Dynamical modeling of molecular constructions and setups for DNA unzipping

Carlo Barbieri[1], **Simona Cocco**[1], **Rémi Monasson**[2]
and Francesco Zamponi[2]

[1] LPSENS, Unité Mixte de Recherche (UMR 8550) du CNRS et de l'ENS,
associée à l'UPMC Université Paris 06, 24 Rue Lhomond, 75231 Paris Cedex 05, France
[2] LPTENS, Unité Mixte de Recherche (UMR 8549) du CNRS et de l'ENS,
associée à l'UPMC Université Paris 06, 24 Rue Lhomond, 75231 Paris Cedex 05, France

Abstract
We present a dynamical model of DNA mechanical unzipping under the action of a force. The model includes the motion of a fork in a sequence-dependent landscape, the trap(s) acting on the bead(s) and the polymeric components of the molecular construction (unzipped single strands of DNA and linkers). Different setups are considered to test the model, and the outcome of the simulations is compared to simpler dynamical models existing in the literature where polymers are assumed to be at equilibrium.

1. Introduction

Over the past 15 years, various single molecule experiments have investigated DNA mechanical and structural properties [1–18] and protein–DNA interactions [19–29]. These experiments provide dynamical information usually hidden in large-scale bulk experiments, such as fluctuations on the scale of the individual molecule. The separation of the two strands of a DNA molecule under a mechanical stress, usually referred to as unzipping, was first carried out by Bockelmann and Heslot in 1997 [8]. The strands are pulled apart at a constant velocity while the force necessary for the opening is measured. The average opening force for the λ-phage sequence is about 15 pN (at room temperature and standard ionic conditions), with fluctuations around this value that depend on the particular sequence content. Bockelmann, Heslot and collaborators have shown that the force signal is correlated to the average sequence on the scale of ten base pairs but could be affected by the mutation of one base pair (bp) adequately located along the sequence [10]. Liphardt *et al* [15] and Danilowicz *et al* [16–18] have performed an analogous experiment, using a constant force setup, on a short RNA and long DNA molecules respectively (figure 1(B)). The distance between the two strand extremities is measured as a function of the time while the molecule is submitted to

a constant force. The separation of DNA strands has also been studied in single molecule experiments by translocation through nanopores [26, 27].

The motivation underlying unzipping experiments of DNA is (at least) twofold. First, the study of unzipping aims at a better understanding of the mechanisms governing the opening of DNA during transcription and replication by proteins such as polymerases, helicases and exonucleases [20, 21, 28, 29]. Simple theoretical models describing the opening as an unidimensional random walk on a sequence-dependent free energy landscape have been proved to describe quite well several experimental effects such as stick–slip motion in the opening at constant velocity [9, 10], the long pauses at a fixed position of unzipping at constant forces [16, 30, 31], the hopping dynamics between two or more states in unzipping at critical forces of short DNA molecules [15, 31–33] and the torsional drag effects in unzipping at large velocity [11, 34]. Moreover, statistical mechanical analyses have been successfully applied to extract from experimental data the sequence-dependent free energy landscape and the height of free energy barriers [35, 36].

Second, unzipping experiments could potentially be useful to extract information on the sequence itself [37]. Recently, single molecule sequencing has been achieved by monitoring a DNA/RNA polymerase in the course of
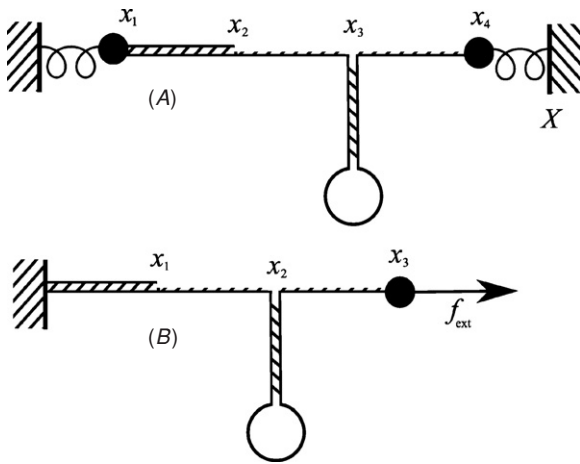
Figure 1. Typical experimental setups that will be described in the following. (*A*) A setup with two optical traps (beads x_1 and x_4) drawn as springs and whose centers are the black vertical lines and (*B*) a setup with a single magnetic bead x_3 that applies a constant force to the molecule attached to a fixed 'wall'. In both cases, the molecular construction is made by a DNA molecule that has to be opened (therefore, one should include two single-strand linkers that are the opened parts of the molecule) and one double-stranded DNA linker. The coordinates x_i are the distances of the corresponding points from the left reference position (which is the center of the left optical trap in case (*A*) and the fixed wall in case (*B*)).

DNA synthesis from a ssDNA template [33, 38]; such single molecule sequencing could become competitive with standard DNA sequencing because they do not require, *a priori*, amplification through polymerase chain reactions. A fundamental question on the possibility of extracting information on the sequence from unzipping experiments is the influence of the experimental setup on the measures and the limitations imposed by the latter [37, 39]. Indeed, characteristic spatio-temporal limitations are the finite rates of data acquisition, the relaxation time of the bead, the limited spatial resolution, the thermal drift and more generally the noise in the instruments. Moreover, the dynamics of the opening fork (figure 1) is influenced by the single strands (open parts) of the molecule and the linkers, and cannot be deduced directly from the observation of the bead from which the force or the position is measured.

The accuracy of unzipping experiments at fixed velocity has improved a lot over the last decade. Initially performed with an optical fiber [8], experiments were then based on the use of simple optical traps [10]. Nowadays, double optical traps [13, 36] allow us to considerably reduce the drift of the setup and to achieve a temporal resolution of the order of 10 kHz, a sub-nanometric spatial resolution, and a precision on measured forces of the order of fraction of pN. Unzipping at fixed force has been performed by a magnetic trap with a low temporal resolution (from 60 Hz to 200 Hz) due to the time needed to extract the position of the bead, the spatial precision being of the order of 10 nm Hz$^{-1/2}$ [28, 29], or by an optical trap also with a low temporal resolution (about 10 Hz) imposed by a feedback mechanism needed to keep the force constant [15]. Recently, a new dumbbell dual optical trap has been developed. It operates without feedback and can

maintain the force constant over distances of about 50 nm [33] with a temporal resolution of 10 kHz and a spatial resolution of 0.1 nm Hz$^{-1/2}$.

Limitations due to the experimental systems were first addressed in [39]. This paper stated the impossibility of inferring the sequence due to ssDNA fluctuations: fluctuations increase with the number of opened base pairs and can become larger than the length of about 1 nm corresponding to the spatial resolution of one open base pair. This problem could however be solved by integrating out the single-strand dynamical fluctuations. Several works have studied the effects of the setup on the hopping dynamics of small RNA molecules [32, 33, 39, 40]. The following effects have been underlined. First, the free energy landscape changes when adding a harmonic potential to the free energy, due to the bead and handles [10, 32, 33, 40]. Therefore, for a given force, the measured separation of the extremities depends on the stiffnesses of the trap and handles. Moreover, the opening and closing rates depend on the stiffness of the optical trap; in particular when the experimental system gets softer the fluctuations of the force gets smaller, and the hopping rates approach their fixed-force values.

In this paper, we introduce a model for the coupled dynamics of the opening fork, the ssDNA strand, the linkers and the bead in the optical or magnetic trap. Essential notions and existing literature are reviewed in section 2. Our dynamical model is presented in section 3. Our program allows us to simulate a generic setup, characterized by bead dimensions, optical stiffness (absent in the case of magnetical tweezers), linker composition (dsDNA or ssDNA) and lengths, and the length of molecule to be unzipped. All the parameters that characterize the different dynamical components can be adjusted in the simulation. The model is then used to simulate fixed-force (section 4) and fixed-extension (section 5) numerical unzippings.

2. Free energies, time scales and effective dynamics

We discuss hereafter the thermodynamic properties of the various parts of the experimental setup (DNA sequence, open part of the molecule, single- or double-strand linkers), as well as the relevant time scales. Finally, we briefly review previous dynamical studies where the linkers and the open portion of the molecules are assumed to be at equilibrium.

2.1. Thermodynamics of the components

2.1.1. Polymeric models for the linkers and open molecule. A polymer model is specified by its free energy as a function of the extension x for a given number n of monomers; we call this quantity $W(x, n)$. When x and n are large, W is an extensive quantity; hence, $W(x, n) = nw(x/n) = nw(l)$, where $l = x/n$ is the extension per monomer. We also define

$$f(l) = \frac{\partial W(x, n)}{\partial x} = w'(l),$$
$$l(f) = \text{inverse of } f(l), \quad (1)$$
$$g(f) = \max_l [fl - w(l)] = fl(f) - w[l(f)],$$

which are, respectively, the force at fixed extension, the average extension at fixed force and the free energy at fixed force. Note that $g(f)$ is simply the integral of $l(f)$. Hence, a polymer model is completely described from the knowledge of the extension versus force characteristic curve, $l(f)$. In the following, we will use some classical models for this function.

- *Gaussian (Hook) model.*

$$l_{\text{Hook}}(f) = \frac{f}{k^m}, \qquad (2)$$

where the stiffness constant k^m is related to the temperature T and the average squared monomer length (at zero force) b^2 through $k^m = k_B T / b^2$.

- *Freely-jointed chain (FJC) model.*

$$l_{\text{FJC}}(f) = \coth\left(\frac{fb}{k_B T}\right) - \frac{k_B T}{fb} \qquad (3)$$

is the extension (per monomer) of a chain of rigid rods of length b, free to rotate around each other. Comparison of this model with force–extension curves for single-stranded DNA shows that a better fit is obtained from a modified FJC:

$$l_{\text{MFJC}}(f) = d\left(1 + \frac{f}{\gamma_{\text{ss}}}\right) \times l_{\text{FJC}}(f), \qquad (4)$$

which takes into account the elasticity effects on the rod length. Standard fit parameters are $d = 0.56$ nm, $b = 1.4$ nm and $\gamma_{\text{ss}} = 800$ pN.

- *Extensible worm-like chain (WLC) model.*

$$l_{\text{WLC}}(f) = L\left[1 - \frac{1}{2}\left(\frac{k_B T}{fA}\right)^{1/2} + \frac{f}{\gamma_{\text{ds}}}\right] \qquad (5)$$

is the formula for the high-force extension of an elastic chain with persistence length equal to A. Experiments show that it is an excellent description of double-stranded DNA at high forces, with $L = 0.34$ nm, $A = 48$ nm and $\gamma_{\text{ds}} = 1000$ pN.

2.1.2. Free-energy landscape for the sequence. Let $b_i = A, T, C$ or G denote the ith base along the $5' \rightarrow 3'$ strand (the other strand is complementary) and $B = \{b_1, b_2, \ldots, b_N\}$. The free-energy excess when the first n bp of the molecule is open with respect to the closed configuration ($n = 0$) is [31]

$$G(n; B) = \sum_{i=1}^{n} g_0(b_i, b_{i+1}), \qquad (6)$$

where $g_0(b_i, b_{i+1})$ is the binding energy of the bp number i; it depends on b_i (pairing interactions) and on the neighboring bp b_{i+1} due to stacking interactions. g_0 is obtained from the MFOLD server [41, 42], and listed in table 1 for 150 mM NaCl, room temperature and pH 7.5. The values of the free energies should be changed for different ionic conditions and temperatures.

As an illustration, we plot the free energy $G(n; \Lambda)$ of the first 50 bases of the λ-phage sequence, $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$, in figure 2 after subtraction of $n g_{\text{ss}}(f)$ for forces $f = 15.9$ and 16.4 pN. $g_{\text{ss}}(f)$ is the work to stretch the two opened single strands when one more bp is opened, and calculated
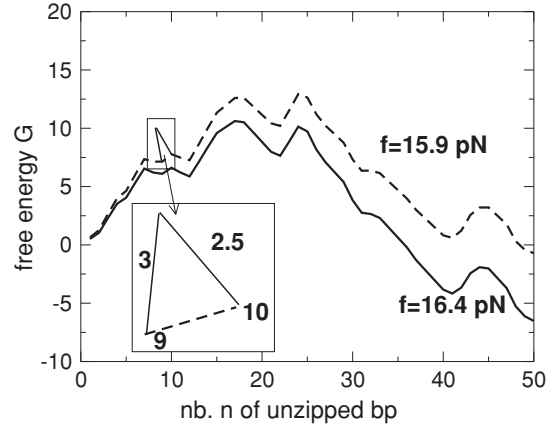

Figure 2. Free energy G (units of $k_B T$) to open the first n base pairs, for the first 50 bases of the DNA λ-phage at forces 15.9 (dashed curve) and 16.4 pN (full curve). For $f = 15.9$ pN, the two minima at bp 1 and bp 50 are separated by a barrier of 12 $k_B T$. Inset: additional barrier representing the dynamical rates (21) to go from base 10 to 9 (barrier equal to $2g_{\text{ss}} = 2.5\ k_B T$) and from base 9 to 10 (barrier equal to $g_0(b_9, b_{10}) = 3\ k_B T$); see text.

Table 1. Binding free energies $g_0(b_i, b_{i+1})$ (units of $k_B T$) obtained from the MFOLD server [41, 42] for DNA at room temperature, pH = 7.5 and an ionic concentration of 0.15 M. The base values b_i and b_{i+1} are given by the line and column, respectively.

g_0	A	T	C	G
A	1.78	1.55	2.52	2.22
T	1.06	1.78	2.28	2.54
C	2.54	2.22	3.14	3.85
G	2.28	2.52	3.90	3.14

from the modified FJC model (4). The subtraction allows us to compare the increase in the free energy due to the opening of the sequence to the gain resulting from the release of ssDNA polymers at a given force.

At these forces, the two global minima in figure 2 are located in $n = 1$ (closed state) and $n = 50$ (partially open state). Experiments on a small RNA molecule, called P5ab [15], have been performed at the critical force f_c such that the closed state has the same free energy as the open one: $G(N; \Lambda) = N g_{\text{ss}}(f_c)$. They showed that, as the barrier between these two minima is not too high, the molecule switches between these two states; see section 2.3.

2.2. Fluctuations at equilibrium

2.2.1. Case of a single polymer. We now consider the orders of magnitude of the fluctuations of the polymer. When submitted to a force of $f = 15$ pN, the average extension of the polymer is $\bar{x} = n x^m$ with $x^m = l(f)$. We use for single-stranded DNA the MFJC model, and for double-stranded DNA the WLC model, with the parameters discussed in section 2.1.1; then we get $x^m_{\text{ss}} = 0.46$ nm and $x^m_{\text{ds}} = 0.33$ nm for ss- and dsDNA respectively. At thermal equilibrium, the extension will fluctuate around these average values. The fluctuations are controlled by the *microscopic effective spring constant* $k^m(l) = w''(l) = 1/l'(f)$. For ds-

Table 2. Fluctuations of single-stranded DNA at $f = 15$ pN and $T = 16.7\,^\circ$C; $\delta\bar{x}/\bar{x} = 0.37/\sqrt{n}$, $\delta\bar{f}/\bar{f} = 1.57/\sqrt{n}$, $\tau = 4.83 \times 10^{-11}\text{s}\,n^2$.

n	$\delta\bar{x}/\bar{x}$	$\delta\bar{f}/\bar{f}$	τ (s)
10	0.117	0.496	4.8×10^{-9}
40	0.058	0.248	7.7×10^{-8}
100	0.037	0.157	4.8×10^{-7}
400	0.018	0.078	7.7×10^{-6}
1000	0.012	0.050	4.8×10^{-5}

Table 3. Fluctuations of double-stranded DNA at $f = 15$ pN and $T = 16.7\,^\circ$C; $\delta\bar{x}/\bar{x} = 0.17/\sqrt{n}$, $\delta\bar{f}/\bar{f} = 4.83/\sqrt{n}$, $\tau = 5.1 \times 10^{-12}\text{s}\,n^2$.

n	$\delta\bar{x}/\bar{x}$	$\delta\bar{f}/\bar{f}$	τ (s)
100	0.017	0.483	5.1×10^{-8}
400	0.0085	0.241	8.1×10^{-7}
1000	0.0054	0.153	5.1×10^{-6}
4000	0.0027	0.076	8.1×10^{-5}
10000	0.0017	0.048	5.1×10^{-4}

and ssDNA we find, respectively, $k_{\text{ds}}^m = 1311$ pN nm^{-1} and $k_{\text{ss}}^m = 138$ pN nm^{-1} according to the above models. For a polymer with n monomers, the stiffness is $k = k^m/n$ since the effective spring constant is given by $k(x, n) = \frac{\partial^2}{\partial x^2} W(x, n) = k^m(x/n)/n$.

Alternatively, the force f exerted on the polymer will fluctuate around its average value $\bar{f}$ if its extremities are kept at a fixed distance x from each other. These fluctuations of force (in the fixed-extension ensemble) and extension (in the fixed-force ensemble) are easily computed by a quadratic expansion of the free energy around the average, i.e. when approximating the polymer with a spring of stiffness k^m/n, with the result

$$\langle \delta x^2 \rangle = \frac{k_B T}{k^m} n, \qquad \langle \delta f^2 \rangle = \frac{k_B T k^m}{n}. \qquad (7)$$

Defining $\delta\bar{x} = \sqrt{\langle\delta x^2\rangle}$ and $\delta\bar{f} = \sqrt{\langle\delta f^2\rangle}$, we get

$$\frac{\delta\bar{x}}{\bar{x}} = \sqrt{\frac{k_B T}{k^m(x^m)^2}}\frac{1}{\sqrt{n}}, \qquad \frac{\delta\bar{f}}{\bar{f}} = \sqrt{\frac{k_B T k^m}{\bar{f}^2}}\frac{1}{\sqrt{n}}. \qquad (8)$$

As expected, the relative fluctuations of both force and extension become smaller and smaller as the number n of monomers increases. Some values are reported in tables 2 and 3.

2.2.2. Case of several polymers (fixed-distance setup). Now consider the case of several polymers, e.g. linker and open part of the molecule attached one after the other. In a fixed-force experiment, the components of the setup are independent (at the level of the saddle-point approximation) and the fluctuations in the extensions simply add up. In the fixed-distance setup, however, correlations between the extensions make the analysis more complicated. As a concrete example, we consider the setup in figure 1(*A*). The linker joining x_1 and x_2 is a double-stranded DNA segment of N_{ds} bases. The two linkers joining (x_2, x_3) and (x_3, x_4) are single-stranded DNA segments of $N_{\text{ss}} = N_{\text{ss}}^0 + n$ bases, where n is the number of opened base pairs.

The centers of the two optical traps are at 0 and X. We call x_1 the position of the first bead and x_4 the position of the second. The probability $P_{\text{eq}}(n, x_1, x_2, x_3, x_4) = \text{e}^{-F/k_B T}$, where the free energy F reads as

$$F(\vec{x}, n) = \tfrac{1}{2}k_1 x_1^2 + W_{\text{ds}}(x_2 - x_1, N_{\text{ds}}) + W_{\text{ss}}(x_3 - x_2, N_{\text{ss}})$$
$$+ W_{\text{ss}}(x_4 - x_3, N_{\text{ss}}) + \tfrac{1}{2}k_2(x_4 - X)^2 + G(n; B), \qquad (9)$$

where $W_{\text{ds}}(x, N_{\text{ds}}) = N_{\text{ds}} w_{\text{ds}}(x/N_{\text{ds}})$ and $W_{\text{ss}}(x, N_{\text{ss}}) = N_{\text{ss}} w_{\text{ss}}(x/N_{\text{ss}})$ are the elongation free energies of the double strand and single strand, respectively.

In order to study the fluctuations in this setup, we first find the maximum of P_{eq} assuming that $G(n; B) = n g_0$, i.e. a uniform sequence B, and treating n as a continuous variable assuming that it is large. At the maximum $x_i = \bar{x}_i$ and we define

$$x_{\text{ds}}^m = \frac{\bar{x}_2 - \bar{x}_1}{N_{\text{ds}}}, \qquad x_{\text{ss}}^m = \frac{\bar{x}_3 - \bar{x}_2}{N_{\text{ss}}} = \frac{\bar{x}_4 - \bar{x}_3}{N_{\text{ss}}}. \qquad (10)$$

The saddle-point condition $\partial_{x_i} F_A = 0$ gives the following equations, which represent the force balance condition along the chain:

$$k_1\bar{x}_1 = w_{\text{ds}}'(x_{\text{ds}}^m) = w_{\text{ss}}'(x_{\text{ss}}^m) = k_2(X - \bar{x}_4) \equiv \bar{f}. \qquad (11)$$

The derivative with respect to n gives, using equations (1) and (11), the condition

$$g_0 = 2[x_{\text{ss}}^m w_{\text{ss}}'(x_{\text{ss}}^m) - w_{\text{ss}}(x_{\text{ss}}^m)] = g_{\text{ss}}(\bar{f}), \qquad (12)$$

which allows us to find the force $\bar{f}$ transmitted along the chain. Once (12) is solved, the extensions of the beads and of the double- and single-stranded parts of DNA ($\bar{x}_1$, $X - \bar{x}_4$, x_{ds}^m and x_{ss}^m respectively) are determined by equation (11). Finally, the number of open bases $\bar{n}$ is determined by

$$\bar{x}_1 + N_{\text{ds}} x_{\text{ds}}^m + 2(N_{\text{ss}}^0 + \bar{n})x_{\text{ss}}^m + (X - \bar{x}_4) = X. \qquad (13)$$

Note that the value of $\bar{f}$ is determined only by g_0.

We work at temperature $T = 16.7\,^\circ$C ($k_B T = 4$ pN nm) and choose a uniform molecule with $g_0 = 2.69 k_B T$, which is a representative value for the pairing free energies in table 1. We use the same models as in section 2.2.1 for the single- and double-stranded DNA, with $N_{\text{ds}} = 3120$ and $N_{\text{ss}}^0 = 40$. Then solving equation (12) we get $\bar{f} = 16.5$ pN, and from equation (11) we get $x_{\text{ss}}^m = 0.47$ nm, $x_{\text{ds}}^m = 0.33$ nm. We choose $k_1 = 0.1$ pN nm^{-1}, then $\bar{x}_1 = 165$ nm, and $k_2 = 0.512$ pN nm^{-1}, then $X - \bar{x}_4 = 32$ nm. Given these values, $\bar{n}$ is defined by X using equation (13):

$$\bar{n} = \frac{X - 1264}{0.94}, \qquad (14)$$

with X expressed in nanometers.

For the same setup, we can compute the fluctuations of n and of the elongations of the elements of the setup. In particular, the fluctuations of the bead positions are measurable in the experiment.

Let us define $\delta x_i = x_i - \bar{x}_i$ and $\delta n = n - \bar{n}$. To simplify the formalism, we also define $\delta x_{\text{ds}} = \delta x_2 - \delta x_1$, $\delta x_{\text{ss}}^L = \delta x_3 - \delta x_2$ and $\delta x_{\text{ss}}^R = \delta x_4 - \delta x_3$. A quadratic expansion of F around its minimum gives

$$\delta F \sim \frac{1}{2}k_1\delta x_1^2 + \frac{1}{2}k_2\delta x_4^2 + \frac{w_{\text{ds}}''(x_{\text{ds}}^m)}{2N_{\text{ds}}}\delta x_{\text{ds}}^2$$
$$+ \frac{w_{\text{ss}}''(x_{\text{ss}}^m)}{2N_{\text{ss}}^0 + \bar{n}}\Big[(\delta x_{\text{ss}}^L - x_{\text{ss}}^m\delta n)^2 + (\delta x_{\text{ss}}^R - x_{\text{ss}}^m\delta n)^2\Big]. \qquad (15)$$

Using (4) and (5), we get $k_{ss}^m = w_{ss}''(x_{ss}^m) = 152$ pN nm^{-1} and $k_{ds}^m = w_{ds}''(x_{ds}^m) = 1416$ pN nm^{-1}.

One should take care of the fact that $\delta x_1 + \delta x_4 + \delta x_{ds} + \delta x_{ss}^L + \delta x_{ss}^R = 0$; it is convenient to express δx_{ss}^R as a function of the others since its fluctuations are identical to those of δx_{ss}^L. The quadratic expansion of the function δF has the form $\delta F = \frac{1}{2}\delta\mathbf{x}A\delta\mathbf{x}$ where $\delta\mathbf{x} = \left(\delta x_1, \delta x_4, \delta x_{ds}, \delta x_{ss}^L, x_{ss}^m\delta n\right)$ and

$$A = \frac{k_{ss}^m}{N_{ss}^0 + \bar{n}}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 0 \\ 1 & 1 & 1 & 0 & 2 \end{pmatrix}$$

$$+ \begin{pmatrix} k_1 & 0 & 0 & 0 & 0 \\ 0 & k_2 & 0 & 0 & 0 \\ 0 & 0 & k_{ds}^m/N_{ds} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (16)$$

The inverse of the matrix A is

$$A^{-1} = \begin{pmatrix} \frac{1}{k_1} & 0 & 0 & -\frac{1}{2k_1} & -\frac{1}{2k_1} \\ 0 & \frac{1}{k_2} & 0 & -\frac{1}{2k_2} & -\frac{1}{2k_2} \\ 0 & 0 & \frac{N_{ds}}{k_{ds}^m} & -\frac{N_{ds}}{2k_{ds}^m} & -\frac{N_{ds}}{2k_{ds}^m} \\ -\frac{1}{2k_1} & -\frac{1}{2k_2} & -\frac{N_{ds}}{2k_{ds}^m} & \frac{1}{4k_{eff}} & \frac{1}{4k_{eff}^s} \\ -\frac{1}{2k_1} & -\frac{1}{2k_2} & -\frac{N_{ds}}{2k_{ds}^m} & \frac{1}{4k_{eff}^s} & \frac{1}{4k_{eff}} \end{pmatrix}, \quad (17)$$

where

$$\frac{1}{k_{eff}^s} = \frac{1}{k_1} + \frac{1}{k_2} + \frac{N_{ds}}{k_{ds}^m}, \qquad \frac{1}{k_{eff}} = \frac{1}{k_{eff}^s} + 2\frac{N_{ss}^0 + \bar{n}}{k_{ss}^m}. \quad (18)$$

This immediately gives

$$k_B T (A^{-1})_{1,1} = \langle \delta x_1^2 \rangle = \frac{k_B T}{k_1}$$

$$k_B T (A^{-1})_{2,2} = \langle \delta x_4^2 \rangle = \frac{k_B T}{k_2}$$

$$k_B T (A^{-1})_{3,3} = \langle \delta x_{ds}^2 \rangle = \frac{k_B T N_{ds}}{k_{ds}^m} \quad (19)$$

$$k_B T (A^{-1})_{4,4} = \langle (\delta x_{ss}^L)^2 \rangle = \frac{k_B T}{4k_{eff}}$$

$$\frac{k_B T}{(x_{ss}^m)^2}(A^{-1})_{5,5} = \langle \delta n^2 \rangle = \frac{k_B T}{4k_{eff}(x_{ss}^m)^2}$$

and shows that the fluctuations of n are dominated by the weakest element of the setup; moreover, the correlation between the bead displacements δx_1, δx_4 and the fluctuations of the number of open base pairs δn is $\langle \delta n \delta x_1 \rangle = -\frac{k_B T}{2k_1 x_{ss}^m}$ and $\langle \delta n \delta x_4 \rangle = -\frac{k_B T}{2k_2 x_{ss}^m}$; the stiffer the optical trap, the weaker is the correlation between the location of the bead and the number of open bases. Examples are given in table 4.

2.3. Effective dynamical models

In the simplest dynamical models, the fork (separating the open and closed portions of the molecule) undergoes a biased random motion in the sequence landscape. The linkers are treated at equilibrium, which is correct if their characteristic time scales are much smaller than the average time needed to open or close a base pair.

Table 4. Saddle-point calculation for the setup in figure 1(A) with a uniform molecule and $k_1 = 0.1$ pN nm^{-1}, $k_2 = 0.512$ pN nm^{-1}, $N_{ds} = 3120$, $N_{ds}^0 = 40$. The force along the molecule is $\bar{f} = 16.5$; then $k_{ss}^m = 152$ pN nm^{-1}, $k_{ds}^m = 1416$ pN nm^{-1} and $k_{eff}^s = 0.07$ pN nm^{-1}.

X	$\bar{n}$	k_{eff}	$\sqrt{\langle \delta n^2 \rangle}$
1273	10^1	0.067	8.2
1358	10^2	0.062	8.5
2204	10^3	0.036	11.2
10664	10^4	0.0068	25.7

2.3.1. Time scales for the polymeric components of the setup. In this section, we recall the typical time scales of the polymeric components in the setup. Assume that the polymers are subject to a Brownian force $\eta(t)$ which is a zero-average Gaussian process with an autocorrelation function $\langle \eta(t)\eta(0) \rangle = 2\Gamma T \delta(t)$. Let Γ be the friction coefficient of the polymer [43], that is, the ratio of the viscous force exerted by the solvent to the velocity. As will be shown in section 3, the friction coefficient scales as $\Gamma = \gamma^m n/3$ with $\gamma_{ss}^m = \gamma_{ds}^m \sim 2 \times 10^{-8}$ pN s nm^{-2}. Then, approximating $f(x, n) \sim k^m x/n$, the relaxation time for an isolated polymer of n bases is given by

$$\tau = \frac{\gamma^m n^2}{3k^m}. \quad (20)$$

Note that the factor 3 in the denominator of the above equation is an approximation for the true factor $\pi^2/4$. The validity of its approximation and the simplification it leads to will be discussed in appendix A.

It is useful to compare the amplitude of the force fluctuations with the noise. To do this, we approximate $\langle \delta f(t)\delta f(0) \rangle \sim 2\tau \langle \delta f^2 \rangle \delta(t) = 2T\Gamma_f \delta(t)$. Then, using equation (7) to estimate $\langle \delta f^2 \rangle$, we get $\Gamma_f = n\gamma^m/3 = \Gamma$, and (not surprisingly) the force fluctuations are of the same order as the noise term.

From table 2, the relaxation time of the unzipped strands is smaller than the typical base-pair opening (or closing) time as long as the number n of unzipped bases is smaller than a few hundreds. This is the case, in particular, for unzipping experiments on short RNA molecules.

2.3.2. Random walk in the sequence landscape. Let us first model the motion of the fork alone, that is, assuming that the other components of the setup are at equilibrium. We consider a DNA molecule unzipped under a fixed force f in the sequence-landscape $G(n; B) - ng_{ss}(f)$ of figure 2. The fork, whose position is denoted by $n(t)$, can move forward ($n \to n + 1$) or backward ($n \to n - 1$) with rates (probability per unit of time) equal to, respectively,

$$r_o(b_{n+1}, b_{n+2}) = r \exp[-\beta g_0(b_{n+1}, b_{n+2})],$$
$$r_c = r \exp[-2\beta g_{ss}(f)], \quad (21)$$

where $\beta = 1/k_B T$; see figure 2. The value of the attempt frequency r is of the order of 10^6 Hz [12, 14, 31]. Expression (21) for the rates is derived from the following assumptions. First, the rates should satisfy detailed balance. Second, we impose that the opening rate r_o depends on the binding free
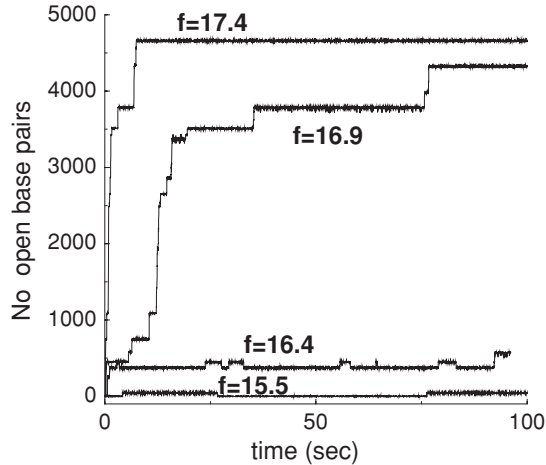
Figure 3. Number of open base pairs as a function of the time for various forces (shown in the figure). Data show one numerical unzipping (for each force) obtained from a Monte Carlo simulation of the random walk motion of the fork with rates (21).

energy, and not on the force, and vice versa for the closing rate r_c. This choice is motivated by the fact that the range for the base-pair interaction is very small: the hydrogen and stacking bonds are broken when the bases are kept apart at a fraction of an Angstrom, while the force work is appreciable on the distance of the opened bases (≈ 1 nm). In contrast, to close the base pairs, one has to first work against the applied force; therefore, the closing rate r_c depends on the force but not on the sequence. This physical origin of the rates is reported in the inset of figure 2. Note that, as room temperature is much smaller than the thermal denaturation temperature, we safely discard the existence of a denatured bubble in the zipped DNA portion.

An example of unzipping dynamics for the λ-phage sequence is shown in figure 3. The characteristic pauses in the unzipping, present in experiments and corresponding to deep local minima in the sequence landscape, are reproduced. The rates (21) lead to a master equation for the probability $\rho_n(t)$ for the fork to be at site n at time t:

$$\frac{\mathrm{d}\rho_n(t)}{\mathrm{d}t} = -\sum_{m=0}^{N} T_{n,m}\rho_m(t), \qquad (22)$$

where the matrix $T_{n,m}$ is tridiagonal with nonzero entries $T_{m-1,m} = -r_c(f)$, $T_{m+1,m} = -r_o(m)$ and $T_{m,m} = r_o(m) + r_c(f)$. Given this transition matrix, the opening dynamics can be simulated with Monte Carlo dynamics. For small RNA or DNA molecules, the transition matrix $T_{n,m}$ can be diagonalized numerically [31]. The smallest non-zero eigenvalue gives the switching time between a closed and open configuration for a hairpin with a free energy barrier such as that plotted in figure 2.

2.3.3. Dynamics of the bead with equilibrated linkers and strands. In a typical experiment, the force is exerted on the molecule through the action of a (magnetic or optical) trap on the bead. While the external force on the bead can be considered as constant (e.g. in a magnetic trap), the force

acting on the fork fluctuates unless the trap (and the molecular construction) is very soft; see equation (8). Therefore, the fixed-force model of the previous section has to be modified. In addition the bead, of size $R \simeq 1$ μm, is a slow component whose dynamics need to be taken into account. Let us denote by k the stiffness of the trap and by γ the friction of the bead in the solvent of viscosity η. Typical values for these quantities are $k = 0.1$–0.5 pN nm^{-1} and $\gamma = 6\pi R\eta = 1.67 10^{-5}$ pN s nm^{-1}. Thus, the characteristic relaxation time of the bead is $\tau = \gamma/k \simeq 0.2$–$1$ ms.

The coupled dynamics of the fork and the bead was considered by Manosas *et al* [14] in the case of small RNA unzipping, with a single optical trap. For such small molecules the relaxation time of the unzipped strands is expected to be much smaller than the characteristic time of the bead, and the molecule can be considered at equilibrium. The dynamical scheme therefore consists in a coupled evolution equation for the location of the bead and of the fork. The bead position obeys a Langevin equation including the external force and the force exerted by the fork through the (equilibrated) linkers and unzipped strands, while the fork moves with rates (21) with a bead location-dependent force.

A main conclusion of [14] is that, in the absence of feedback imposing a fixed force on the molecule, the trap stiffness must be as low as possible to detect jumps between closed and open configurations of the RNA molecule. We will discuss the validity of this statement in an information-theoretic setting in section 5.2.

3. Dynamical modeling of the setup and its components

The assumption that the linkers and the unzipped strands are at equilibrium as the unzipping proceeds is correct for short molecules as was the case in [14]. For long DNA molecules, the relaxation time of the unzipped strands may become large and dynamical modeling of the polymers involved in the molecular construction cannot be avoided.

The purpose of this section is to describe how such a dynamical model can be implemented. We hereafter denote by 'setup' the full molecular construction that is used in a given experiment, including linkers, beads, etc, while the word 'molecule' refers to the part of DNA which has to be opened. In an idealized description, the state variable is a vector $\vec{x} = (x_1, \ldots, x_p)$ whose elements are the distances from a reference position (that can be either the center of an optical trap or a fixed 'wall' to which the polymers are attached) of the extremities of the polymeric components in the setup. In addition to $\vec{x}$, the number of open base pairs n is needed to complete the description of the state of the setup.

As discussed in section 2.1, the total free energy $F(\vec{x}, n)$ of a setup is the sum of different contributions coming from all the elements of the setup. A typical example is given in equation (9).

Our aim is thus to construct a dynamical model that holds on intermediate time scales, $t \gtrsim 10^{-6}$ s, and

(i) gives the correct equilibrium Gibbs measure $P_{\mathrm{eq}}(\vec{x}, n) = \exp(-F(\vec{x}, n)/(k_{\mathrm{B}}T))$,

(ii) reproduces the relaxation times for the different elements of the setup, as discussed below,

(iii) gives reasonable dynamical correlations between different elements of the setup.

It is worth stressing at this point that ours is a coarse-grained model which does not take into account the motion of the individual monomers. It is expected that the dynamics on time scales smaller than the typical sojourn time of the fork on a base ($\gtrsim 10^{-6}$ s) is not relevant to our study of unzipping.

3.1. Langevin dynamics for the polymers and the beads

First, we consider the dynamics of $\vec{x}$ at fixed n. In appendix A, we show that for long enough times the dynamics of the setup can be described by a system of coupled Langevin equations:

$$\Gamma_{ij}\dot{x}_j = -\frac{\partial F}{\partial x_i} + \eta_i, \qquad (23)$$

where $i, j = 1, \ldots, p$, and

- the free energy $F(\vec{x})$ is the sum of a contribution coming from each element of the setup:
  - (i) each optical trap contributes $\frac{1}{2}k\Delta x^2$, where Δx is its elongation;
  - (ii) a bead in position i subjected to a constant force gives a contribution $-fx_i$;
  - (iii) a polymer gives a contribution $W_i(\Delta x, N_i)$, with Δx being its elongation and N_i its number of monomers.

For example, the total free energies of the setups in figure 1 are

$$
\begin{aligned}
F_A(\vec{x}) &= \tfrac{1}{2}k_1 x_1^2 + W_{\mathrm{ds}}(x_2 - x_1, N_{\mathrm{ds}}) + W_{\mathrm{ss}}(x_3 - x_2, N_{\mathrm{ss}}) \\
&\quad + W_{\mathrm{ss}}(x_4 - x_3, N_{\mathrm{ss}}) + \tfrac{1}{2}k_2(x_4 - X)^2, \\
F_B(\vec{x}) &= W_{\mathrm{ds}}(x_1, N_{\mathrm{ds}}) + W_{\mathrm{ss}}(x_2 - x_1, N_{\mathrm{ss}}) \\
&\quad + W_{\mathrm{ss}}(x_3 - x_2, N_{\mathrm{ss}}) - fx_3.
\end{aligned} \qquad (24)
$$

- $\vec{\eta}$ is a Gaussian white noise with zero average and variance $\langle \eta_i(t)\eta_j(0)\rangle = 2k_B T \Gamma_{ij}\delta(t)$, as requested by the fluctuation–dissipation relation.
- the matrix Γ is a tridiagonal matrix such that
  - (i) the diagonal element Γ_{ii} is the sum of three contributions:
    - (a) a term $\gamma_{i-1}^m N_{i-1}/3 + \gamma_i^m N_i/3$ coming from the adjacent polymers (if any);
    - (b) a term γ coming from the bead (if any) attached to x_i;
    - (c) a term taking into account the viscosity of the N_c base pairs of the DNA molecule attached to the fork (x_3 and x_2 in figures 1(A) and (B) respectively) that are not open; this term has the Fleury form $\gamma_{\mathrm{mol}} = \gamma' N_c^{3/5}$ and has to be added to the diagonal element of Γ corresponding to the fork position;
  - (ii) the offdiagonal elements are zero, except $\Gamma_{i,i+1} = \Gamma_{i+1,i} = \gamma_{i+1}^m \frac{N_{i+1}}{6}$ that get a contribution from the polymer joining x_i and x_{i+1}.

For instance, the setups in figure 1 correspond to the matrices:

$$
\Gamma_B = \begin{pmatrix} \gamma_{\mathrm{ds}}^m \frac{N_{\mathrm{ds}}}{3} + \gamma_{\mathrm{ss}}^m \frac{N_{\mathrm{ss}}}{3} & \gamma_{\mathrm{ss}}^m \frac{N_{\mathrm{ss}}}{6} & 0 \\ \gamma_{\mathrm{ss}}^m \frac{N_{\mathrm{ss}}}{6} & 2\gamma_{\mathrm{ss}}^m \frac{N_{\mathrm{ss}}}{3} + \gamma' N_c^{3/5} & \gamma_{\mathrm{ss}}^m \frac{N_{\mathrm{ss}}}{6} \\ 0 & \gamma_{\mathrm{ss}}^m \frac{N_{\mathrm{ss}}}{6} & \gamma + \gamma_{\mathrm{ss}}^m \frac{N_{\mathrm{ss}}}{3} \end{pmatrix},
$$

$$
\Gamma_A = \begin{pmatrix} \gamma + \gamma_{\mathrm{ds}}^m \frac{N_{\mathrm{ds}}}{3} & \gamma_{\mathrm{ds}}^m \frac{N_{\mathrm{ds}}}{6} & 0 & 0 \\ \gamma_{\mathrm{ds}}^m \frac{N_{\mathrm{ds}}}{6} & & & \\ 0 & & \Gamma_B & \\ 0 & & & \end{pmatrix}. \qquad (25)
$$

A detailed derivation of these results and in particular of the form of the matrix Γ can be found in appendix A.

3.2. Fork dynamics

The Langevin equation for the polymer dynamics at fixed n must be complemented with transition rates for the dynamics of n. To this aim, we discretize the Langevin equation with time step Δt, and at each time step we allow the opening $n \to n+1$ or closing $n \to n-1$ of a base pair at most.

The dynamics takes the form of a discrete time Markov chain, with transitions $(\vec{x}, n) \to (\vec{x}', n')$ and $n' \in \{n, n \pm 1\}$. The total free energy $F(\vec{x}, n) = F_{\mathrm{setup}}(\vec{x}, n) + G(n; B)$, where the first contribution has been discussed in the previous section and $G(n; B)$ is the pairing free energy of the molecule, as discussed in section 2.1.2. In appendix B, we show that in order to satisfy the detailed balance condition with respect to $P_{\mathrm{eq}}(\vec{x}, n) = \exp(-F(\vec{x}, n)/(k_B T))$, one should perform a single step following the procedure.

(i) Choose whether to stay ($n' = n$), to open ($n' = n+1$) or to close ($n' = n-1$) a base, with rates $r^{s,o,c}(\vec{x}, n)$ respectively:

$$
\begin{aligned}
r^o(\vec{x}, n) &= r\Delta t\, e^{\beta[G(n;B)-G(n+1;B)]}, \\
r^c(\vec{x}, n) &= r\Delta t\, e^{\beta F(\vec{x},n)-\beta F(\vec{x},n-1)}, \\
r^s(\vec{x}, n) &= 1 - r^o(\vec{x}, n) - r^c(\vec{x}, n).
\end{aligned} \qquad (26)
$$

(ii) If the choice was to open, *first* perform a discrete Langevin step $\vec{x} \to \vec{x}'$ at fixed n and *then* increase n by one.

(iii) If the choice was to close, *first* decrease n by one and *then* perform a discrete Langevin step $\vec{x} \to \vec{x}'$ at fixed $n' = n-1$.

(iv) If the choice was to stay, just perform a discrete Langevin step $\vec{x} \to \vec{x}'$ at fixed n.

The Langevin equation is discretized in a standard way by integrating equation (23) over a time Δt:

$$
x_i(t + \Delta t) = x_i(\Delta t) + \Gamma_{ij}^{-1}\left[-\frac{\partial F(\vec{x})}{\partial x_j}\Delta t + E_j \right], \qquad (27)
$$

where $E_j = \int_0^{\Delta t} \eta_j(t)\,\mathrm{d}t$ are Gaussian variables with zero average and variance

$$
\langle E_i E_j \rangle = 2k_B T \Gamma_{ij} \Delta t \qquad (28)
$$

that are independently drawn at each discrete time step.

3.3. Free energy at finite n

In section 2.1, we discussed some models for the free energy $W(x, n)$ of a polymer with n monomers and extension x. In the limit $x, n \to \infty$ at fixed extension per monomer, $l = x/n$, the free energy enjoys an extensivity property: $W(x, n) = nw(l)$. However, in our simulations we might be interested in regimes where n is small, typically of the order of 10–40 for small RNA molecules. In this case, knowledge of the free energy per monomer, w, is not sufficient, and a more detailed expression for W is necessary to avoid inconsistencies.

As a starting point of the analysis, we consider a polymer made of N identical monomers whose endpoints are denoted by $u_i, i = 1, \ldots, N$ with $u_0 = 0$. The Hamiltonian of the chain is the sum of pairwise interactions $\varphi(u_i - u_{i-1})$ and the free energy reads, for $x = u_N$, as

$$\mathrm{e}^{-\beta W(x,n)} = \ell_0^{-N+1} \int \mathrm{d}u_1, \ldots, \mathrm{d}u_{N-1}\, \mathrm{e}^{-\beta \sum_i \varphi(u_i - u_{i-1})}, \quad (29)$$

where ℓ_0 is a reference microscopic length scale. From the above relation, the Chapman–Kolmogorov equation follows:

$$\mathrm{e}^{-\beta W(x,n+m)} = \ell_0^{-1} \int \mathrm{d}y\, \mathrm{e}^{-\beta W(y,n) - \beta W(x-y,m)}. \quad (30)$$

We first consider for simplicity the Gaussian model, $\varphi(x) = \frac{1}{2}k^m x^2$. Then it is easy to show that

$$W(x, n) = \frac{k^m}{2n}x^2 - \frac{k_B T}{2} \log\left[\frac{k\ell_o^2}{2\pi k_B T n}\right]. \quad (31)$$

In the limit of large polymers, one obtains the free energy of a monomer of extension l through

$$w(l) = \lim_{n \to \infty} \frac{1}{n} W(x = ln, n) = \varphi(l) \quad (32)$$

as expected and consistent with the discussion of section 2.1. The logarithmic term in (31) contributes neither to w nor to the Langevin equation for x. However it does contribute to the rate to close a base pair (see equation (26)) and should be taken into account in order to recover the correct rates. An example of the effect of this term is obtained by computing the equilibrium probability of n. Consider the (unrealistic) case of a homopolymer, $G(n; B) = ng_0$, subject to a constant force and using a Gaussian model for the open part of the molecule; then

$$P_{\mathrm{eq}}(n) = \frac{1}{Z} \int \mathrm{d}x\, \mathrm{e}^{-n\beta g_0 - \beta W(x,2n) + \beta f x}$$
$$= \frac{1}{Z'} \mathrm{e}^{-n\beta g_0 + \frac{n}{k} f^2}. \quad (33)$$

Therefore $P_{\mathrm{eq}}(n)$ is a pure exponential, while if the correction were neglected one would have obtained wrong behavior at small n.

For a generic model of $\varphi(x)$, one cannot compute $W(x, n)$. Still we found that for our purposes ($n \gtrsim 40$), a consistent approximation is obtained by keeping only the first correction to the $n \to \infty$ result, i.e. by defining

$$\mathrm{e}^{-\beta W(x,n)} = \mathrm{e}^{-\beta n w(x/n)} \sqrt{\frac{\beta k(x/n)\ell_o^2}{2\pi n}}, \quad (34)$$

where $k(l) = w''(l)$. One can check that this expression satisfies equation (30) with corrections in the exponent of $O(1)$, while the terms $O(\log n + \log m)$ are taken into account. Within this approximation, the error in $\log r_c(x, n)$ in equation (26) is $O(1/n^2)$ while if the first corrections are neglected it is $O(1/n)$.

In the following, we will make use of definition (34) unless otherwise stated. We will discuss an example where the effects of neglecting the corrections are clearly observable.

3.4. Details of the numerical simulations

We performed numerical simulations of the molecular constructions depicted in figure 1, with the following specifications.

- The total free energies of the two setups are given by equation (24) plus the term $G(n; B)$.

- The free energy of each polymer includes the saddle-point corrections, i.e. it is given by equation (34). The relation $l(f)$ (see section 2.1) is numerically inverted to obtain $w(l)$ and $k(l)$ that enter in equation (34).

- For the single-stranded DNA we used the MFJC model, equation (4), with $d = 0.56$ nm, $b = 1.4$ nm and $\gamma_{\mathrm{ss}} = 800$ pN.

- For the double-stranded DNA we used the WLC model in equation (5), with a small regularization term to avoid a divergence for $f \to 0$, which is however irrelevant for values of forces to be discussed in the following, and with $A = 48$ nm, $L = 0.34$ nm and $\gamma_{\mathrm{ds}} = 1000$ pN.

- Unless otherwise stated, the double-stranded DNA linker is made of $N_{\mathrm{ds}} = 3120$ bps, while the two single-stranded linkers are made of $N_{\mathrm{ss}} = 40 + n$ bases each, where n is the number of open DNA bases (in other words, we included on each side a 40-base single-stranded linker).

- We worked at fixed temperature $k_B T = 4$ pN nm, corresponding to $T = 16.7\,^\circ$C.

- We used the dynamical equations for the polymers defined above, equations (23), within the discrete procedure illustrated in section 3.2 and with transition rates (26) for the fork with the attempt rate $r = 10^6$ Hz.

- The matrices Γ corresponding to the setups in figure 1 are given in equation (25); we used $\gamma_{\mathrm{ds}}^m = \gamma_{\mathrm{ss}}^m = \gamma' = 2 \times 10^{-8}$ pN s nm^{-1}. We used a value $\gamma = 1.67 \times 10^{-5}$ pN s nm^{-1} for the viscosity of the beads.

- The time step was fixed at $\Delta t = 10^{-8}$ s; this value ensures a correct integration of the equation of motion in all the regimes discussed below. Even if in some cases a larger integration step could be used, we decided to keep it fixed in order to be sure that discretization biases are not present.

The values of the spring constants k_1 and k_2 and of the force f in equation (24) varied in different simulation runs, and will be specified later.

The program we used for the numerical simulations can be downloaded from http://www.lpt.ens.fr/ zamponi. A user-friendly version will be made available as soon as possible.

3.5. Limits of validity of the dynamical model

Our model of the polymer dynamics suffers from two main limitations.

First, we keep only one collective coordinate for each polymer (its extension) associated with the longest relaxation mode. Faster modes are discarded. The approximation is justified provided there is no other mode slower than the typical sojourn time on a base pair. From the discussion of section 2.3.1, the number of unzipped base pairs, n, cannot be well above a thousand.

Another upper limit on n comes from the assumption that the force is uniform along the polymer. In principle the force is a function of the time t and the location y along the polymer, which obeys a diffusion equation with a microscopic diffusion coefficient $D_{ss}^m \simeq (x_{ss}^m)^2 / \tau_{ss}^m$, where x_{ss}^m is the length of a monomer and $\tau_{ss}^m = \gamma_{ss}^m / k_{ss}^m$ is its relaxation time. Assume that, at time 0, a base pair closes and the polymer is stretched at the extremity $x = 0$ by x_{ss}^m. Then the force, initially equal to $f(x, t = 0) = k_{ss}^m x_{ss}^m \delta(x)$, will decay following the Gaussian diffusion kernel. At time t, the force density at the extremity is $f(x, t) = k_{ss}^m x_{ss}^m / \sqrt{2\pi D_{ss}^m t}$. The relaxation is over when this force excess is of the same order of magnitude as the typical thermal fluctuations δf calculated in (8), that is, for times

$$t > n \frac{k_{ss}^m (x_{ss}^m)^2}{2\pi k_B T} \tau_{ss}^m \simeq 2 \times 10^{-10} n \text{ ps}. \qquad (35)$$

When $n \sim 1000$, the corresponding relaxation time is of the order of the sojourn time on a base.

In conclusion, our dynamical model is adapted to ssDNA polymers whose length ranges from a few hundred to a few thousand bases. Shorter polymers can be considered at equilibrium, while longer polymers cannot be modeled without taking into account the space dependence of forces. A simple way to tackle this difficulty consists in arbitrarily cutting long polymers into 1000-base long segments, each modeled as above. This procedure will be followed in section 5.1.

4. Unzipping at fixed force

4.1. Quasi-equilibrium unzipping

Before turning to the more interesting case of out-of-equilibrium unzipping, we focus on the case of a small molecule which is subject to a constant force close to the critical force. In this situation, the molecule is able to visit all the possible configurations.

We performed a set of numerical simulations at constant force $\bar{f} = 16.45$ pN, with the setup described in figure 1(B). The DNA molecule is a uniform segment of $N = 500$ base pairs, with pairing free energy $G(n; B) = n g_0$ and $g_0 = 2.69 k_B T$. The entropic free energy per base of the two open single strands is $2 g_{ss}(\bar{f}) = 2.684 k_B T$. Therefore, the infinite molecule would stay close; we are slightly below the critical force. To the right and left open portions of the molecule, two single-stranded DNA linkers of $N_{ss}^0 = 40$ bases each are attached; therefore, the total length of the single-stranded linkers is $N_{ss} = N_{ss}^0 + n$, where n is as usual the
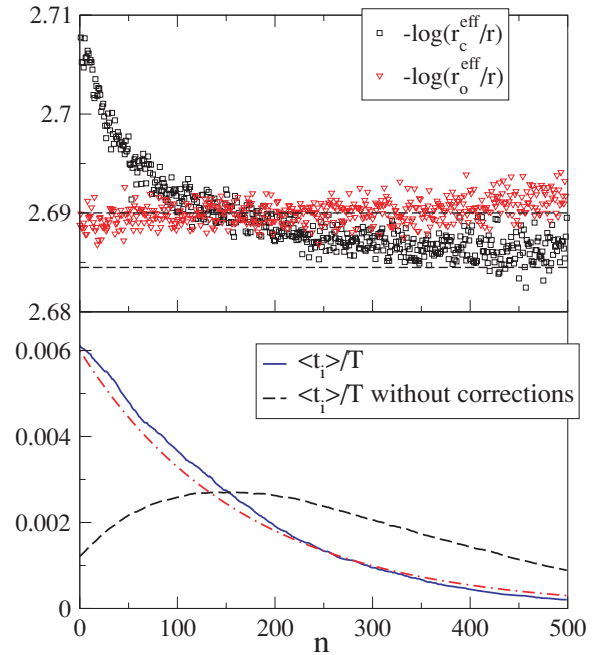

Figure 4. Bottom: average fraction of the time spent on each base. The full (blue) curve corresponds to equation (34) while the dashed (black) curve corresponds to equation (34) without the saddle-point corrections (the square-root term). The dot-dashed (red) line is $P_{eq}(n) \propto \exp[-n \Delta g]$ with $\Delta g = 0.006$. Top: effective rates (squares and triangles) estimated from the maximization of the probability in equation (36) ($r = 10^6$ Hz) without saddle-point corrections (full curve of the lower panel). The dashed lines are the asymptotic values of the rates; see text. We do not report the rates corresponding to the full equation (34) since they are essentially independent of n.

number of open base pairs. The leftmost linker is a double-stranded DNA of $N_{ds} = 3120$ base pairs, whose presence is however irrelevant for the scope of this section. The total length of the simulation was $T = 7200$ s, i.e. 2 h.

4.1.1. A test of the model. The average fraction of time spent on each base, corresponding to the equilibrium probability distribution $P_{eq}(n)$, is reported in the lower panel of figure 4. We expect that in the large n limit, $P_{eq}(n) \sim \exp[-n(g_0 - 2g_{ss}(f))] = \exp[-n \Delta g]$, with $\Delta g \sim 0.006$. This is expected to break down when N_{ss} is so small that the second-order corrections to the saddle-point in equation (34) become important. As can be seen in figure 4, the exponential form correctly describes the data.

We performed additional simulations in which the square-root term in equation (34) was removed. As one can see, in this case the small n deviations are much more pronounced. It is worth noting that for a non-Gaussian polymer, one expects a deviation from the exponential form at small enough n. However, this analysis shows that taking into account the small n corrections to $W(x, n)$ systematically reduces this effect. Estimating its real order of magnitude therefore requires an exact expression for $W(x, n)$, which could be in principle obtained from the recurrence equation (30). However, this is a complicated numerical task that goes beyond the scope of this

paper. What we want to stress here is that the inclusion of the square-root term in equation (34) gives significant differences when $n \lesssim 200$ and should therefore be included if one wants to analyze the unzipping of small molecules.

4.1.2. Effective dynamics of the fork. In a situation where the linkers are short, such that their relaxation time is faster than the mean time spent on a base, the linkers are able to reach equilibrium before n changes. Therefore one might hope to define an *effective dynamics* for the fork, where n changes according to effective rates that depend on the variation in the free energy of the setup on closing or opening a base.

To this aim we considered the model for the fork dynamics described in section 2.3, but assuming n-dependent opening and closing rates. Within this model, the probability of a trajectory of the fork is a function of the number of upward (u_n)/downward (d_n) jumps and the time spent on base n, t_n:

$$P_{\text{eff}}[n(t)] = \prod_{n=1}^{N} \left(r_c^{\text{eff}}(n)\Delta t\right)^{d_n} \left(r_o^{\text{eff}}(n)\Delta t\right)^{u_n}$$
$$\times \left(1 - \Delta t \left(r_c^{\text{eff}}(n) + r_o^{\text{eff}}(n)\right)\right)^{t_n}. \tag{36}$$

Given the values of u_n, d_n, t_n measured along our trajectory of duration T, we can infer the effective rates by maximizing the above probability. Assuming that $r^{\text{eff}}\Delta t \ll 1$, we obtain

$$r_c^{\text{eff}}(n) = \frac{d_n}{t_n}, \qquad r_o^{\text{eff}}(n) = \frac{u_n}{t_n}, \tag{37}$$

as estimates for the effective rates. For the full expression (34), the rates are almost independent of n; on the other hand, if the first-order correction is neglected, one obtains n-dependent rates, consistent with the observation that $P_{\text{eq}}(n)$ is not exponential. These are reported in the upper panel of figure 4. In both cases, the rates are consistent with the detailed balance condition $r_c^{\text{eff}}(n)P_{\text{eq}}(n) = r_o^{\text{eff}}(n-1)P_{\text{eq}}(n-1)$.

4.2. Out-of-equilibrium opening

For long molecules, the barrier between the closed and open states may become very large, e.g. $\sim 3000\, k_{\text{B}}T$ for the 50 000 bases λ–DNA at the critical force $f_c = 15.5$ pN [31]. The time necessary to cross this barrier is huge, and full opening of the molecule never happens during experiments. To open a finite fraction of the molecule, the force has to be chosen to be larger than its critical value. The opening can then be modeled as a transient random walk, characterized by pauses at local minima of the free energy and rapid jumps in between [16].

4.2.1. Analytical calculation of the average time spent by the fork on a base. First consider the case of a fixed force acting on the fork while all the other components are at equilibrium as in section 2.3. In the transient random walk, the opening fork spends a finite time around a position n before escaping away and never coming back again in n. The number u_n of opening transitions $n \rightarrow n + 1$ is stochastic and varies from experiment to experiment and base to base. The total number of times the fork visits the base pair n before escaping is given by the sum of the number u_n of transitions from $n - 1$ to n

and of the number $u_{n+1} - 1$ of transitions from $n + 1$ to n. Therefore, the average time spent in n is

$$t_n = \frac{\langle u_n \rangle + \langle u_{n+1} \rangle - 1}{r_o + r_c(n)}, \tag{38}$$

where $1/(r_o + r_c(n))$ is the average time spent in n before each opening or closing step. Let us introduce the probability E_{n+1}^n of never reaching back position n starting from position $n + 1$. The probability P of the number u_n of opening transitions $n \rightarrow n + 1$ during a single unzipping simply reads as

$$P(u_n) = \left(1 - E_{n+1}^n\right)^{u_n - 1} E_{n+1}^n. \tag{39}$$

From equation (39), we have that the average number of openings of bp n is

$$\langle u_n \rangle = \sum_{u_n \geqslant 1} P(u_n)u_n = \frac{1}{E_{n+1}^n}. \tag{40}$$

We are thus left with the calculation of E_{n+1}^n. For infinite force, $E_{n+1}^n = 1$ since the fork never moves backward. For finite force, we write a recursive equation for the probability E_m^n that the fork never comes back to base n starting from base $m (\geqslant n + 1)$:

$$E_m^n = q_m E_{m-1}^n + (1 - q_m) E_{m+1}^n, \tag{41}$$

where

$$q_n = \frac{e^{g_{\text{ss}}(f)}}{e^{g_{\text{ss}}(f)} + e^{g_0(b_n, b_{n+1})}} \tag{42}$$

is the probability of closing base n and $1 - q_n$ is the probability of opening it at each step. Note that for forces larger than the critical force, we have $q_n < \frac{1}{2}$: the random walk is submitted to a forward drift and is transient. The boundary conditions for equation (41) are $E_n^n = 0$ and $E_m^n = 1$ for $m \rightarrow \infty$.

For a homogeneous sequence, the escape probability is $E = (1 - 2q)/(1 - q)$. For a heterogeneous sequence by defining $\rho_m^n = \frac{E_m^n}{E_{m+1}^n}$, we obtain the Riccati recursion relation:

$$\rho_n^n = 0; \qquad \rho_{m+1}^n = \frac{1 - q_{m+1}}{1 - q_{m+1}\rho_m^n} \qquad \text{for } n \geqslant m. \tag{43}$$

Equation (43) can be solved numerically for a given sequence. Then, the escape probability starting from $n + 1$ is

$$E_{n+1}^n = \prod_{m \geqslant n+1} \rho_m^n, \tag{44}$$

and the average time spent in the base n is then obtained from (40) and (38).

4.2.2. Results from the dynamical model. To check whether these theoretical predictions are affected by dynamical fluctuations of the bead, linkers and unzipped strands, we have carried out simulations with the model of section 3. We have carried out 160 unzippings of the λ-phage sequence at a force of 17 pN for $T = 100$ s (physical time), with the same molecular construct of section 4.1 ($N_{\text{ds}} = 3120$ base pairs of dsDNA linkers on a side plus $N_{\text{ss}}^0 = 40$ bases of the ssDNA linker at each side of the DNA to be open). For such a construct, the equilibrium extension of the polymers for n open base pairs is $2N_{\text{ss}}l_{\text{ss}} + N_{\text{ds}}l_{\text{ds}}$, where $l_{\text{ds}} = 0.3337$ nm, $l_{\text{ss}} = 0.4758$ nm and $N_{\text{ss}} = N_{\text{ss}}^0 + n$. The stiffness of the polymers
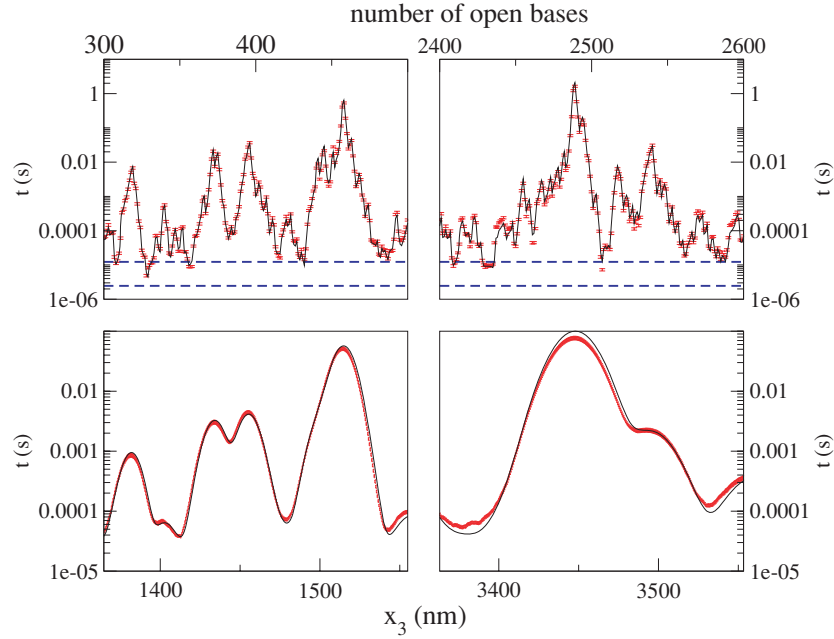
Figure 5. Top: average time spent by the fork on position *n*. Bottom: time spent by the whole setup at an extension between x_3 and $x_3 + \Delta x$, with $\Delta x = 0.5$ nm. The black line in both figures represents the theoretical predictions from section 4.2.1. The red points are the results from the simulation. Standard deviations are represented by error bars in the top panels and by the thickness of the red curves in the bottom panels.

is $1/k_{\text{eff}} = N_{\text{ss}}/k_{\text{ss}}^m + N_{\text{ds}}/k_{\text{ds}}^m$ with $k_{\text{ss}}^m = 160.5$ pN nm^{-1} and $k_{\text{ds}}^m = 1450$ pN nm^{-1}. The relaxation times of the polymers are of the order of 0.1 ms for about 400 unzipped bases and 1 ms for about 2500 open bases, and are larger than the characteristic times of about 2×10^{-6} s needed to open a weak base and of about 10^{-5} s needed to open a strong base.

We plot in figure 5 the average time spent by the fork at location *n* for two portions of the sequence, corresponding to about 400 and 2500 open base pairs. The agreement between the theoretical and numerical estimates of the times is excellent, meaning that the fluctuations of extensions of the polymers and the dynamics of the bead induce negligible changes on the rates of opening and closing, as seen close to the critical force in section 4.1.

As experiments do not give direct acces to the time spent by the fork at location *n*, we show in figure 5 (bottom) the time $t(x_3)$ spent by the unzipped ssDNA between extensions x_3 and $x_3 + \text{d}x$. These times are compared to their values assuming that the positions x_3 of the beads are randomly drawn from the equilibrium measure:

$$t(x_3) = \sum_n t_n P(x_3|n), \qquad (45)$$

where t_n is calculated from (38) and $P(x_3|n)$ is calculated from an argument similar to that used in section 2.2.1 and can be written up to the quadratic order around the saddle point as

$$P(x_3|n) = \sqrt{\frac{\beta k_{\text{eff}}(f)}{2\pi}} \, \mathrm{e}^{-\beta \frac{k_{\text{eff}}(f)}{2}(x_3 - N_{\text{ds}}l_{\text{ds}}(f) - 2N_{\text{ss}}l_{\text{ss}}(f))^2}. \qquad (46)$$

The agreement is, again, excellent.

Figure 5 and equation (45) show that $t(x_3)$ gets contributions from the times spent by the fork on a set of bases whose number depends on the magnitude of the equilibrium

fluctuations of the linkers. These equilibrium fluctuations increase with the length of ssDNA, e.g. $\delta x_3 \simeq 5$ nm for 400 unzipped base pairs and $\delta x_3 \simeq 12$ nm for 2500 unzipped bases. Therefore, as the number *n* of unzipped base pairs increases, the characteristic curve of $t(x_3)$ gets more and more convoluted (compare left-bottom and right-bottom panels in 5).

In figure 6 we compare the value of the ssDNA extension from one unzipping, x_3, to its average value at equilibrium, x_3^{eq}, as a function of the number of unzipped base pairs *n*. The fluctuations in the extension are compatible with the equilibrium deviations. Again, no clear out-of-equilibrium effect is observed. The reason is that, even if the single strand is not relaxed in the opening time of a base, the fork goes back and forward around a given location before moving away. Therefore, the quantities we have measured are averaged on the number of times a base pair is opened and are close to their mean value even in a single unzipping. This can be deduced from figure 5 by comparing the total time spent on a base (points) with the time to open a base (dashed lines)

5. Unzipping at fixed extremities

5.1. Correlation functions

One of the main advantages of considering the dynamics of the linkers and of the beads is that it allows us to compute autocorrelation functions and to explore the interaction between different parts of the setup, a task which would be impossible from *a priori* calculations.

We have performed a few simulations with the setup shown in figure 1(A) where the spring constant of the first optical trap of extension x_1 is 0.1 pN nm^{-1} and the second
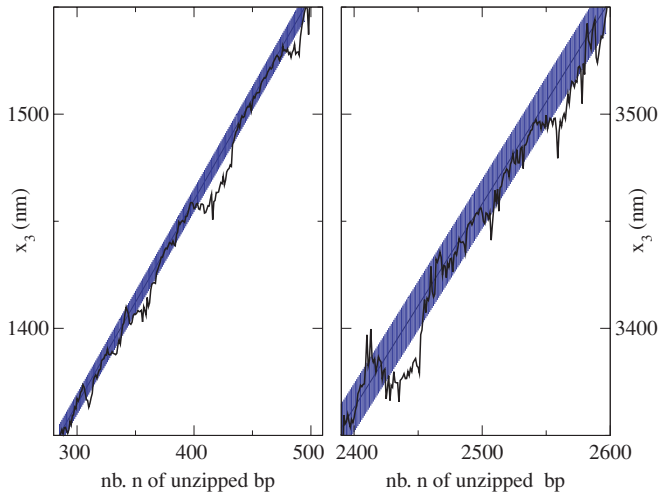
Figure 6. Total extension x_3 of the setup in figure 1(B) at a fixed number n of unzipped bases for a single unzipping (black line). If the fork visits the same base n twice or more, we plot the average of the extension values. The gray strip represents the average value at equilibrium, $x_3^{eq}(n)$, and the standard deviation around its value at equilibrium.

(x_4) has stiffness 0.512 pN nm^{-1}. The molecule in the fork is uniform with $g_0 = 2.69 k_B T$. The only parameter that is varied across simulations is the distance between the optical traps and thus the typical number of open bases. In figure 7, we show two typical cases. What is evident is that the single strand has two time scales: one which is proper to the fluctuations at n fixed and another which is of the same order of magnitude as the correlation time of the fork. As the number of open bases grows, the fast time scale also grows until it becomes impossible to distinguish the two.

As remarked in section 3.5, our model cannot in principle be used when the linkers are made of $n \gtrsim 1000$ monomers. To check for the importance of force propagation effects, we ran a simulation for $N_{ss} = 9700$ (bottom panel of figure 7) where we cut each linker into nine subunits of 1000 bases each plus a final unit which is connected to the opening fork. Overall, the correlation functions are not much affected by this modification and in particular the correlation times are unaffected within numerical errors. The main effect of cutting the long linkers is that the correlation function of the linker becomes more stretched (i.e. if they are fit with $\exp[-(t/\tau)^{\beta_s}]$, the exponent β_s is slightly smaller). This is to be expected since by cutting the polymer we include more relaxation modes, each with its relaxation time. A wider distribution of relaxation times implies a smaller exponent β_s. In table 5, we compare the results of the numerical simulation with the predictions of section 2.2.1 which do not take into account the interactions between different parts of the setup. While the simulated results for the single-stranded and the double-stranded DNA are not too far off from the prediction, the two springs show a much greater deviation from the theoretical estimates. This prompted us to analyze further the relationship between the fork and the bead position as will be discussed later.

The potential acting on the fork position, in the case of a uniform molecule, is dictated by the stiffness of the rest of
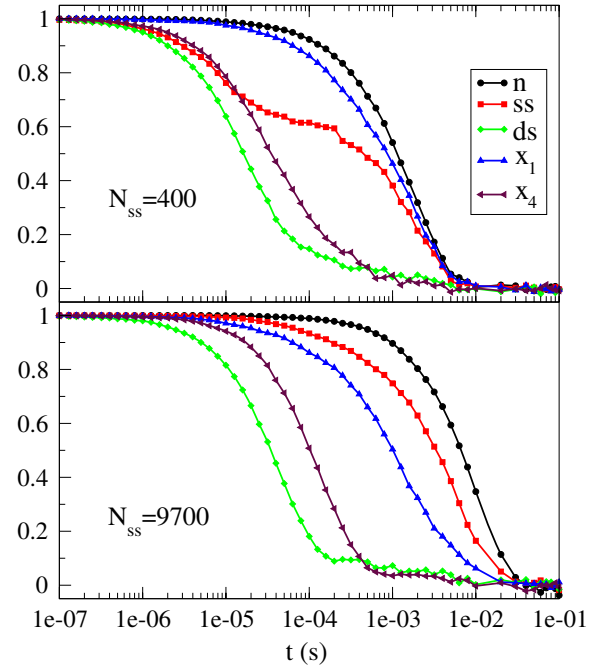

Figure 7. Correlation functions for the setup in figure 1(A) at two different values of the number of open bases, $N_{ss} = 40 + n$.

Table 5. Comparison between the correlation times of the setup in figure 1(A) as computed for an isolated element and the result of a complete numerical simulation. In the case of the fork, we reported as the theoretical value $1/k_{eff}$, which must be multiplied by a viscosity to obtain the relaxation time; it turns out that a viscosity $\sim 8 \times 10^{-5}$ pN s nm^{-1} matches the theoretical and numerical results.

	Theoretical (s)	Numerical (s)
Single strand	$4.83 \times 10^{-11} N_{ss}^2$	$5.4 \times 10^{-11} N_{ss}^2$
Double strand	4.96×10^{-5}	$\sim 3 \times 10^{-5}$
Spring x_1	1.67×10^{-4}	$\sim 1.5 \times 10^{-3}$
Spring x_4	3.26×10^{-4}	$\sim 7 \times 10^{-5}$
Fork N_{ss}	$\propto 14.2 + 0.013 N_{ss}$	$1.3 \times 10^{-3} + 8.4 \times 10^{-7} N_{ss}$

the setup only as seen in section 2.2.1. That is to say that n experiences a harmonic potential with the spring constant proportional to k_{eff}; this in turn predicts correlation times that are proportional to $\frac{1}{k_{eff}}$ which has a linear dependence on n. This behavior is in very good agreement with the data that have been extracted from numerical simulations.

5.2. Mutual information between the bead position and fork location

Figure 9 shows the dynamical correlations of the fork and bead positions. The two beads have different correlation functions due to the difference in their stiffnesses: $k = 0.5$ pN nm^{-1} for bead 1 and $k = 0.1$ pN nm^{-1} for bead 2. After an initial decay (taking place over a time proportional to $1/k$ from section 2.3.3), the bead correlations exhibit a quasi-plateau behavior whose height is roughly proportional to $1/k$. The plateau reflects the correlation between the motion of the bead and that of the fork on time scales of the order of the equilibration
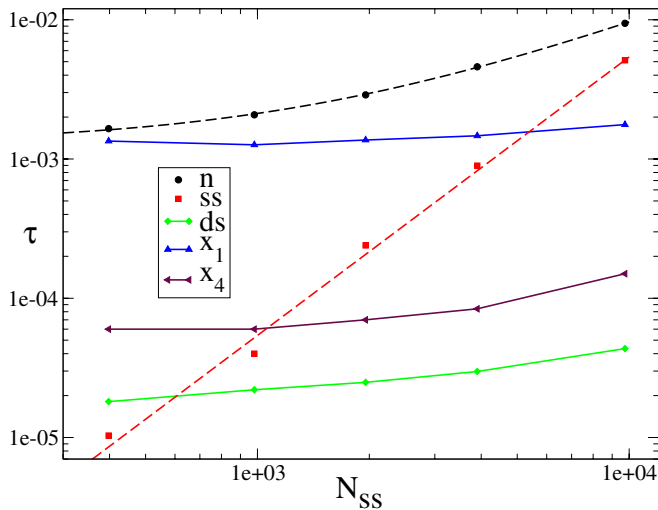
Figure 8. Relaxation times of the correlation functions in figure 7 as a function of the number of open bases. In the case of the single strand (ss), only the fast relaxation time is plotted. For the fork and the single strand, dashed lines indicate a fit to $\tau_n = A + BN_{ss}$ (with $A = 1.3 \times 10^{-3}$ and $B = 8.4 \times 10^{-7}$) and $\tau_{ss} = CN_{ss}^2$ (with $C = 5.4 \times 10^{-11}$ s). For the others, full lines are guides to the eye.

time of the fork. It appears that soft beads allow one to track the location of the fork better than stiffer beads.

In the following, we will give a closer look at the dependence of these correlations on the optical trap stiffness; to do so we construct a setup as in figure 1(*A*), but where the stiffness of the optical trap on the left is kept constant at 0.512 pN nm^{-1} while the stiffness of that on the right is varied across two orders of magnitude[3].

To give quantitative support to this statement we define the mutual information I between the position of the bead in the optical trap, x_4, and the number of open base pairs, n:

$$I(x_4, n) = \sum_n \int dx_4 P(x_4, n) \log \left(\frac{P(x_4, n)}{P(x_4) P(n)} \right), \quad (47)$$

where $P(x_4, n)$ is the joint probability density for the bead to be at position x_4 while there are n open base pairs; $P(n)$ and $P(x_4)$ are the two marginals. Note that the definition of mutual information does not suffer from the problems which arise with entropy when we switch between a continuous and a discrete definition; that is to say that binning with sufficiently small bins does not change the mutual information.

I can be easily computed by keeping track of the times passed at a given bead position and the given number of open bases during a run of the simulation. As stressed before, the fact that the x_4 coordinate must be binned has negligible effects on the computation of entropy. For very large stiffnesses the amplitude of the oscillations of the bead can become very small, and thus a lack of sensitivity in the measure of the position of the bead could become an issue. Fortunately, the current state of the art in the optical trap cannot attain stiffnesses larger than, say, 1 pN nm^{-1} with

---

[3] An attentive reader might have noted that we changed the stiffness of the right bead compared to what it was in the previous section; the rationale behind this choice is to keep its value at the center of the range in which we will vary the other.
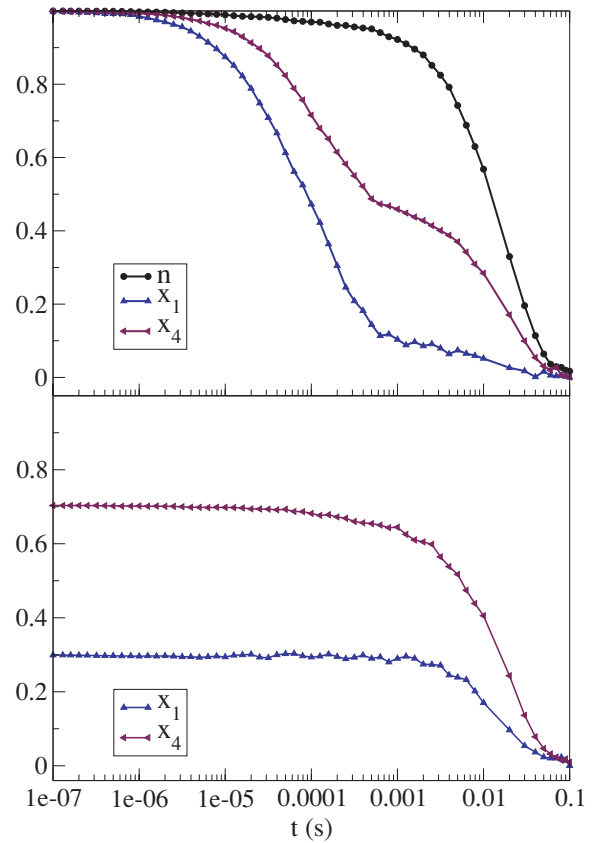

Figure 9. Top: autocorrelation functions for the setup in figure 1(*A*) when the molecule to unzip is a block copolymer composed of alternating stretches of ten strong pairs and ten weak pairs. This way the fork correlation time is greatly increased allowing us to view effects on the two traps of different optical stiffnesses. Bottom: correlation functions between one of the two beads and the number of open base pairs. Values have been normalized so that the value at zero time difference is $\rho = \langle x_i n \rangle / \sqrt{\langle x_i^2 \rangle \langle n^2 \rangle}$.

micrometer beads [32]. In this regime, the fluctuations of the bead are dominated by the stiffness of the trap and thus we can say that $\langle \delta x_4^2 \rangle \sim (\beta k_2)^{-1}$; see equation (19). Comparing the fluctuations of the bead position with the sub-nanometer precision Δ over its location yields

$$\frac{\sqrt{\langle \delta x_4^2 \rangle}}{\Delta} \simeq 10\text{–}50, \quad (48)$$

which is much larger than unity.

Figure 10 shows that the mutual information I only weakly depends on the sequence but strongly depends on the stiffness k of the trap. This behavior can be understood very intuitively. Right after a base pair opens or closes, the whole setup in a fixed-force experiment has to give way; the less rigid an element of the setup is compared to the rest, the more it will accommodate for the change in n.

We conclude that, in a single measurement, soft traps give more information on the fork location than stiff traps. However, I is the mutual information between the fork and bead locations *per measure*. As we have seen in section 5.1, the correlation times extracted from the simulations decrease with k and, as k grows, more and more uncorrelated measures
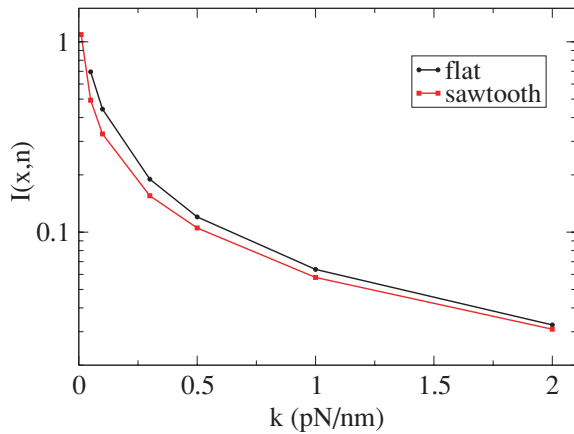
Figure 10. Mutual information *I* between x_4 and *n* as a function of the trap stiffness, *k*. Black circles are computed on a uniform sequence, while red squares are measured on the sawtooth potential described in the caption to figure 9.

can be done in the same amount of time. It is thus expected that information *per unit of time* is not maximal for small values of *k*. In other words, stiffer traps give worse quality but more frequent signals on the location of the fork. Finding the optimal value of *k* would require a detailed analysis of the correlation times of the bead and of the fork. In particular, the size of the bead would affect the optimal value for *k* through the viscosity coefficient, but not the information per measure, *I*. However this dependence should not be crucial since the bead size cannot be much varied in experiments: it can be neither too small to exert a sufficient force nor too large due to the size of the physical setup.

6. Conclusion

This paper has been devoted to the presentation of a dynamical model for the different components of the setups used in the unzipping of single DNA molecules under a mechanical action. Compared to previous studies, our model does not assume *a priori* that the polymers in the molecular construction are at equilibrium but takes into account their relaxation dynamics. It is important to stress out that the dynamical description for the linkers and the unzipped part of DNA is coarse grained: the basic unity is the polymers themselves and not the monomers they are made of.

As a consequence, each polymer is associated with a unique relaxation time. The assumption is justified as long as these times are comparable to the typical opening or closing time of a single base pair. Longer polymeric chains, e.g. ssDNA strands with a few thousand bases, need to be modeled in a more detailed way; more precisely, they should be divided into short enough segments along which the force can be considered as uniform on the time scales associated with the fork motion. Although in this paper we did not observe any important force propagation effect, these might be more important in strongly nonequilibrium situations such as opening at constant (high) velocity. We plan to simulate unzippings with such molecular constructions in the near future to understand how force propagation across the

polymeric segments can affect the effective rates for closing base pairs in such situations.

One of our results is that one has to be very careful with the expression of the free energies (entering the dynamical rates) for short polymers, be they linkers or ssDNA unzipped strands. Use of the free energy per monomer, obtained from force–extension measures on long molecules, as usually done in the literature, can lead to erroneous results. We have shown that finite-size corrections to the energetic contributions and the dynamical rates have to be taken into account.

As a main advantage, the code we have developed is versatile: we can easily change setups, for example use a fixed-force or fixed-position ensemble, and change the number and types of linkers and of traps for the beads. We have found that, in fixed-force unzippings, the opening and closing rates for the fork are not affected by the force fluctuations coming from the polymeric chains. For small linkers and a number of unzipped base pairs, indeed, force fluctuations are large but fast, and are averaged out on the characteristic opening–closing time of a base pair. For large linkers or a number of unzipped bases force fluctuations are slow but small, and therefore do not change the dynamic of the opening fork. We have also performed unzipping simulations at large forces where the opening dynamics is transient, and found that the average time spent by the unzipped strands at a given extension is accurately predicted from the time spent by the fork on a base convoluted by the equilibrium fluctuations of ssDNA. Moreover, the extension between the extremities at a fixed number of open base pairs in a single unzipping experiment is compatible with equilibrium fluctuations of ssDNA and linkers. The program could be easily adapted to unzipping at constant velocity, where non-equilibrium effects are likely to be more important.

Our study suggests that one measure of the position of the bead in soft traps gives more information on the location of the fork than in the case of stiffer traps. This statement is however to be considered with caution. Beads in stiffer traps reach equilibrium on shorter time scales, and the overall rate of information per unit time could be higher in stiffer traps. While purely qualitative at this stage, such a statement is relevant to the study of the inverse problem of unzipping, that is, inferring the sequence of the DNA molecule from the unzipping signal. We hope that the present dynamical modeling will be useful to assess the rate at which information on the sequence could be acquired from mechanical single molecule experiments.

Appendix A. Langevin dynamics of coupled polymers

One of the simplest models of polymer dynamics is that proposed by Rouse [44], where the polymer is described as

a chain of beads which are modeled as Brownian particles, linked by harmonic springs.

While it is true that this model is very crude because it ignores hydrodynamic interactions and exclude volume effects, it has the huge advantage of being largely solvable. Therefore, we will now use it as the basis for a few considerations that will then be generalized to more realistic models.

Our aim is to write a system of coupled equations for the time evolution of a certain number of marked points on a (hetero)polymer. One of these points will be for instance the location of the opening fork. In the case of a double DNA strand attached to a single strand, one point will mark the location where the two different polymers are attached (see the examples in figure 1). Note that if the marked points we focus on are far apart, only the slower modes of the system will be relevant, as the fast modes describe local relaxations of the chain. Therefore, in the following, we want to focus on a long wavelength/long time effective description of the chain.

A.1. The dynamics of a single polymer

A.1.1. The model and its normal modes. As the simplest case we consider a polymer composed of N identical springs, each with an identical link at one end. The first is connected to a wall that has infinite mass (or, better still in this framework, infinite viscosity) and on the last a force f is exerted. The Langevin equations describing such a polymer can be written as

$$
\begin{cases}
\gamma_m \dot{u}_1 = -2k_m u_1 + k_m u_2 + \eta_1 \\
\quad\vdots \\
\gamma_m \dot{u}_n = -2k_m u_n + k_m u_{n-1} + k_m u_{n+1} + \eta_n \\
\quad\vdots \\
\gamma_m \dot{u}_N = -k_m u_N + k_m u_{N-1} + f + \eta_N,
\end{cases}
\tag{A.1}
$$

where η_i are white Gaussian noises of zero mean and variance:

$$
\langle \eta_i(t)\eta_j(0) \rangle = 2k_{\mathrm{B}}T\delta_{ij}\delta(t). \tag{A.2}
$$

Let us for the moment neglect the noise term. Then, defining $\tau_m = \gamma_m/k_m$, we can formally rewrite these equations as

$$
\tau_m \dot{u}_n = -2u_n + u_{n-1} + u_{n+1}, \qquad \forall\, n, \tag{A.3}
$$

supplemented by the boundary conditions

$$
u_0 \equiv 0, \qquad u_{N+1} \equiv u_N + f/k_m. \tag{A.4}
$$

A standard way to find the normal modes of the above linear system is to search for solutions of the form $u_n(t) = u_n(0)\exp(-\lambda t/\tau_m)$. One can easily show that the general solution satisfying the first boundary condition $u_0 = 0$ has the form

$$
u_n(t) \propto \sin(qn)\exp(-\lambda(q)t/\tau_m),
$$
$$
\lambda(q) = 2(1 - \cos(q)). \tag{A.5}
$$

The second boundary condition (A.4) requires that $u_{N+1}(t) - u_N(t) = f/k_m = \text{const}$. Since we can always add the constant value to $u_{N+1}(t)$, we can replace this boundary condition by $u_{N+1}(t) = u_N(t)$. This requires that $\sin(qN) \sim \sin(q(N+1))$; then $q = (\pi/2 + p\pi)/N$. The slowest mode then corresponds to $q = \pi/2/N$, which for large N gives a relaxation time

$$
\tau(N) = \tau_m/\lambda(\pi/2/N) \sim \frac{4}{\pi^2}\tau_m N^2, \tag{A.6}
$$

which proves the validity of the scaling in equation (20).

A.1.2. Recurrence equations for a fixed end. We now want to write a system of coupled equations for a certain number of points on the polymer by integrating out us we are not interested in. To begin, we focus on the end point u_N.

It is convenient to perform a Laplace transformation and write

$$
u_n(t) = \int_0^\infty d\lambda\, u_n(\lambda)\, e^{-\lambda t/\tau_m}. \tag{A.7}
$$

Then equation (A.5) becomes, in Laplace space,

$$
(2 - \lambda)u_n(\lambda) = u_{n+1}(\lambda) + u_{n-1}(\lambda), \tag{A.8}
$$

with the same boundary conditions $u_0(\lambda) \equiv 0$, and $u_{N+1}(\lambda) - u_N(\lambda) = (f/k_m)\delta(\lambda)$. For $\lambda \neq 0$, the latter condition reduces to $u_{N+1}(\lambda) = u_N(\lambda)$ as discussed above for the normal mode analysis.

We introduce a function

$$
\zeta_{n-1}(\lambda) = u_{n-1}(\lambda)/u_n(\lambda). \tag{A.9}
$$

Substituting the latter relation in (A.8), we get

$$
(2 - \lambda - \zeta_{n-1}(\lambda))u_n(\lambda) = u_{n+1}(\lambda), \tag{A.10}
$$

from which we get a Riccati recurrence equation

$$
\begin{cases}
\zeta_0(\lambda) = 0 & \text{(due to } u_0 = 0), \\
\zeta_n(\lambda) = \dfrac{1}{2 - \lambda - \zeta_{n-1}(\lambda)}.
\end{cases}
\tag{A.11}
$$

This recurrence can be solved and the function $\zeta_n(\lambda)$ computed for all n.

Since we are interested in the large time limit, we can expand the function $\zeta_n(\lambda)$ for small λ; we obtain

$$
\zeta_n(\lambda) = \frac{n}{n+1} + \frac{n(1+2n)}{6(1+n)}\lambda + \frac{n(6 + 19n + 16n^2 + 4n^3)}{180(1+n)}\lambda^2
$$
$$
+ O(\lambda^3). \tag{A.12}
$$

One obtains the effective equation for u_N by substituting the above expression in (A.10) and setting $n = N$. Keeping only the linear term in λ and the leading terms in $N \gg 1$, we get

$$
\left(1 + \frac{1}{N} - \lambda\frac{N}{3}\right)u_N(\lambda) = u_{N+1}(\lambda). \tag{A.13}
$$

Moving back to the time domain, we obtain

$$
\tau_m\frac{N}{3}\dot{u}_N = -\frac{1}{N}u_N + (u_{N+1} - u_N), \tag{A.14}
$$

which is equivalent, using the boundary condition $u_{N+1} - u_N = f/k_m$, to

$$
\frac{\gamma_m N}{3}\dot{u}_N = -\frac{k_m}{N}u_N + f. \tag{A.15}
$$

In this way, we got an effective equation for the endpoint of the polymer that is still a linear first-order differential equation and takes into account only the slowest mode of the chain.

There is however an inconvenience: in fact a straightforward computation shows that the relaxation time obtained from equation (A.15) is $\tau(N) = \tau_m N^2/3$ that differs by a factor $\pi^2/12$ from the correct value given by equation (A.6). The origin of this discrepancy clearly lies in the fact that the expansion we made in equation (A.12) is not convergent at fixed λ for $n \to \infty$, as successive terms in the series are of order $n^{2p-1}\lambda^p$.
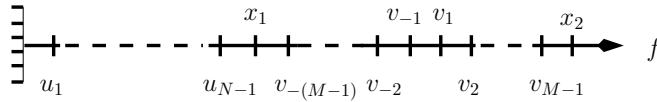
Figure A1. Two joint polymers subjected to an external force f. x_1 marks the endpoint of the first polymer made of N links whose endpoints are $u_1, u_2, \ldots, u_{N-1}, u_N \equiv x_1$. The second polymer originates from x_1 and is made of $2M - 1$ links, whose endpoints are $v_{-(M-1)}, v_{-(M-2)}, \ldots, v_{-1}, v_1, \ldots, v_{M-2}, v_{M-1}, x_2$.

Let us then go back to the computation of the normal modes of the system within this formalism. The second boundary condition $u_{N+1}(\lambda) = u_N(\lambda)$ implies $\zeta_N(\lambda) = 1$. The normal modes are the solutions of this equation with respect to λ. One can show from the exact expression of $\zeta_N(\lambda)$ that

$$\lim_{N \to \infty} N[\zeta_N(\widetilde{q}^2/N^2) - 1] = -\widetilde{q} \cot(\widetilde{q}) \equiv \widetilde{\zeta}(\widetilde{q}). \quad (A.16)$$

The zeroes of this function are $\widetilde{q} = \pi/2 + k\pi$; therefore, the solutions of $\zeta_N(\lambda) = 1$ tend for large N to $\lambda = (\pi/2 + p\pi)^2/N^2$, in agreement with the exact result of the previous section. An inspection of equations (A.12) and (A.16) shows that the small λ expansion of $\zeta_N(\lambda)$ is equivalent to performing a small $\widetilde{q}$ expansion of $\widetilde{\zeta}(\widetilde{q})$ in order to find its first zero. This indeed yields $\widetilde{\zeta}(\widetilde{q}) \sim -1 + \widetilde{q}^2/3$ that gives $\widetilde{q} = \sqrt{3}$ for the first zero that gives back $\tau(N) = \tau_m N^2/3$.

Then one can check that a higher order expansion in λ (or equivalently in $\widetilde{q}$) produces a more accurate result; indeed the series of $\widetilde{\zeta}(\widetilde{q})$ converges for $\widetilde{q} < \pi$ while the zero is located at $\widetilde{q} = \pi/2$. It is easy to show that if one truncates the series to order p, the difference between the solution and the true zero is exponentially small in p.

A.1.3. Discussion The conclusion of this section is that equation (A.15) is a correct description of the dynamics of the end of the polymer in the limit of large N and large times. While it captures the correct scaling with N of the relaxation time, the coefficient is wrong by a factor of $\pi^2/12 \sim 0.82$. Still, this is quite satisfactory for our purposes since the experimental error in the determination of τ_m is of the same order of magnitude. Better approximations can be obtained by truncating the expansion of $\zeta_N(\lambda)$ to higher orders in λ, therefore obtaining a higher order differential equation for $u_N(t)$.

In the following, we will derive the coupled equation for many marked points along the chain, limiting ourselves to the first-order truncation. This produces first-order differential equations of the Langevin type.

A.2. Dynamics of two coupled polymers

We will now show how to use this formalism to derive coupled equations for different points on a composite polymer. We continue neglecting the noise, which we will reintroduce at the end of this section.

As a simple example, let us consider the polymer drawn in figure A1. It is composed of N monomers of type 'U' linked to $2M - 1$ monomers of type 'V'. The two types of monomers might differ in the value of the microscopic spring constant,

bead viscosity, etc. If the monomers are identical, then we are just marking a point in the middle of a polymer.

The effective equation for the endpoint of polymer U can be derived following the analysis of the previous section. We denote $x_1 \equiv u_N$ and we get

$$\gamma_m^U \frac{N}{3} \dot{x}_1(t) = -\frac{k_m^U}{N} x_1(t) + k_m^V(v_{-(M-1)}(t) - x_1(t)), \quad (A.17)$$

where the last term is the 'external' force that the polymer V exerts on U.

A.2.1. Integration of the V polymer. Now we want to integrate out all the monomers $v_{-(M-1)}, \ldots, v_{M-1}$ in order to obtain the coupling between x_1 and x_2. To this aim, and in order to keep the formalism symmetric, we can start from the middle of the polymer V by integrating simultaneously v_{-1} and v_1 in order to obtain effective equations for v_{-2} and v_2, and so on. In Laplace space (note that now in equation (A.7) $\tau_m = \tau_m^V$), the equations for $v_{\pm 1}$ have the form

$$\begin{aligned} (2 - \lambda) v_{-1}(\lambda) &= v_{-2}(\lambda) + v_1(\lambda), \\ (2 - \lambda) v_1(\lambda) &= v_2(\lambda) + v_{-1}(\lambda). \end{aligned} \quad (A.18)$$

These can be easily solved to get $v_{\pm 1}$ as a function of $v_{\pm 2}$. Iteration leads to the following form for the equation after n steps:

$$\begin{aligned} \xi_n(\lambda) v_{-n-1}(\lambda) &= v_{-n-2}(\lambda) + \eta_n(\lambda) v_{n+1}(\lambda), \\ \xi_n(\lambda) v_{n+1}(\lambda) &= v_{n+2}(\lambda) + \eta_n(\lambda) v_{-n-1}(\lambda). \end{aligned} \quad (A.19)$$

One can check that this form is stable under one step of iteration and the following recursion relations are obtained:

$$\begin{cases} \xi_0 = 2 - \lambda, \\ \eta_0 = 1, \\ \xi_{n+1} = 2 - \lambda - \dfrac{\xi_n}{\xi_n^2 - \eta_n^2}, \\ \eta_{n+1} = \dfrac{\eta_n}{\xi_n^2 - \eta_n^2}, \end{cases} \quad (A.20)$$

where the initial values are determined by consistency between (A.18) and (A.19) for $n = 0$. These recurrences are easily solved by introducing the two quantities $A_n = 1/(\xi_n - \eta_n)$ and $B_n = 1/(\xi_n + \eta_n)$ respectively; these satisfy the same recurrence in (A.11) except for the initial condition which is different and determined according to (A.20).

At the leading order in $n \to \infty$ and at first order in λ, we get

$$\xi_n(\lambda) = 1 + \frac{1}{2n} - \frac{2n}{3}\lambda, \qquad \eta_n(\lambda) = \frac{1}{2n} + \frac{2n}{6}\lambda. \quad (A.21)$$

Finally, one obtains from this procedure a coupled equation for $v_{-(M-1)}$ and v_{M-1} where $x_1 \equiv v_{-M}$ and $x_2 \equiv v_M$ also appear.

A.2.2. Coupled effective equations. To obtain the coupled effective equations, one starts from the following system:

$$\begin{cases} -\gamma_m^U \dfrac{N}{3} \dfrac{\lambda}{\tau_m^V} x_1 = -\dfrac{k_m^U}{N} x_1 + k_m^V(v_{-M+1} - x_1), \\ \xi_{M-2}(\lambda) v_{-M+1}(\lambda) = x_1 + \eta_{M-2}(\lambda) v_{M-1}(\lambda), \\ \xi_{M-2}(\lambda) v_{M-1}(\lambda) = x_2 + \eta_{M-2}(\lambda) v_{-M+1}(\lambda), \\ (1 - \lambda) x_2 = v_{M-1} + f, \end{cases} \quad (A.22)$$

where the first equation is just the Laplace transform of equation (A.17) (recall that we use the definition of Laplace transform (A.7) with $\tau_m = \tau_m^V$), the second and third equations are equation (A.19) for $n = M - 2$ and the last equation is the Laplace transform of the equation for x_2, which in the time domain reads as $\gamma_m^V \dot{x}_2 = -k_m^V(x_2 - v_{M-1}) + f$.

Eliminating v_{-M+1} and v_{M-1} from these equations, using the recurrence equations (A.20) and the result (A.21) we finally get the coupled equations:

$$\begin{cases} \left(\gamma_m^U \dfrac{N}{3} + \gamma_m^V \dfrac{2M}{3}\right) \dot{x}_1 + \gamma_m^V \dfrac{2M}{6} \dot{x}_2 \\ \quad = -\dfrac{k_m^U}{N} x_1 + \dfrac{k_m^V}{2M}(x_2 - x_1), \\ \gamma_m^V \dfrac{2M}{3} \dot{x}_2 + \gamma_m^V \dfrac{2M}{6} \dot{x}_1 = -\dfrac{k_m^V}{2M}(x_2 - x_1) + f. \end{cases} \quad (A.23)$$

At this point we reintroduce the free energy of the polymer chain, defining $N_1 \equiv N$ and $N_2 \equiv 2M - 1 \sim 2M$:

$$F(x_1, x_2) = \frac{k_m^U}{2N_1} x_1^2 + \frac{k_m^V}{2N_2}(x_2 - x_1)^2, \quad (A.24)$$

and a matrix

$$\Gamma \equiv \begin{pmatrix} \gamma_m^U \frac{N_1}{3} + \gamma_m^V \frac{N_2}{3} & \gamma_m^V \frac{N_2}{6} \\ \gamma_m^V \frac{N_2}{6} & \gamma_m^V \frac{N_2}{3} \end{pmatrix} \quad (A.25)$$

so that we can write the above system as

$$\Gamma_{ij} \dot{x}_j = -\frac{\partial F}{\partial x_i} + f_i + \eta_i, \quad (A.26)$$

where $\vec{f} = (0, f)$ is the external force vector and we reintroduced the noise term $\vec{\eta}$ that we neglected before.

The correlation function of the noise at this point is determined by the requirement that the fluctuation–dissipation relation is verified. This imposes that

$$\langle \eta_i(t)\eta_j(0) \rangle = 2k_{\rm B}T \Gamma_{ij}\delta(t). \quad (A.27)$$

A.3. Beads

At this point, we should add the beads that are used for the optical manipulation of polymers. These beads are optically tweezed or subjected to magnetic fields in order to apply forces to the polymers. In the former case, the force acting on the bead is a harmonic force $f = -k(x - X)$, while in the latter it is constant, $f = f_{\rm ext}$. Each bead is characterized by a friction coefficient that can be computed using the Stokes law; we denote it by γ. Typically they are of the order of 10^{-5} pN s nm^{-1}, i.e. much bigger than the microscopic viscosity of the polymers $\gamma_m \sim 10^{-8}$ pN s nm^{-1}.

In the presence of a bead attached to the endpoint of a polymer, the equations of motion (A.1), (A.18), etc, remain valid, but one should add the contribution of γ to the viscosity of the coordinate describing the position of the bead. For instance, if there is a bead attached to the endpoint u_N, the last equation of (A.1) reads as

$$(\gamma + \gamma_m)\dot{u}_N = -k_m u_N + k_m u_{N-1} + f + \eta_N. \quad (A.28)$$

Then the above derivation still holds because the last equation is not used until the end. The only modification will be the

inclusion of γ on the diagonal element Γ_{ii} corresponding to the coordinate of the bead.

Therefore to describe the beads attached to the end of the molecular construction in figure 1, we modify the matrix Γ as above, and in case A, we add to the free energy a term $\frac{1}{2}k(x_4 - X)^2$, while in case B we add a term $-f_{\rm ext}x_3$.

In the case of figure 1(A), one also has to include the left bead. In this case, if we call V the first polymer after the bead, we can start from a system of equations identical to (A.22), but with the first equation replaced by

$$-\gamma \dot{x}_1 = -kx_1 + k_m^V(v_{-M+1} - x_1). \quad (A.29)$$

This will again lead to (A.26) with

$$\Gamma \equiv \begin{pmatrix} \gamma + \gamma_m^V \frac{N_2}{3} & \gamma_m^V \frac{N_2}{6} \\ \gamma_m^V \frac{N_2}{6} & \gamma_m^V \frac{N_2}{3} \end{pmatrix} \quad (A.30)$$

and

$$F(x_1, x_2) = \frac{k}{2} x_1^2 + \frac{k_m^V}{2N_2}(x_2 - x_1)^2. \quad (A.31)$$

A.4. Description of a generic setup

The arguments of the previous section suggest that in the general case, a bead can be treated 'as a particular instance of a polymer'. In other words, we can consider the setups in figure 1 as chains of p joint elements $U = U_1, U_2, \ldots, U_p$; each element can be an 'optical trap' (i.e. a spring) or a polymer of $N_1, N_2, \ldots, N_p$ monomers respectively (in the case of an optical trap, we set by default $N_i = 1$). The endpoint of each element is denoted by x_i, and $\vec{x} \equiv (x_1, x_2, \ldots, x_p)$ is the state vector of the system (we also define $x_0 \equiv 0$).

Then, the total free energy is $F(\vec{x}) = \sum_{i=1}^p W_{U_i}(x_i - x_{i-1})$ where $W_{U_i}(x) = \frac{1}{2}kx^2$ for an optical trap of stiffness k. Then equation (A.26) holds, with i, j running from 1 to p and the noise correlation matrix is given by (A.27).

The matrix Γ must be constructed as follows. Each diagonal term Γ_{ii}, related to x_i, is the sum of a Stokes term coming from a bead possibly attached to x_i and the contribution coming from the two elements adjacent to x_i (except for $i = p$ when there is only one contribution):

$$\Gamma_{ii} = \gamma + \gamma_m^{U_i} \frac{N_i}{3} + \gamma_m^{U_{i+1}} \frac{N_{i+1}}{3}(1 - \delta_{ip}) \quad (A.32)$$

(the first term is present only if there is a bead attached to x_i). All the off-diagonal elements are zero except those adjacent to the diagonal (i.e. connecting x_i and x_{i+1}) which get a contribution from the polymer connecting these two ends:

$$\Gamma_{i,i+1} = \Gamma_{i+1,i} = \gamma_m^{U_{i+1}} \frac{N_{i+1}}{6}, \qquad i = 1, \ldots, p-1. \quad (A.33)$$

Note that this final formulation is independent of the Gaussian form of $F(\vec{x})$ that we assumed in the derivation; therefore, we will also use it for non-Gaussian polymers substituting the appropriate form of $F(\vec{x})$ in equation (A.26).

To conclude this section, note that a further check of the quality of the first-order approximation can be done as follows. If we consider a single polymer made of $N_1 + N_2$ bases, the corresponding relaxation time is predicted to be $\tau = \tau_m(N_1 + N_2)^2/3$. On the other hand, we could

consider two coupled polymers of N_1 and N_2 bases following equation (A.23) for $y_m^{U,V} = \gamma_m$ and $k_m^{U,V} = k_m$ respectively. The coupled equation can be exactly solved and yields two distinct relaxation times (that typically differ by a factor of 10); the slowest relaxation time can be compared with $\tau = \tau_m (N_1 + N_2)^2/3$. We found that the difference is at most 20%, and the error is maximal for $N_1 \sim N_2$ while it decreases when one of the two polymers is much longer than the other.

Appendix B. Transition rates for the fork dynamics

We now consider a fork n attached to the polymers. For simplicity, we consider the case of a single polymer whose extension is x and free energy is $W(x, n)$. We want to construct a stochastic process that samples the equilibrium distribution $P_{eq}(x, n) = e^{-\beta W(x,n) - G(n;B)}/Z$, where $-G(n; B)$ is the free energy gain in closing the first n bases of DNA, as defined in equation (6).

The random process is constructed as follows. The Langevin equation discussed in the previous section is discretized with time step Δt. If at a given time t the system is in a state (x, n), we allow three possible transitions:

- $(x, n) \to (x + \Delta x, n)$ with rate $H^s(x, n, \Delta x)$,
- $(x, n) \to (x + \Delta x, n + 1)$ with rate $H^o(x, n, \Delta x)$,
- $(x, n) \to (x + \Delta x, n - 1)$ with rate $H^c(x, n, \Delta x)$.

We must have

$$\int \mathrm{d}\Delta x\, H^s(x, n, \Delta x) + H^o(x, n, \Delta x) + H^c(x, n, \Delta x) = 1. \tag{B.1}$$

Moreover we can define rates $r^{s,o,c}(x, n) = \int \mathrm{d}\Delta x\, H^{s,o,c}(x, n, \Delta x)$ that represent the rates to stay, open or close n independent of Δx. In a practical implementation we first decide whether to open, close or stay according to $r^{s,o,c}$, and then extract Δx from the distribution $H^{s,o,c}(x, n, \Delta x)/r^{s,o,c}(x, n)$.

The detailed balance conditions read as

$$
\begin{aligned}
&P(n, x) H^o(x, n, \Delta x) \\
&\quad = P(n + 1, x + \Delta x) H^c(n + 1, x + \Delta x, -\Delta x) \\
&P(n, x) H^c(x, n, \Delta x) \\
&\quad = P(n - 1, x + \Delta x) H^o(n - 1, x + \Delta x, -\Delta x) \\
&P(n, x) H^s(x, n, \Delta x) \\
&\quad = P(n, x + \Delta x) H^s(n, x + \Delta x, -\Delta x).
\end{aligned} \tag{B.2}
$$

We assume that the rate for opening is given by the product of a term that only depends on the binding free energy as in equation (21) and a term corresponding to a standard Langevin step:

$$
\begin{aligned}
H^o(x, n, \Delta x) &= r\,\Delta t\, e^{G(n;B) - G(n+1;B)} \sqrt{\frac{4\pi T \Delta t}{\gamma_n}} \\
&\quad \times \exp\left[-\frac{\gamma_n}{4T\Delta t}\left(\Delta x - \frac{f(x, n)\Delta t}{\gamma_n}\right)^2\right].
\end{aligned} \tag{B.3}
$$

Note that integrating over Δx we find $r^o(x, n) = r\Delta t\, e^{G(n;B) - G(n+1;B)} = r\Delta t\, e^{-g_0(b_{n+1}, b_{n+2})}$, consistent with equation (21).

Now it is easy to show that the following expression for $H^c(x, n, \Delta x)$ follows from the second detailed balance condition:

$$
\begin{aligned}
H^c(x, n, \Delta x) &= r\,\Delta t\, e^{\beta W(x,n) - \beta W(x+\Delta x, n-1)} \sqrt{\frac{4\pi T \Delta t}{\gamma_{n-1}}} \\
&\quad \times \exp\left[-\frac{\gamma_{n-1}}{4T\Delta t}\left(\Delta x + \frac{f(x + \Delta x, n - 1)\Delta t}{\gamma_{n-1}}\right)^2\right]
\end{aligned} \tag{B.4}
$$

and that the first condition is then automatically satisfied. Up to now, we did not specify the form for $f(x, n)$. However for a generic $f(x, n)$, the above rate is not Gaussian. To obtain a Gaussian rate, we assume that

$$f(x, n) = -\frac{\partial W(x, n)}{\partial x}, \tag{B.5}$$

and perform the following simplifications assuming that Δt is small:

$$
\begin{aligned}
H^c(x, n, \Delta x) &= r\,\Delta t\, e^{\beta W(x,n) - \beta W(x, n-1)} \\
&\quad \times e^{\beta W(x,n-1) - \beta W(x+\Delta x, n-1) - \beta f(n-1, x+\Delta x)} \sqrt{\frac{4\pi T \Delta t}{\gamma_{n-1}}} \\
&\quad \times \exp\left[-\frac{\gamma_{n-1}}{4T\Delta t}\left(\Delta x - \frac{f(x + \Delta x, n - 1)\Delta t}{\gamma_{n-1}}\right)^2\right] \\
&\sim r\,\Delta t\, e^{\beta W(x,n) - \beta W(x, n-1)} \sqrt{\frac{4\pi T \Delta t}{\gamma_{n-1}}} \\
&\quad \times \exp\left[-\frac{\gamma_{n-1}}{4T\Delta t}\left(\Delta x - \frac{f(x + \Delta x, n - 1)\Delta t}{\gamma_{n-1}}\right)^2\right. \\
&\quad \left. + \frac{\beta}{2}\frac{\partial^2 W(x, n - 1)}{\partial x^2}\Delta x^2\right].
\end{aligned} \tag{B.6}
$$

Neglecting $O(\Delta x^3)$ one obtains a Gaussian distribution for Δx, and computing the first and second moments of the Gaussian one can see that at the lowest order in Δt it is equivalent to

$$
\begin{aligned}
H^c(x, n, \Delta x) &= r\,\Delta t\, e^{\beta W(x,n) - \beta W(x, n-1)} \sqrt{\frac{4\pi T \Delta t}{\gamma_{n-1}}} \\
&\quad \times \exp\left[-\frac{\gamma_{n-1}}{4T\Delta t}\left(\Delta x - \frac{f(x, n - 1)\Delta t}{\gamma_{n-1}}\right)^2\right].
\end{aligned} \tag{B.7}
$$

From the above expression, we deduce that the rate for closing is $r^c(x, n) = r\Delta t\, e^{\beta W(x,n) - \beta W(x, n-1)}$, and one *first* has to close and then perform a Langevin step with force $f(x, n - 1)$ and friction γ_{n-1}.

Finally, the rate at constant n is simply given by

$$
\begin{aligned}
H^s(x, n, \Delta x) &= [1 - r^o(x, n) - r^c(x, n)] \sqrt{\frac{4\pi T \Delta t}{\gamma_n}} \\
&\quad \times \exp\left[-\frac{\gamma_n}{4T\Delta t}\left(\Delta x - \frac{f(x, n)\Delta t}{\gamma_n}\right)^2\right],
\end{aligned} \tag{B.8}
$$

and it is easy to see that this verifies the third detailed balance equation if equation (B.5) holds and higher orders in Δt are neglected.

To resume, the implementation of the algorithm is as follows.

(1) Choose whether to stay, open or close, with rates $r^{s,o,c}(x,n)$ respectively.

(2) If open, *first* perform a Langevin step at n and *then* increase n by 1.

(3) If close, *first* decrease n by 1 and *then* perform a Langevin step at $n-1$.

(4) If stay, just perform a Langevin step at n.

(5) Go to 1.

The extension of the above derivation to a case where many polymers are present is straightforward, since the only polymers whose rates are coupled with n are the two adjacent ones. All the other polymers are not influenced by n, and one can use standard discretized Langevin dynamics.

References

[1] Turner P C, McLennan A G, Bates A D and White M R H 2000 *Molecular Biology* (Berlin: Springer)

[2] Bloomfield V A, Crothers D M and Tinoco I 2000 *Nucleic Acids: Structures, Properties, and Functions* (Mill Valley, CA: University Science Books)

[3] Bustamante C, Bryant Z and Smith S B 2003 Ten years of tension: single-molecule DNA mechanics *Nature* **421** 423–7

[4] Marko J F and Cocco S 2003 The micromechanics of DNA *Phys. World* **16** 37–41

[5] Smith S B, Finzi L and Bustamante C 1992 Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads *Science* **258** 1122–6

[6] Cluzel P, Lebrun A, Heller C, Lavery R, Viovy J L, Chatenay D and Caron F 1996 DNA: an extensible molecule *Science* **271** 792

[7] Smith S B, Cui Y and Bustamante C 1996 Overstretching B-dNA: the elastic response of individual double-stranded and single-stranded DNA molecules *Science* **271** 795

[8] Essevaz-Roulet B, Bockelmann U and Heslot F 1997 Mechanical separation of the complementary strands of DNA *Proc. Natl Acad. Sci.* **94** 11935–40

[9] Bockelmann U, Essevaz-Roulet B and Heslot F 1998 DNA strand separation studied by single molecule force measurements *Phys. Rev.* E **58** 2386–94

[10] Bockelmann U, Thomen P, Essevaz-Roulet B, Viasnoff V and Heslot F 2002 Unzipping DNA with optical tweezers: high sequence sensitivity and force flips *Biophys. J.* **82** 1537–53

[11] Thomen P, Bockelmann U and Heslot F 2002 Rotational drag on DNA: a single molecule experiment *Phys. Rev. Lett.* **88** 248102

[12] Bockelmann U, Thomen P and Heslot F 2004 Dynamics of the DNA duplex formation studied by single molecule force measurements *Biophys. J.* **87** 3388–96

[13] Manosas M, Collin D and Ritort F 2006 Force-dependent fragility in RNA hairpins *Phys. Rev. Lett.* **96** 218301

[14] Manosas M, Wen J D, Li P T X, Smith S B, Bustamante C, Tinoco I Jr and Ritort F 2007 Force unfolding kinetics of RNA using optical tweezers: II. Modeling experiments *Biophys. J.* **92** 3010

[15] Liphardt J, Onoa B, Smith S B, Tinoco I and Bustamante C 2001 Reversible unfolding of single RNA molecules by mechanical force *Science* **292** 733–7

[16] Danilowicz C, Coljee V W, Bouzigues C, Lubensky D K, Nelson D R and Prentiss M 2003 DNA unzipped under a constant force exhibits multiple metastable intermediates *Proc. Natl Acad. Sci.* **100** 1694–9

[17] Danilowicz C, Kafri Y, Conroy R S, Coljee V W, Weeks J and Prentiss M 2004 Measurement of the phase diagram of DNA unzipping in the temperature–force plane *Phys. Rev. Lett.* **93** 78101

[18] Weeks J D, Lucks J B, Kafri Y, Danilowicz C, Nelson D R and Prentiss M 2005 Pause point spectra in DNA constant-force unzipping *Biophys. J.* **88** 2752–65

[19] Harlepp S, Marchal T, Robert J, Léger J F, Xayaphoummine A, Isambert H and Chatenay D 2003 Probing complex RNA structures by mechanical force *Eur. Phys. J.* E **12** 605–15

[20] van Oijen A M, Blainey P C, Crampton D J, Richardson C C, Ellenberger T and Xie X S 2003 Single-molecule kinetics of λ exonuclease reveal base dependence and dynamic disorder *Science* **301** 1235–8

[21] Perkins T T, Dalal R V, Mitsis P G and Block S M 2003 Sequence-dependent pausing of single lambda exonuclease molecules *Science* **301** 1914–8

[22] Wuite G J, Smith S B, Young M, Keller D and Bustamante C 2000 Single-molecule studies of the effect of template tension on T7 DNA polymerase activity *Nature* **404** 103–6

[23] Maier B, Bensimon D and Croquette V 2000 Replication by a single DNA polymerase of a stretched single-stranded DNA *Proc. Natl Acad. Sci.* **97** 12002

[24] Levene M J, Korlach J, Turner S W, Foquet M, Craighead H G and Webb W W 2003 Zero-mode waveguides for single-molecule analysis at high concentrations *Science* **299** 682–6

[25] Lang M J, Fordyce P M and Block S M 2003 Combined optical trapping and single-molecule fluorescence *J. Biol.* **2**

[26] Sauer-Budge A F, Nyamwanda J A, Lubensky D K and Branton D 2003 Unzipping kinetics of double-stranded DNA in a nanopore *Phys. Rev. Lett.* **90** 238101

[27] Mathé J, Visram H, Viasnoff V, Rabin Y and Meller A 2004 Nanopore unzipping of individual DNA hairpin molecules *Biophys. J.* **87** 3205–12

[28] Lionnet T, Dawid A, Bigot S, Barre F X, Saleh O A, Heslot F, Allemand J F, Bensimon D and Croquette V 2006 DNA mechanics as a tool to probe helicase and translocase activity *Nucl. Acids Res.* **34** 4232

[29] Lionnet T, Spiering M M, Benkovic S J, Bensimon D and Croquette V 2007 Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism *Proc. Natl Acad. Sci.* **104** 19790

[30] Lubensky D K and Nelson D R 2002 Single molecule statistics and the polynucleotide unzipping transition *Phys. Rev.* E **65** 31917

[31] Cocco S, Marko J F, Monasson R, Sarkar A and Yan J 2003 Force–extension behavior of folding polymers *Eur. Phys. J.* E **10** 249–63

[32] Mangeol P, Côte D, Bizebard T, Legrand O and Bockelmann U 2006 Probing DNA and RNA single molecules with a double optical tweezer *Eur. Phys. J.* E **19** 311–7

[33] Greenleaf W J, Woodside M T, Abbondanzieri E A and Block S M 2005 Passive all-optical force clamp for high-resolution laser trapping *Phys. Rev. Lett.* **95** 208102

[34] Cocco S, Monasson R and Marko J F 2002 Force and kinetic barriers to initiation of DNA unzipping *Phys. Rev.* E **65** 41907

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

[36] Woodside M T, Behnke-Parks W M, Larizadeh K, Travers K, Herschlag D and Block S M 2006 Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins *Proc. Natl Acad. Sci.* **103** 6190–5

[37] Baldazzi V, Cocco S, Marinari E and Monasson R 2006 Inference of DNA sequences from mechanical unzipping: an ideal-case study *Phys. Rev. Lett.* **96** 128102

[38] Harris T D *et al* 2008 Single-molecule DNA sequencing of a viral genome *Science* **320** 106

[39] Thompson R E and Siggia E D 1995 Physical limits on the mechanical measurement of the secondary structure of bio-molecules *Europhys. Lett.* **31** 335

[40] Hyeon C, Morrison G and Thirumalai D 2008 Force-dependent hopping rates of RNA hairpins can be estimated from accurate measurement of the folding landscapes *Proc. Natl Acad. Sci.* **105** 9604–9

[41] Zuker M 2000 Calculating nucleic acid secondary structure *Curr. Opin. Struct. Biol.* **10** 303–10

[42] SantaLucia J 1998 A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics *Proc. Natl Acad. Sci.* **95** 1460–5

[43] Doi M and Edwards S F 1986 *The Theory of Polymer Dynamics (International Series of Monographs on Physics* vol 73) (Oxford: Clarendon)

[44] Rouse P E Jr 1953 A theory of the linear viscoelastic properties of dilute solutions of coiling polymers *J. Chem. Phys.* **21** 1272–80