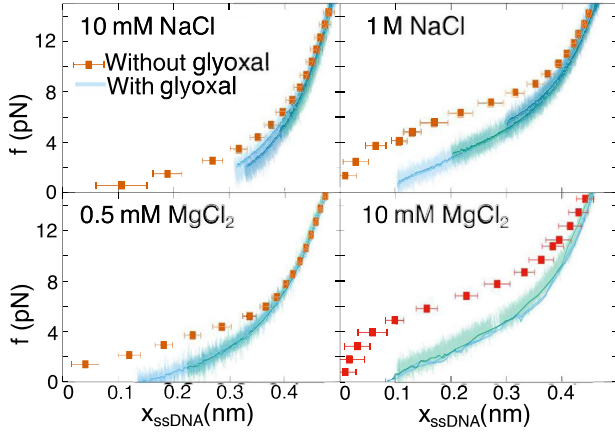


FIG. 9. ssDNA FECs with glyoxal. Rescaled force-extension curves for different salt concentrations in the presence and absence of glyoxal. The red points correspond to the rescaled averaged FEC for the H_{13680} ssDNA molecule. The blue and green lines show releasing cycles for different tethers of the H_{7138} ssDNA molecule after the secondary structure is suppressed by glyoxal. Experiments are performed in 10 mM NaCl (top left), 1M NaCl (top right), 0.5 mM $MgCl_2$ (bottom left), and 10 mM $MgCl_2$ (bottom right). The dark lines show the averaged signal using a 10 Hz filter.

APPENDIX B: EXPERIMENTS WITH GLYOXAL

Experiments with chemically treated ssDNA are performed in order to check the characterization of the free ssDNA elasticity. Following Refs. [28,54,55], we carry out experiments with glyoxal using the following protocol. We first unzip dsDNA in the absence of glyoxal to generate ssDNA. Since glyoxal prevents base pairing, the blocking loop oligonucleotide is not necessary to prevent hairpin re-zipping. Once unzipped, we keep the force constant above the unzipping force (approximately 20 pN at 1M NaCl and 10 mM $MgCl_2$, approximately 15 pN at 10 mM NaCl and 0.5 mM $MgCl_2$) and flowed glyoxal at 1M concentration.

We wait for 15 min for glyoxal to covalently attach to the exposed ssDNA bases. Glyoxal changes the index of refraction of the solvent, modifying force calibration. Therefore, once glyoxal coats ssDNA, we wash the chamber by flowing the original buffer to remove glyoxal and measure the glyoxal-ssDNA FECs. Figure 9 shows the results we obtain for different molecules at the lowest and highest salt conditions in sodium and magnesium. For comparison, we also show the results obtained without glyoxal. The differences between the FECs obtained in the two conditions (with and without glyoxal) are due to the presence or absence of secondary structure.

In Fig. 10, we show the FECs averaged over the molecules at each condition together with three elastic models: an extensible freely jointed chain model (with parameters from Ref. [32]), the WLC model in Eq. (2), and the Saleh formula in Eq. (3). The results show that at high

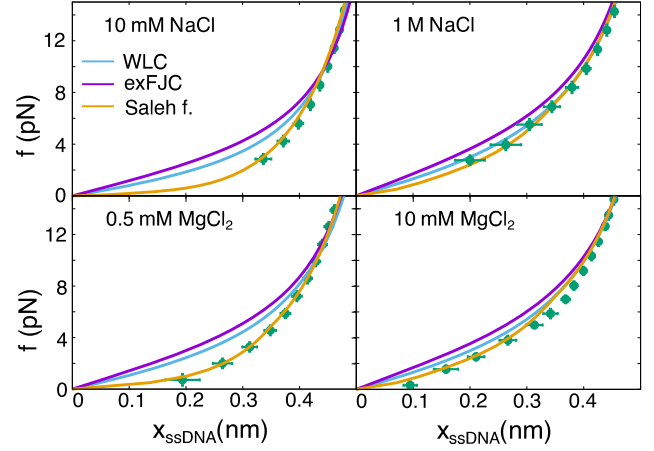


FIG. 10. Comparison between elastic models and the experimental FEC with glyoxal. Green points show the averaged rescaled FECs of the H_{7138} ssDNA molecule for different salt concentrations (from 2–3 different tethers). The continuous lines correspond to the elastic curve obtained from the WLC (blue, elastic parameters from Ref. [32]), extensible freely jointed chain (magenta, elastic parameters from Ref. [32]), and Saleh formula (orange, described in Sec. II E 2).

salts the inextensible-WLC and the Saleh formula are both compatible with the data. However, there are differences at low salts, where the Saleh formula works better. This difference confirms the results shown in Fig. S4 in Supplemental Material [45], where the largest differences among the elastic models appear at the lowest salt concentrations.

APPENDIX C: HAIRPIN ORIENTATION

A DNA hairpin has a helix diameter of $d = 2$ nm, corresponding to that of dsDNA (B form) [83]. Under an externally applied force, the hairpin extension can be modeled as a single bond of length d that orients along the stretching direction [50,71,84]:

$$x_d(f) = d \left[\coth \left(\frac{df}{k_B T} \right) - \frac{k_B T}{df} \right], \quad (C1)$$

where k_B is the Boltzmann constant and T is the temperature.

APPENDIX D: C-DOMAIN LENGTH

C domains are modeled as rigid dipoles of length δ that orient in the presence of force [Eq. (C1), $d = \delta$]. By assuming that terminal bases in C domains form a base pair [as in hairpinlike secondary structures, Fig. 6(a)], then $\delta \sim 2$ nm, the double-helix diameter.

We investigate the effect of varying δ in our model. Starting from $\delta = 0$ and upon increasing its value, we find that the fitting energy parameters ϵ and γ remain almost unchanged (Fig. 11): ϵ and γ increase $\lesssim 3\%$ when $0 \leq \delta \leq 0.6$ nm. We conclude that the effect of δ is very small in the obtained fitting parameters. For the sake of

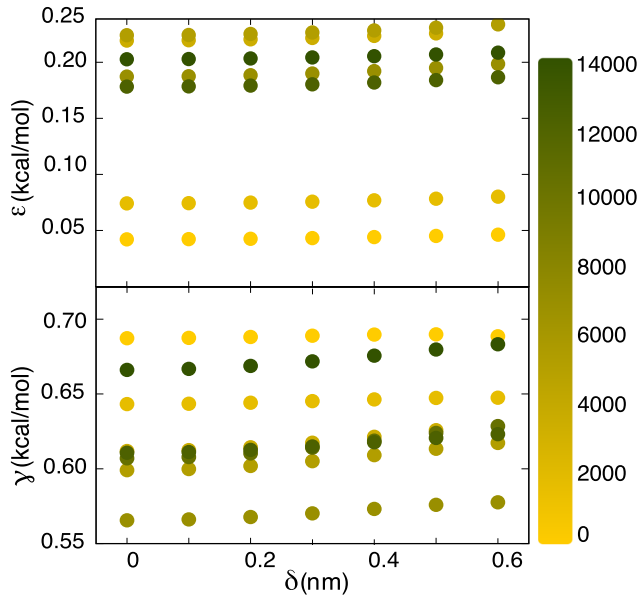


FIG. 11. Effect of the domain size in the fitting parameters. Best-fitting values for the model parameters ϵ and γ as a function of the length δ . The effect is very small: ϵ (approximately 3% variation, top) and γ (approximately 3% variation, bottom). The color code indicates the molecular length, from short (yellow) to long (dark green).

simplicity, we use $\delta = 0$ nm throughout the analysis presented in this work.

Notice that our model, as presented in Sec. II F, is applicable only for $\delta \leq l \sim 0.7$ nm; otherwise, for $\delta > l$ the model predicts, at sufficiently high forces, a structure made of alternating C and F domains of only one base ($\phi_C = 1/2$, $n_C = 1$). This prediction is clearly incorrect, as secondary structure is fully disrupted at sufficiently high forces. This inaccuracy of the helix-coil model is a consequence of its simplicity and might be corrected by imposing a minimum number of bases, n_{\min} , to form a C domain such that $n_C > n_{\min}$. With this restriction, the helix-coil model becomes too complicated and is not analytically solvable anymore. This fact and the negligible effect of δ on ϵ and γ (Fig. 11) explain why we take $n_{\min} = 1$ and $\delta = 0$, keeping the simplicity of our model.

APPENDIX E: GC CONTENT EFFECT ON SECONDARY STRUCTURE

Sequences with high GC content present the shoulder in the FEC at larger forces [28]. The molecules studied in this work (Sec. III B) present small differences in the GC content (44%–53%). For long molecules ($N \gtrsim 500$ b, squares), ϵ as a function of the GC content increases (Fig. 12), compatible with the larger stability of secondary structures with high GC content, as shown in Fig. 12. For shorter molecules ($N \lesssim 500$ b, yellow circles), ϵ decreases strongly, which we interpret as finite-size effects (see the main text).

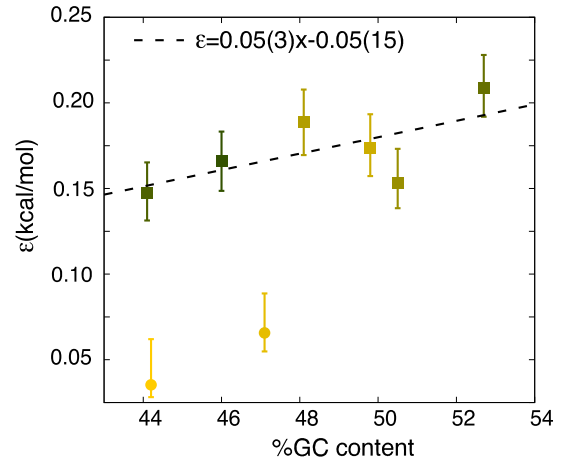


FIG. 12. Effect of the GC content in ϵ . Best-fitting values for ϵ as a function of the GC content of the sequence, for the eight molecules studied. Long molecules ($N \gtrsim 500$ b) are shown in squares, whereas short molecules ($N \lesssim 500$ b), deviating from the collapsed FEC trend observed in Fig. 1(d), are shown in circles. The linear fit (dashed line) performed with long molecules ($700\text{b} \leq N \leq 14$ kb) shows an increase of ϵ with GC content, as expected from the higher stability of GC bps as compared to AT bps. The color code is as in Fig. 11.

APPENDIX F: MODEL VALIDATION

To test the model, we use experimental FEC data of an 8022-base ssDNA molecule obtained by unpeeling a

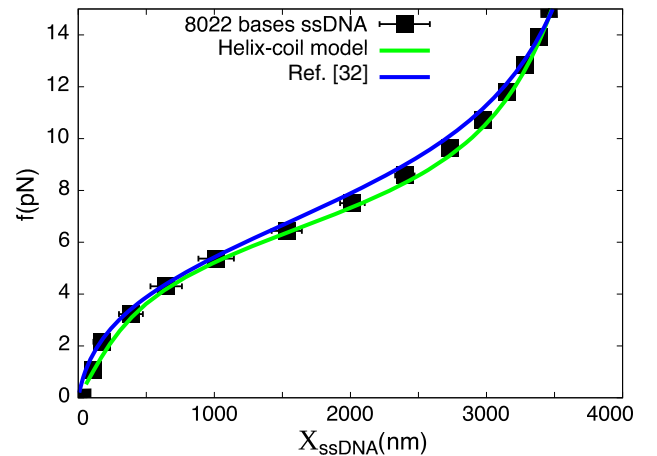


FIG. 13. Comparison of the model with a ssDNA FEC obtained using unpeeling of dsDNA. Average FEC of an 8022-base ssDNA (black points) in 20 mM Tris pH 7.5, 10 mM MgCl_2 , and 20 mM NaCl. The errors are the standard errors obtained from averaging three different tethers. The experimental extension of each ssDNA tether is corrected with a small multiplying factor (from 0.9 to 1.1) to match the WLC elasticity at 15 pN. The FEC predicted by the model (green line) is obtained using Eq. (13) with ϵ given by Eq. (14) (main text) and $\gamma = 0.61$ kcal/mol (average value) [Fig. 3(c)]. The blue curve is the phenomenological formula proposed in Ref. [32] (see Sec. S10 in Supplemental Material [45] for the used parameters).

dsDNA molecule using a different optical tweezers setup [27] (Fig. 13). In this method, one strand of the dsDNA molecule is attached by its 5' and 3' ends to the beads. Pulling the molecule above 80–100 pN (approximately 200 nm/s) induces the mechanical denaturalization of the double helix and promotes the releasing of the free strand (the one not attached to the beads) [46,48]. We use our model without any fitting parameters to predict the FEC of this molecule at the experimental conditions (10 mM MgCl₂ and 20 mM NaCl). The good match between the model prediction and the experimental data (Fig. 13) further validates the model.

APPENDIX G: HELIX-COIL MODEL DERIVATION

The model Hamiltonian, given by Eq. (5), can be written as

$$\mathcal{H} = -A \sum_i^N \sigma_i - B \sum_i^N \sigma_i \sigma_{i+1} - C(f) - D, \quad (\text{G1})$$

where

$$A = -\frac{\epsilon}{2} + \frac{1}{2} \int_0^f x_F(f') df', \quad (\text{G2})$$

$$B = -\frac{1}{4} \int_0^f d_C(f') df' + \gamma, \quad (\text{G3})$$

$$C = \frac{N}{2} \left(\int_0^f x_F(f') df' + \frac{1}{2} \int_0^f d_C(f') df' \right), \quad (\text{G4})$$

$$D = \frac{N}{2} \epsilon. \quad (\text{G5})$$

D is a constant and, therefore, is not considered in the calculations below. The partition function reads as

$$Z = \sum_{\{\sigma_i\}} e^{-\beta \mathcal{H}(\{\sigma_i\})} = \sum_{\sigma_1} V^N \sigma_1 \sigma_1 = \text{tr}(V^N), \quad (\text{G6})$$

where tr is the trace and V is the transfer matrix given by

$$V = e^{C/N} \begin{pmatrix} e^{\beta(A+B)} & e^{-\beta B} \\ e^{-\beta B} & e^{-\beta(A-B)} \end{pmatrix}. \quad (\text{G7})$$

By diagonalizing V , we find the eigenvalues

$$\lambda_{\pm} = e^{\beta C/N} e^{\beta B} \cdot \left[\cosh(\beta A) \pm \sqrt{e^{-4\beta B} + \sinh^2(\beta A)} \right]. \quad (\text{G8})$$

The partition function [Eq. (G6)] can then be written as

$$Z_N = \lambda_+^N \left[1 + \left(\frac{\lambda_-}{\lambda_+} \right)^N \right], \quad (\text{G9})$$

which, in the thermodynamic limit $N \rightarrow \infty$ ($\lambda_- < \lambda_+$), leads to

$$Z_N = \lambda_+^N. \quad (\text{G10})$$

1. Free-energy computation

From the partition function [Eq. (G10)], we can compute the free energy per monomer (base) as

$$\begin{aligned} \frac{G}{N} &= -\frac{1}{\beta} \log(\lambda_+) = -\frac{C}{N} - B \\ &\quad - \frac{1}{\beta} \log \left[\cosh(\beta A) + \sqrt{e^{-4\beta B} + \sinh^2(\beta A)} \right], \end{aligned} \quad (\text{G11})$$

with A , B , and C given in Eqs. (G2)–(G4). The free energy per monomer at zero force, G_0/N , is obtained by imposing $f = 0$ in Eq. (G11):

$$\frac{G_0}{N} = -\gamma - \frac{1}{\beta} \log \left[\cosh(-\beta \epsilon / 2) + \sqrt{e^{-4\beta \gamma} + \sinh^2(-\beta \epsilon / 2)} \right]. \quad (\text{G12})$$

For a given DNA sequence, MFold gives the free energy of formation of secondary structures with respect the random coil (i.e., configuration with all bases free). In order to compare the free energy of the helix-coil model with MFold predictions, we subtract from Eq. (G12) the energy per base of the random coil at zero force [Eq. (5) with $f = 0$, $\sigma_i = 1$, $\forall i$], $\epsilon/2 - \gamma$:

$$\frac{\Delta G_0}{N} = -\frac{\epsilon}{2} - \frac{1}{\beta} \log \left[\cosh\left(\frac{\beta \epsilon}{2}\right) + \sqrt{e^{-4\beta \gamma} + \sinh^2\left(\frac{\beta \epsilon}{2}\right)} \right]. \quad (\text{G13})$$

2. Computation of $\phi_F(f)$, $\phi_C(f)$, and $M_C(f)$

The fraction of bases in F domains can be written as

$$\phi_F(f) = \frac{N_F(f)}{N} = \frac{1 + \langle \sigma_n \rangle(f)}{2}, \quad (\text{G14})$$

where $\langle \rangle$ stands for the ensemble average.

Using the transfer matrix definition in Eq. (G7), the average σ for the first monomer reads as

$$\begin{aligned} \langle \sigma_1 \rangle(f) &= \frac{1}{Z} \sum_{\sigma_1 \dots \sigma_N} \sigma_1 V(\sigma_1, \sigma_2) V(\sigma_2, \sigma_3) \dots V(\sigma_N, \sigma_1) \\ &= \frac{1}{Z} \sum_{\sigma_1 \dots \sigma_N} V' V^{N-1}, \end{aligned} \quad (\text{G15})$$

where

$$V' = e^{C/N} \begin{pmatrix} e^{\beta(A+B)} & e^{-\beta B} \\ -e^{\beta B} & -e^{-\beta(A-B)} \end{pmatrix} \quad (\text{G16})$$

and A , B , and C are given in Eqs. (G2)–(G4). Considering Pauli matrices $V' = \sigma_z V$, we can write

$$\langle \sigma_1 \rangle(f) = \frac{1}{Z} \text{tr}(\sigma_z V^N). \quad (\text{G17})$$

Using the cyclic property of the trace, Eq. (G17) can be extended to any site n :

$$\langle \sigma_n \rangle(f) = \frac{1}{Z} \text{tr}(\sigma_z V^N). \quad (\text{G18})$$

By diagonalizing V , we can write

$$\langle \sigma_n \rangle(f) = \frac{1}{Z} (\langle v_{\pm} | \sigma_z | v_{\pm} \rangle \lambda_{\pm}^N + \langle v_{\pm} | \sigma_z | v_{\pm} \rangle \lambda_{\mp}^N), \quad (\text{G19})$$

where $|v_{\pm}\rangle$ are the eigenvectors:

$$\begin{aligned} |v_{+}\rangle &= \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix}, \\ |v_{-}\rangle &= \begin{pmatrix} -\sin(\theta/2) \\ \cos(\theta/2) \end{pmatrix}, \end{aligned} \quad (\text{G20})$$

with the angle θ given by

$$\tan \theta = \frac{e^{-2\beta B}}{\sinh(\beta A)}. \quad (\text{G21})$$

In the thermodynamic limit, $Z = \lambda_{+}^N$ and Eq. (G19) reduces to

$$\langle \sigma_n \rangle(f) = \langle v_{\pm} | \sigma_z | v_{\pm} \rangle. \quad (\text{G22})$$

By substituting Eq. (G20) in Eq. (G22), we obtain

$$\begin{aligned} \langle \sigma_n \rangle(f) &= \cos^2(\theta/2) - \sin^2(\theta/2) = \cos(\theta) \\ &= \frac{\sinh(\beta A)}{\sqrt{e^{-4\beta B} + \sinh^2(\beta A)}}. \end{aligned} \quad (\text{G23})$$

Finally, using Eq. (G23) in Eq. (G14) leads to

$$\phi_F(f) = \left(\frac{1}{2} + \frac{\sinh(\beta A)}{2\sqrt{e^{-4\beta B} + \sinh^2(\beta A)}} \right). \quad (\text{G24})$$

The fraction of bases in C domains $\phi_C(f) = [1 - N_F(f)]/N$ then reads as

$$\phi_C(f) = \left(\frac{1}{2} - \frac{\sinh(\beta A)}{2\sqrt{e^{-4\beta B} + \sinh^2(\beta A)}} \right). \quad (\text{G25})$$

On the other hand, the number of compact domains, M_C , is given by

$$\langle M_C \rangle = \frac{N}{4} (1 - \langle \sigma_i \sigma_{i+1} \rangle). \quad (\text{G26})$$

Using the transfer matrix formalism, we can write $\langle \sigma_n \sigma_{n+r} \rangle$ as

$$\langle \sigma_n \sigma_{n+r} \rangle(f) = \frac{\langle v_{+} | \sigma_z V^r \sigma_z | v_{+} \rangle}{\langle v_{+} | V^r | v_{+} \rangle} = \cos^2(\theta) + \left(\frac{\lambda_{-}}{\lambda_{+}} \right)^r \sin^2(\theta). \quad (\text{G27})$$

The number of compact domains [Eq. (G26)] then reads as

$$\begin{aligned} \langle M_C \rangle &= \frac{N}{4} \sin^2 \theta \left(1 - \frac{\lambda_{-}}{\lambda_{+}} \right) \\ &= \frac{N}{2} \frac{\sinh(\beta A)}{\cosh(\beta A) + \sqrt{e^{-4\beta B} + \sinh^2(\beta A)}}, \end{aligned} \quad (\text{G28})$$

where we use Eq. (G23).

-
- [1] H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira *et al.*, *Molecular Cell Biology* (MacMillan, London, 2008).
 - [2] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, P. W. K. Roberts *et al.*, *Molecular Biology of the Cell* (Garland Science, Taylor & Francis Group, London, 2018).
 - [3] U. Nagaswamy, N. Voss, Z. Zhang, and G. E. Fox, *Database of Non-canonical Base Pairs Found in Known RNA Structures*, *Nucl. Acids Res.* **28**, 375 (2000).
 - [4] J. Florián, J. Šponer, and A. Warshel, *Thermodynamic Parameters for Stacking and Hydrogen Bonding of Nucleic Acid Bases in Aqueous Solution: Ab Initio/Langevin Dipoles Study*, *J. Phys. Chem. B* **103**, 884 (1999).
 - [5] J. Šponer, J. Leszczynski, and P. Hobza, *Hydrogen Bonding, Stacking and Cation Binding of DNA Bases*, *J. Mol. Struct.* **573**, 43 (2001).
 - [6] D. W. Staple and S. E. Butcher, *Pseudoknots: RNA Structures with Diverse Functions*, *PLoS Biol.* **3**, e213 (2005).
 - [7] D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, and S. Balasubramanian, *The Regulation and Functions of DNA and RNA G-Quadruplexes*, *Nat. Rev. Mol. Cell Biol.* **21**, 459 (2020).
 - [8] Y. Wan, M. Kertesz, R. C. Spitale, E. Segal, and H. Y. Chang, *Understanding the Transcriptome through RNA Structure*, *Nat. Rev. Genet.* **12**, 641 (2011).
 - [9] K. J. Marians, *Prokaryotic DNA Replication*, *Annu. Rev. Biochem.* **61**, 673 (1992).
 - [10] S. P. Bell and A. Dutta, *DNA Replication in Eukaryotic Cells*, *Annu. Rev. Biochem.* **71**, 333 (2002).

- [11] A. M. van Oijen and J. J. Loparo, *Single-Molecule Studies of the Replisome*, *Annu. Rev. Biophys.* **39**, 429 (2010).
- [12] R. G. Roeder, *The Complexities of Eukaryotic Transcription Initiation: Regulation of Preinitiation Complex Assembly*, *Trends Biochem. Sci.* **16**, 402 (1991).
- [13] E. C. Friedberg, G. C. Walker, W. Siede, and R. D. Wood, *DNA Repair and Mutagenesis* (American Society for Microbiology, Washington, DC, 2005).
- [14] E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, and S. M. Block, *Direct Observation of Base-Pair Stepping by RNA Polymerase*, *Nature (London)* **438**, 460 (2005).
- [15] M. Stracy, S. Uphoff, F. Garza de Leon, and A. N. Kapanidis, *In Vivo Single-Molecule Imaging of Bacterial DNA Replication, Transcription, and Repair*, *FEBS Lett.* **588**, 3585 (2014).
- [16] D. Thirumalai and C. Hyeon, *RNA and Protein Folding: Common Themes and Variations*, *Biochemistry* **44**, 4957 (2005).
- [17] K. A. Dill and J. L. MacCallum, *The Protein-Folding Problem, 50 Years on*, *Science* **338**, 1042 (2012).
- [18] G. Portella and M. Orozco, *Multiple Routes to Characterize the Folding of a Small DNA Hairpin*, *Angew. Chem., Int. Ed. Engl.* **49**, 7673 (2010).
- [19] D. B. Ritchie and M. T. Woodside, *Probing the Structural Dynamics of Proteins and Nucleic Acids with Optical Tweezers*, *Curr. Opin. Struct. Biol.* **34**, 43 (2015).
- [20] F. Schneider, N. Möritz, and H. Dietz, *The Sequence of Events during Folding of a DNA Origami*, *Sci. Adv.* **5**, eaaw1412 (2019).
- [21] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, *Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis*, *Proteins* **21**, 167 (1995).
- [22] L. Wallace and C. Matthews, *Sequential vs. Parallel Protein-Folding Mechanisms: Experimental Tests for Complex Folding Reactions*, *Biophys. Chem.* **101–102**, 113 (2002).
- [23] E. M. Mahen, P. Y. Watson, J. W. Cottrell, and M. J. Fedor, *mRNA Secondary Structures Fold Sequentially but Exchange Rapidly In Vivo*, *PLoS Biol.* **8**, e1000307 (2010).
- [24] M. Karplus, *Behind the Folding Funnel Diagram*, *Nat. Chem. Biol.* **7**, 401 (2011).
- [25] S. W. Englander and L. Mayne, *The Nature of Protein Folding Pathways*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15873 (2014).
- [26] K. C. Neuman and A. Nagy, *Single-Molecule Force Spectroscopy: Optical Tweezers, Magnetic Tweezers and Atomic Force Microscopy*, *Nat. Methods* **5**, 491 (2008).
- [27] S. B. Smith, Y. Cui, and C. Bustamante, *Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules*, *Science* **271**, 795 (1996).
- [28] M.-N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette, *Stretching Single Stranded DNA, a Model Polyelectrolyte*, *Phys. Rev. Lett.* **89**, 248102 (2002).
- [29] A. Buhot and A. Halperin, *Effects of Stacking on the Configurations and Elasticity of Single-Stranded Nucleic Acids*, *Phys. Rev. E* **70**, 020902(R) (2004).
- [30] C. Bustamante, J. Marko, E. Siggia, and S. Smith, *Entropic Elasticity of λ -Phage DNA*, *Science* **265**, 1599 (1994).
- [31] A. Montanari and M. Mézard, *Hairpin Formation and Elongation of Biomolecules*, *Phys. Rev. Lett.* **86**, 2178 (2001).
- [32] A. Bosco, J. Camunas-Soler, and F. Ritort, *Elastic Properties and Secondary Structure Formation of Single-Stranded DNA at Monovalent and Divalent Salt Conditions*, *Nucl. Acids Res.* **42**, 2064 (2014).
- [33] B. H. Zimm and J. Bragg, *Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains*, *J. Chem. Phys.* **31**, 526 (1959).
- [34] S. Lifson and A. Roig, *On the Theory of Helix-Coil Transition in Polypeptides*, *J. Chem. Phys.* **34**, 1963 (1961).
- [35] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic, New York, 1970).
- [36] M. Fixman and J. J. Freire, *Theory of DNA Melting Curves*, *Biopolymers* **16**, 2693 (1977).
- [37] Y. Kafri, D. Mukamel, and L. Peliti, *Why Is the DNA Denaturation Transition First Order?*, *Phys. Rev. Lett.* **85**, 4988 (2000).
- [38] T. Garel and H. Orland, *Generalized Poland-Scheraga Model for DNA Hybridization*, *Biopolymers* **75**, 453 (2004).
- [39] A. Badasyan, A. Grigoryan, E. S. Mamasakhlisov, A. Benight, and V. Morozov, *The Helix-Coil Transition in Heterogeneous Double Stranded DNA: Microcanonical Method*, *J. Chem. Phys.* **123**, 194701 (2005).
- [40] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, *Glassy Transition in a Disordered Model for the RNA Secondary Structure*, *Phys. Rev. Lett.* **84**, 2026 (2000).
- [41] M. Müller, F. Krzakala, and M. Mézard, *The Secondary Structure of RNA under Tension*, *Eur. Phys. J. E* **9**, 67 (2002).
- [42] T. R. Einert, H. Orland, and R. R. Netz, *Secondary Structure Formation of Homopolymeric Single-Stranded Nucleic Acids Including Force and Loop Entropy: Implications for DNA Hybridization*, *Eur. Phys. J. E* **34**, 1 (2011).
- [43] S. B. Smith, Y. Cui, and C. Bustamante, *Optical-Trap Force Transducer That Operates by Direct Measurement of Light Momentum*, in *Methods in Enzymology, Biophotonics, Part B*, Vol. 361 (Academic, New York, 2003), pp. 134–162.
- [44] J. Gieseler, J. R. Gomez-Solano, A. Magazzù, I. P. Castillo, L. P. García, M. Gironella-Torrent, X. Viader-Godoy, F. Ritort, G. Pesce, A. V. Arzola, K. Volke-Sepúlveda, and G. Volpe, *Optical Tweezers—From Calibration to Applications: A Tutorial*, *Adv. Opt. Photonics* **13**, 74 (2021).
- [45] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevX.11.031037> for detailed description of the sequences, synthesis protocols, experimental procedures and comparison between the elastic models used in the work.
- [46] B. Ibarra, Y. R. Chemla, S. Plyasunov, S. B. Smith, J. M. Lzaro, M. Salas, and C. Bustamante, *Proofreading Dynamics of a Processive DNA Polymerase*, *EMBO J.* **28**, 2794 (2009).
- [47] J. M. Huguet, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort, *Single-Molecule Derivation of Salt Dependent Base-Pair Free Energies in DNA*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15431 (2010).
- [48] A. Candelli, T. P. Hoekstra, G. Farge, P. Gross, E. J. Peterman, and G. J. Wuite, *A Toolbox for Generating*

- Single-Stranded DNA in Optical Tweezers Experiments*, *Biopolymers* **99**, 611 (2013).
- [49] D. B. McIntosh, G. Duggan, Q. Gouil, and O. A. Saleh, *Sequence-Dependent Elasticity and Electrostatics of Single-Stranded DNA: Signatures of Base-Stacking*, *Biophys. J.* **106**, 659 (2014).
- [50] N. Forns, S. de Lorenzo, M. Manosas, K. Hayashi, J. M. Huguet, and F. Ritort, *Improving Signal/Noise Resolution in Single-Molecule Experiments Using Molecular Constructs with Short Handles*, *Biophys. J.* **100**, 1765 (2011).
- [51] X. Zhang, K. Halvorsen, C.-Z. Zhang, W. P. Wong, and T. A. Springer, *Mechanoenzymatic Cleavage of the Ultralarge Vascular Protein von Willebrand Factor*, *Science* **324**, 1330 (2009).
- [52] J. F. Marko and E. D. Siggia, *Stretching DNA*, *Macromolecules* **28**, 8759 (1995).
- [53] C. Bouchiat, M. Wang, J.-F. Allemand, T. Strick, S. Block, and V. Croquette, *Estimating the Persistence Length of a Worm-like Chain Molecule from Force-Extension Measurements*, *Biophys. J.* **76**, 409 (1999).
- [54] O. A. Saleh, D. B. McIntosh, P. Pincus, and N. Ribbeck, *Nonlinear Low-Force Elasticity of Single-Stranded DNA Molecules*, *Phys. Rev. Lett.* **102**, 068301 (2009).
- [55] D. B. McIntosh and O. A. Saleh, *Salt Species-Dependent Electrostatic Effects on ssDNA Elasticity*, *Macromolecules* **44**, 2328 (2011).
- [56] D. R. Jacobson, D. B. McIntosh, M. J. Stevens, M. Rubinstein, and O. A. Saleh, *Single-Stranded Nucleic Acid Elasticity Arises from Internal Electrostatic Tension*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5095 (2017).
- [57] C. V. Bizarro, A. Alemany, and F. Ritort, *Non-specific Binding of Na^+ and Mg_2^+ to RNA Determined by Force Spectroscopy Methods*, *Nucl. Acids Res.* **40**, 6922 (2012).
- [58] A. Severino, A. M. Monge, P. Rissone, and F. Ritort, *Efficient Methods for Determining Folding Free Energies in Single-Molecule Pulling Experiments*, *J. Stat. Mech.: Theory Exp.* **2019**, 124001 (2019).
- [59] ϕ_F equals the fraction of unpaired bases ϕ defined in Ref. [32].
- [60] W. Saenger, *Principles of Nucleic Acid Structure* (Springer Science, New York, 2013).
- [61] B. Maier, D. Bensimon, and V. Croquette, *Replication by a Single DNA Polymerase of a Stretched Single-Stranded DNA*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12002 (2000).
- [62] M. L. Bochman, K. Paeschke, and V. A. Zakian, *DNA Secondary Structures: Stability and Function of G-Quadruplex Structures*, *Nat. Rev. Genet.* **13**, 770 (2012).
- [63] A mild trend with varying GC content is observed; see Appendix E.
- [64] D. Marenduzzo and C. Micheletti, *Thermodynamics of DNA Packaging inside a Viral Capsid: The Role of DNA Intrinsic Thickness*, *J. Mol. Biol.* **330**, 485 (2003).
- [65] N. M. Toan, D. Marenduzzo, and C. Micheletti, *Inferring the Diameter of a Biopolymer from Its Stretching Response*, *Biophys. J.* **89**, 80 (2005).
- [66] N. Toan and C. Micheletti, *Inferring the Effective Thickness of Polyelectrolytes from Stretching Measurements at Various Ionic Strengths: Applications to DNA and RNA*, *J. Phys. Condens. Matter* **18**, S269 (2006).
- [67] At low forces, the measured extension is slightly larger than predicted by the WLC model, suggesting excluded volume effects.
- [68] J. SantaLucia, *A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1460 (1998).
- [69] J. Hooyberghs, P. Van Hummelen, and E. Carlon, *The Effects of Mismatches on Hybridization in DNA Microarrays: Determination of Nearest Neighbor Parameters*, *Nucl. Acids Res.* **37**, e53 (2009).
- [70] F. Landuzzi, X. Viader-Godoy, F. Cleri, I. Pastor, and F. Ritort, *Detection of Single DNA Mismatches by Force Spectroscopy in Short DNA Hairpins*, *J. Chem. Phys.* **152**, 074204 (2020).
- [71] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block, *Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid*, *Science* **314**, 1001 (2006).
- [72] N. Peyret, *Prediction of Nucleic Acid Hybridization: Parameters and Algorithms*, Ph.D. dissertation, Wayne State University, Department of Chemistry, Detroit, MI, 2000.
- [73] M. Zuker, *Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction*, *Nucl. Acids Res.* **31**, 3406 (2003).
- [74] J. M. Huguet, M. Ribezzi-Crivellari, C. V. Bizarro, and F. Ritort, *Derivation of Nearest-Neighbor DNA Parameters in Magnesium from Single Molecule Experiments*, *Nucl. Acids Res.* **45**, 12921 (2017).
- [75] The 7138 and 13680 bases sequences are not studied due to the huge computational time (which grows like $\sim N^3$).
- [76] Y. Seol, G. M. Skinner, K. Visscher, A. Buhot, and A. Halperin, *Stretching of Homopolymeric RNA Reveals Single-Stranded Helices and Base-Stacking*, *Phys. Rev. Lett.* **98**, 158103 (2007).
- [77] B. D. Sattin, W. Zhao, K. Travers, S. Chu, and D. Herschlag, *Direct Measurement of Tertiary Contact Cooperativity in RNA Folding*, *J. Am. Chem. Soc.* **130**, 6085 (2008).
- [78] B. Gracia, H. M. Al-Hashimi, N. Bisaria, R. Das, D. Herschlag, and R. Russell, *Hidden Structural Modules in a Cooperative RNA Folding Transition*, *Cell Rep.* **22**, 3240 (2018).
- [79] L. Konermann, J. Pan, and Y.-H. Liu, *Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics*, *Chem. Soc. Rev.* **40**, 1224 (2011).
- [80] E. Largy and V. Gabelica, *Native Hydrogen/Deuterium Exchange Mass Spectrometry of Structured DNA Oligonucleotides*, *Anal. Chem.* **92**, 4402 (2020).
- [81] P. C. Anthony, A. Y. Sim, V. B. Chu, S. Doniach, S. M. Block, and D. Herschlag, *Electrostatics of Nucleic Acid Folding under Conformational Constraint*, *J. Am. Chem. Soc.* **134**, 4607 (2012).
- [82] H. X. Zhou and X. Pang, *Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation*, *Chem. Rev.* **118**, 1691 (2018).
- [83] R. R. Sinden, *DNA Structure and Function* (Gulf Professional Publishing, New York, 1994).
- [84] A. Alemany and F. Ritort, *Determination of the Elastic Properties of Short ssDNA Molecules by Mechanically Folding and Unfolding DNA Hairpins*, *Biopolymers* **101**, 1193 (2014).