



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

## Hairpins in the conformations of a confined polymer

Downloaded from: <https://research.chalmers.se>, 2019-05-11 18:50 UTC

Citation for the original published paper (version of record):

Werner, E., Jain, A., Muralidhar, A. et al (2018)

Hairpins in the conformations of a confined polymer

Biomicrofluidics, 12(2)

<http://dx.doi.org/10.1063/1.5018787>

N.B. When citing this work, cite the original published paper.

## Hairpins in the conformations of a confined polymer

E. Werner,<sup>1</sup> A. Jain,<sup>2</sup> A. Muralidhar,<sup>2</sup> K. Frykholm,<sup>3</sup> T. St Clere Smithe,<sup>1</sup>  
 J. Fritzsche,<sup>4</sup> F. Westerlund,<sup>3</sup> K. D. Dorfman,<sup>2</sup> and B. Mehlig<sup>1</sup>

<sup>1</sup>*Department of Physics, University of Gothenburg, Origovägen 6B, 412 96 Göteborg, Sweden*

<sup>2</sup>*Department of Chemical Engineering and Materials Science, University of Minnesota–Twin Cities, 421 Washington Avenue SE, Minneapolis, Minnesota 55455, USA*

<sup>3</sup>*Department of Biology and Biological Engineering, Chalmers University of Technology, 412 96 Göteborg, Sweden*

<sup>4</sup>*Department of Physics, Chalmers University of Technology, 412 96 Göteborg, Sweden*

(Received 9 December 2017; accepted 21 February 2018; published online 9 March 2018)

If a semiflexible polymer confined to a narrow channel bends around by 180°, the polymer is said to exhibit a hairpin. The equilibrium extension statistics of the confined polymer are well understood when hairpins are vanishingly rare or when they are plentiful. Here, we analyze the extension statistics in the intermediate situation via experiments with DNA coated by the protein RecA, which enhances the stiffness of the DNA molecule by approximately one order of magnitude. We find that the extension distribution is highly non-Gaussian, in good agreement with Monte-Carlo simulations of confined discrete wormlike chains. We develop a simple model that qualitatively explains the form of the extension distribution. The model shows that the tail of the distribution at short extensions is determined by conformations with one hairpin. © 2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5018787>

### I. INTRODUCTION

A semiflexible polymer confined to a very narrow channel is almost perfectly extended and aligned with the channel axis. Small deviations from a perfect alignment yield an extension slightly smaller than the contour length of the polymer.<sup>1</sup> In this so-called Odijk regime, the equilibrium statistics of the polymer extension approach a Gaussian distribution in the limit of large contour length. The mean and variance of this distribution are known to high precision.<sup>2</sup> For a wormlike chain with persistence length  $\ell_p$  confined to a wider rectangular channel with the largest side length  $D_W$ , this regime is obtained when  $D_W \ll \ell_p$ .<sup>3–5</sup>

For wider channels ( $D_W \approx \ell_p$ ), the polymer can easily turn around, forming a C-shaped “hairpin” of length  $\ell$  as illustrated in Fig. 1. As a result, the conformational statistics depend on the interaction between hairpin segments. Commonly, the strength of interaction between segments in self-avoiding polymers is parameterized by an effective width,  $w$ . The limit  $w=0$  corresponds to an ideal polymer where self-avoidance does not matter. In this case, the typical length of a hairpin defines a length scale  $g$ , the global persistence length.<sup>3,6</sup> If the contour length  $L$  is much larger than  $g$ , there are multiple hairpins in any channel segment, and the distribution of the extension approaches that of a one-dimensional random walk of  $\approx L/g$  steps of length  $\approx g$ .<sup>3,4</sup> What happens when  $w$  is not zero? The importance of the parameter  $w$  is quantified by Odijk’s scaling parameter

$$\xi = \frac{gw}{D_H D_W^{2/3} \ell_p^{1/3}}, \quad (1)$$

where the channel height  $D_H$  is assumed to be smaller than the channel width,  $D_H \leq D_W$ . This parameter measures the expected number of overlapping points between the two strands of a

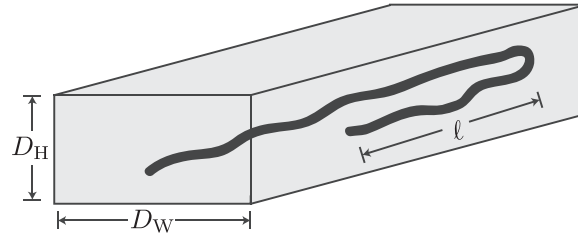


FIG. 1. Schematic of a confined wormlike chain confined in a channel of width  $D_W$  and height  $D_H < D_W$ , showing a configuration with a single hairpin of length  $l$ .

hairpin of length  $g$ .<sup>3-5</sup> The effect of self-avoidance depends on the magnitude of  $\zeta$ . If  $\zeta \gg 1$ , hairpins are rarely observed, as it would be very difficult for the two strands of a hairpin to avoid overlapping with each other. As a result, in this regime the extension statistics are in approximate agreement with the predictions of the Odijk regime.<sup>3,4</sup> If, on the other hand,  $\zeta \ll 1$ , then self-avoidance has a negligible effect on the likelihood of forming a hairpin. Yet for a long polymer the effect on macroscopic observables such as the extension can be very significant. For example, while the average extension of a long ideal polymer ( $w = 0$ ) grows as  $\langle X \rangle \sim \sqrt{L}$ , for a self-avoiding polymer the scaling of the extension is always linear in  $L$  in the limit  $L \rightarrow \infty$ , regardless of how small  $w > 0$  is.

The results summarized above apply provided that  $L \gg g$ . However, since the global persistence length  $g$  increases exponentially as a function of  $\ell_p/D$  for large values of  $\ell_p/D$ ,<sup>4,6,7</sup> the inequality  $L \gg g$  can be violated even for very long polymers that are strongly confined or have a large persistence length, so that  $\ell_p/D$  is large.

In this paper, we describe experiments with nanoconfined DNA that is covered with the protein RecA. The experimental setup is similar to the one in Ref. 8. The persistence length of the DNA-RecA filament is an order of magnitude larger than that of bare DNA. We measured the extension distributions of DNA-RecA filaments in nanochannels of different widths, ranging from 600 nm to 3  $\mu\text{m}$ , and found that the extension distribution is highly non-Gaussian. This conclusion is supported by direct numerical simulations of confined discrete wormlike chains. Comparison between experiments and simulations indicates that the persistence length of the DNA-RecA filaments is approximately  $\ell_p = 2 \mu\text{m}$  (this value is larger than most previous measurements reported in the literature<sup>8,9</sup>). The simulations also show that the global persistence length  $g$  is larger than the contour length  $L$  of the filaments. This leads us to conclude that the extension of the filament in the wider channels is dominated by hairpin configurations such as that shown in Fig. 1. Because  $g > L$ , none of the theories summarized above apply in our case. We therefore formulate a simple model that qualitatively explains the shape of the extension distributions. The model relies on an expansion of the equilibrium distribution into a series of terms corresponding to conformations with no or one hairpin. We find that the main peak is determined by conformations without hairpins, and that the tail of the distribution at short extensions is due to conformations with one hairpin.

## II. METHODS

### A. Experiments

Our experiments used nanochannels with rectangular cross sections, fabricated as described in Ref. 10. The channel width  $D_W$  ranged from 600 nm to 3000 nm, the channel height was  $D_H = 140 \text{ nm}$ . The channels were passivated using a lipid bilayer as previously described.<sup>11</sup> Double-stranded DNA, either  $\lambda$ -DNA (New England Biolabs) or T4-DNA (Wako Chemicals), was coated with fluorescently labeled RecA protein as described in Ref. 8. The resulting DNA-RecA filaments were moved between channels of different widths using pressure-driven flow. In this way, we could observe *the same* DNA-RecA filament in channels of different widths

$D_w$ . After moving the filament into a given channel, fluorescence imaging of the filament was performed while uniform pressure was maintained in the nanochannel. We used a Zeiss Axiovision microscope equipped with a 100 W mercury lamp, a Photometrics Evolve EMCCD camera, and a 100 $\times$  oil immersion TIRF objective (NA = 1.46) from Zeiss. One video of 400 frames was recorded for each channel width, for each filament. The time interval between frames was 0.11 s. The contour lengths of the T4-DNA-RecA filaments ranged from 7  $\mu\text{m}$  to 21  $\mu\text{m}$ , and those of the  $\lambda$ -DNA-RecA filaments ranged from 11  $\mu\text{m}$  to 23  $\mu\text{m}$ .

We measured the span of the DNA-RecA filaments as an estimate of the extension (as opposed to the end-to-end distance). To extract the span as a function of time, we fitted the brightness values for each frame to the curve  $\alpha + \beta\{\text{erf}[\gamma(x - x_0)] - \text{erf}[\varepsilon(x - x_1)]\}$ , where  $x$  is the location along the channel, and  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\varepsilon$ ,  $x_0$ , and  $x_1$  are the fitting parameters.<sup>12</sup>

Kymographs (Fig. 2) were produced by stacking intensity profiles into columns, so that each row of the column represents a single frame of the experimental recording. To produce the intensity profiles, the section of the frame containing the molecule was identified by locating the region with the maximal brightness. Next, the pixel intensity was averaged over the direction perpendicular to the channel, resulting in a row of pixel values which we interpret as the intensity profile along the channel. In practice, we found that we could obtain somewhat more robust results for the extension by modifying the algorithm for computing the kymographs. Instead of creating the intensity profile by averaging across the channel, we first smoothed each video frame by a median filter of radius two pixels, and defined the profile as the maximum pixel intensity in each column. Further, before fitting to the box curve described above, we smoothed the resulting kymograph by a moving time average with a window size of three frames and a moving spatial average with a window size of four pixels. Kymographic representations of all experimental videos are available in the [supplementary material](#).

## B. Direct numerical simulations

Our equilibrium simulations used a discretized wormlike chain model.<sup>13</sup> The discretization was achieved using a touching-bead model, with  $N + 1$  beads of diameter  $w$  connected by rigid bonds. The bending potential

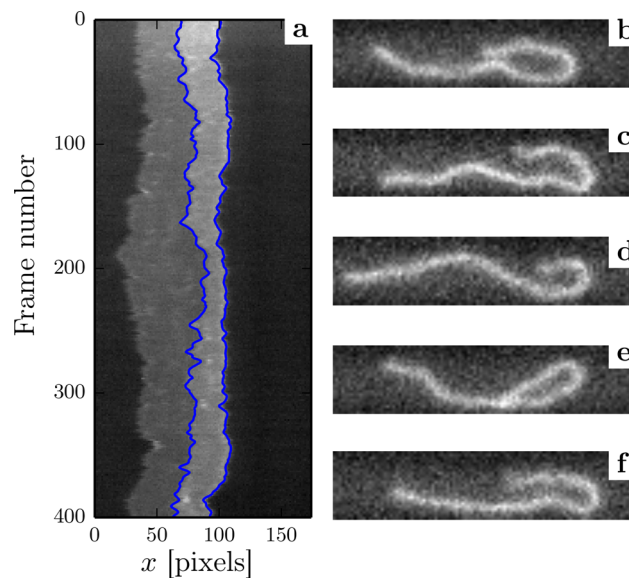


FIG. 2. (a) Kymograph,  $\lambda$ -DNA coated with RecA, channel width  $D_w = 2.4 \mu\text{m}$ . The blue lines indicate the edges of the hairpin, as identified by the algorithm described in Sec. II A. Panels (b)–(f) show cropped microscope images of the confined DNA-RecA filament (frame numbers 1, 100, 200, 300, and 350).

$$U(\theta_1, \dots, \theta_N) = k_B T \kappa \sum_{n=1}^N (1 - \cos \theta_n) \quad (2)$$

was used to model the stiffness of the chain. In Eq. (2), the index  $n$  ranges over bonds,  $\theta_n$  is the angle between bonds  $n$  and  $n+1$ , and  $\kappa$  is the bending constant. The persistence length in this model is defined as<sup>14,15</sup>

$$\ell_p = w \sum_{k=0}^{\infty} \langle \mathbf{t}_n \cdot \mathbf{t}_{n+k} \rangle, \quad (3)$$

where  $\mathbf{t}_n$  is the tangent vector to the chain at bond  $n$ , and the average is computed for an ideal and unconfined chain. The persistence length can be expressed in terms of  $\kappa$  and  $w$  as<sup>16</sup>

$$\frac{\ell_p}{w} = \frac{\kappa}{\kappa - \kappa \coth \kappa} \approx \kappa \quad (4)$$

for large values of  $\kappa$ . Details of the definition of  $\ell_p$  and the derivation of Eq. (4) are given in Ref. 15.

Excluded-volume interactions were incorporated into the model by imposing an infinite energy barrier for bead-bead overlap and bead-wall overlap. We simulated the model using the pruned-enriched Rosenbluth method (PERM),<sup>17,18</sup> following the approach described in our previous work.<sup>19,20</sup> PERM is a biased chain growth method that provides, *inter alia*, information on the extension statistics while avoiding the attrition problem in self-avoiding random walks. The PERM simulations were conducted to grow chains up to 1501 beads using the parameters given in Table I. Distributions of the mean span were obtained from  $2 \times 10^6$  tours, corresponding to  $1.7 \times 10^7$  configurations.

### III. RESULTS AND DISCUSSION

#### A. Extension distribution-experiments and simulations

Figure 3 compares the distribution of the extension  $X$  of the DNA-RecA filament along the channel direction measured in experiments against the results from simulations. It is customary to normalize the extension  $X$  by dividing by the contour length  $L$  of the polymer. However, since it is hard to determine the experimental contour length precisely enough, we normalized the polymer extension by the median polymer extension in the narrowest channel  $D_W = 600$  nm. We must therefore restrict our analysis to the subset of filaments that were imaged in the narrowest channel.

To compare simulations against experiments, it is necessary to determine the value of the persistence length  $\ell_p$  that best describes the DNA-RecA filament. To this end, we measured the median extension for each channel size, and compared the results against simulations for different values of  $\ell_p$ . The results are shown in Fig. 4. We observe the best fit between simulations and experiment when  $\ell_p$  is close to  $2 \mu\text{m}$ . We therefore used this value in Fig. 3. Figure 4 shows that the value  $\ell_p = 1.15 \mu\text{m}$  quoted in Ref. 8 is not consistent with our simulation data at large values of  $D_W$ . In Ref. 8, the persistence length was determined by fitting Odijk's

TABLE I. Parameters used in the computer simulations of the DNA-RecA filaments. The contour length used in the simulations is a representative value for the DNA-RecA filaments observed in experiments such as those in Fig. 1. The value of the persistence length was selected by comparison between the simulation data and experiment (see text). The value of the effective width represents an approximation for the thickening of naked DNA from coating with RecA.

$L$	$\ell_p$	$w$	$D_H$	$D_W$
15 $\mu\text{m}$	2 $\mu\text{m}$	10 nm	140 nm	0.6, 0.9, 1.3, 1.8, 2.4, 3.0 $\mu\text{m}$

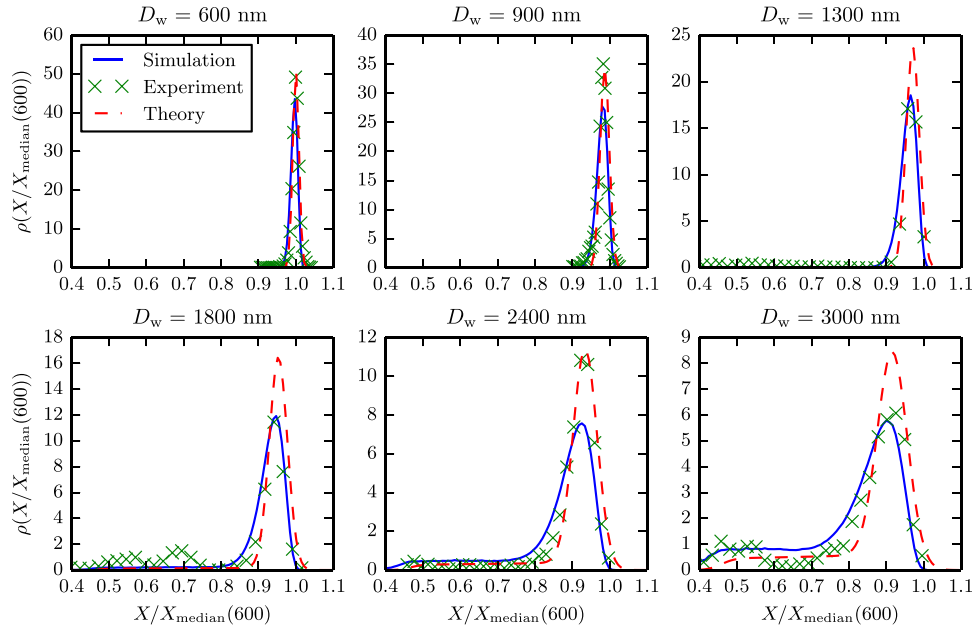


FIG. 3. Distributions of  $\rho[X/X_{\text{median}}(600)]$  of the span  $X$  relative to the median span  $X_{\text{median}}(600)$  in the 600 nm channel from computer simulations (solid lines) and experiments (symbols). Simulation parameters  $L = 15 \mu\text{m}$  and  $\ell_p = 2 \mu\text{m}$ . Since we cannot measure the contour length to sufficient accuracy in experiments, we instead normalize the span by dividing by the median span in the narrowest channel ( $D_w = 600 \text{ nm}$ ). The simulations yield  $X_{\text{median}}(600) \approx 14.18 \mu\text{m}$ . The experimental data consist of the subset of experiments for which the same filament was measured both in the channel under consideration and in the narrowest channel ( $D_w = 600 \text{ nm}$ ), and further where  $X_{\text{median}}(600)$  agrees to within 20% with the value found in the simulation. Further, we removed two filaments that exhibited a hairpin in the narrowest channel, in which case, the median span is not a good proxy for the contour length. Also shown are the model predictions (Sec. III C), dashed lines, for  $L = 15 \mu\text{m}$  and  $\ell_p = 2 \mu\text{m}$ . For the narrowest channel, we assume  $g = \infty$  in the theory, i.e., no hairpins.

expression Eq. (7) to the experimental data, using  $L$  and  $\ell_p$  as fitting parameters. This procedure is quite insensitive to  $\ell_p$  and does not allow to rule out  $\ell_p = 2 \mu\text{m}$  (Appendix C).

Our estimate of the persistence length comes with a large uncertainty, because the minimum in Fig. 4(b) is quite shallow. However, our estimate appears to be inconsistent with the value  $\ell_p \approx 1 \mu\text{m}$  found in Ref. 9 obtained using an entirely different technique. At present, we do not know the reason for this discrepancy. One previous study,<sup>30</sup> on the other hand, found a persistence length of  $2.1 \pm 0.1 \mu\text{m}$  for single-stranded DNA covered with RecA, albeit under experimental conditions that differ from ours.

Figure 3 shows that there is good qualitative agreement between the simulations for  $\ell_p = 2 \mu\text{m}$  and the experiments. This indicates that the discretized wormlike chain model is a good approximation for the conformations of nanoconfined DNA-RecA filaments. In addition to

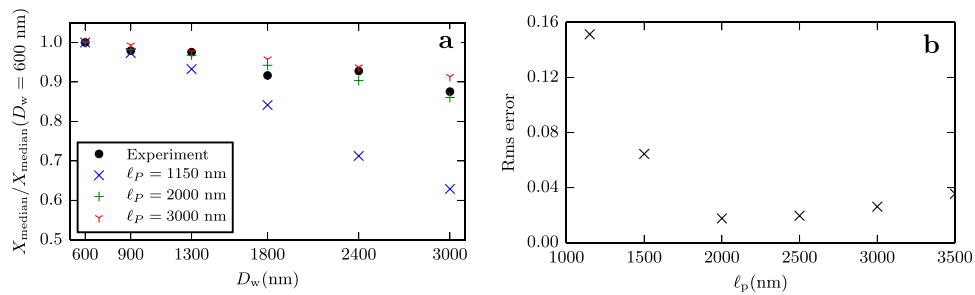


FIG. 4. Comparison of simulations against experiments, to establish which value of the persistence length  $\ell_p$  fits best. (a) Comparison of the median DNA-RecA filament span from experiments against simulations for different values of  $\ell_p$ . The median span is normalized by the value measured in the narrowest channel (as in Fig. 3). (b) Root-mean squared difference between the simulated and experimental values shown in panel (a), averaged over all channels except the smallest ones (where the difference is zero by construction).

the main peak, i.e., the peak located near  $X/X_{\text{median}}(600) \approx 0.9$ , the distributions in the wider channels exhibit pronounced tails at short spans. Using the computer simulations, we verified that the small- $X$  tails correspond to conformations with at least one hairpin (as depicted in Fig. 1).

Figure 3 also shows quantitative differences between simulations and experiments. These differences may be due to the fact that the DNA-RecA filament conformations may have been insufficiently sampled in the experiments. The experimentally observed conformation dynamics (discussed below) indicate that the main peak in the extension distribution is likely to be well sampled, but the tails at short filament extensions (spans) may not be well sampled. Another possible source of error is that the experimental method may have caused a bias in the initial conformations, because the filaments were moved between different regions of the channels between the measurements. However, we expect this bias to be relatively small in our case when compared to injection from a reservoir into a nanoslit,<sup>22</sup> since the change in confinement between two values of  $D_W$  is relatively small. We have insufficient data to evaluate a possible bias with statistical analysis, but we checked at least that in none of the cases analyzed in this section did a filament ending in a left (right) hairpin state in one channel begin with a left (right) hairpin in the next channel. This indicates that the initial conditions were in equilibrium.

Since only the protein RecA fluoresces, parts of DNA that are not coated by RecA are invisible in our experiments. It is therefore possible that such invisible DNA is attached at one or both ends of the visible filament. This could have a small effect on the equilibrium statistics, as excluded volume interactions between the bare and coated DNA could reduce the probability of observing hairpins. More importantly, hydrodynamic friction between the bare DNA and the surrounding solution might slow down the dynamics.

## B. Global persistence length

According to Eq. (1), it is the parameter  $\zeta$  (the scaled global persistence length) that determines how frequently hairpins can form. Therefore, we computed this parameter in our simulations. Odijk<sup>3,6</sup> proposed a mechanical theory for computing the global persistence length  $g$ . However, recent simulations of confined wormlike chains suggest that Odijk's theory overestimates the global persistence length.<sup>4,21,23</sup> While the functional form predicted by Odijk's theory is reasonable for square and circular channels,<sup>4,23</sup> it is not for rectangular channels.<sup>21</sup> Therefore, we decided to determine the values of  $g$  for our particular channels via simulation, following the approach for rectangular channels described in Ref. 21.

Table II presents the values of  $g$  obtained in this way. For channels with  $D_W = 900$  nm and wider, we could reliably extract a value of the global persistence length. For the smallest channel, we could not compute a value of  $g$  from the simulations. This suggests that  $g$  is larger than the maximal contour length that we could simulate. In principle, this limitation can be surmounted by increasing the contour length of the simulated ideal chains. In practice, we expect that  $g$  increases exponentially with decreasing  $D_W/\ell_P$ .<sup>6</sup> As a result, the requisite contour length quickly becomes infeasible to simulate, even using a biased-growth method like PERM. In what follows, we simply set  $g = \infty$  for the smaller channels. At any rate, Table II shows that  $g \gg L$  for all but the widest channel.

Table II also gives the corresponding values of the parameter  $\zeta$  in Eq. (1). We see that the values of  $\zeta$  are of order unity for the wider channels. This is consistent with the shapes of the extension distributions that exhibit pronounced tails at short spans due to hairpin-conformations for these channels. In the following, we describe a simple model that qualitatively explains the shapes of the distributions in this limit.

## C. Model calculation

When the parameter  $\zeta$  is of order unity, the conformations may sometimes exhibit hairpins, and sometimes not. Since the mechanisms determining the polymer span  $X$  are quite different if hairpins are present or absent, it is necessary to consider these possible configurations separately, before combining them into a single distribution  $\rho(X)$

TABLE II. Summary of simulation data and corresponding theory parameters for  $L = 15 \mu\text{m}$ . The values of  $g$  are obtained from simulations of ideal chains for the parameters in Table I, as described in Ref. 21. The entries... indicate channel sizes where  $g$  could not be measured. The values of the scaling parameter  $\xi$  are computed from Eq. (1). The values of the self avoidance length scale  $\ell_{\text{SA}}$  are obtained from Eq. (10) using the value of  $\alpha$  computed from Fig. 6.

$D_{\text{W}} (\mu\text{m})$	0.6	0.9	1.3	1.8	2.4	3.0
$g (\mu\text{m})$	...	13000	350	90	37	22
$\xi$	...	780	17	3.5	1.2	0.61
$\ell_{\text{SA}} (\mu\text{m})$	...	9.4	12	15	18	21

$$\rho(X) = \sum_{j=0}^{\infty} \rho_j(X) P_j. \quad (5)$$

Here,  $\rho_j(X) \equiv \rho(X|N_{\text{h}} = j)$  is the distribution of  $X$  conditional on that the number  $N_{\text{h}}$  of hairpins equals  $j$ , and  $P_j \equiv \text{prob}(N_{\text{h}} = j)$  is the probability of observing  $j$  hairpins. Since this study is concerned with the situation where hairpins are rare, we restrict the following discussion to the two cases of either zero or one hairpin.

When  $L \gg \lambda \approx D_{\text{W}}^{2/3} \ell_{\text{P}}^{1/3}$ , the distribution  $\rho_0(X)$  for a polymer without hairpins is Gaussian

$$\rho_0(x) = (2\pi\sigma_{\text{Od}}^2)^{-1/2} \exp\left[-\frac{(X - \mu_{\text{Od}})^2}{2\sigma_{\text{Od}}^2}\right] \quad (6)$$

with mean and variance which we approximate by the conformation statistics of the Odijk regime<sup>2</sup>

$$\mu_{\text{Od}} = L \left(1 - 0.091 \frac{D_{\text{H}}^{2/3} + D_{\text{W}}^{2/3}}{\ell_{\text{P}}^{2/3}}\right), \quad (7)$$

$$\sigma_{\text{Od}}^2 = 0.0048L \left(\frac{D_{\text{H}}^2 + D_{\text{W}}^2}{\ell_{\text{P}}}\right). \quad (8)$$

In [Appendix A](#), we derive an expression for  $\rho_1(X)$

$$\rho_1(X) = \mathcal{N} \left[ \text{erf}\left(\frac{2X - \mu_{\text{Od}}}{2\sigma_{\text{Od}}}\right) - \text{erf}\left(\frac{X - \mu_{\text{Od}}}{\sqrt{2}\sigma_{\text{Od}}}\right) \right] \exp\left(\frac{X - L/2}{\ell_{\text{SA}}}\right). \quad (9)$$

The calculation in [Appendix A](#) assumes that hairpins are equally likely to form anywhere on the ideal polymer, that each of the ideal hairpin strands fluctuates according to Eqs. (6)–(8). Self avoidance is taken into account by penalizing hairpins with long hairpin lengths  $\ell$ . We assume that the probability of collisions between hairpin strands is proportional to the length of the shorter strand. The prefactor  $\mathcal{N}$  in Eq. (9) is determined by the condition that  $\rho_1(X)$  is normalized to unity on  $[0, L]$ . The length scale  $\ell_{\text{SA}}$  quantifies the hairpin length at which self-avoidance becomes important, penalizing configurations with large  $\ell$ . The probability of overlaps between two hairpin strands can be estimated by a mean-field argument,<sup>3,21</sup> yielding

$$P_{\text{no overlap}}(\ell) = \exp(-\alpha\xi\ell/g) \equiv \exp(-\ell/\ell_{\text{SA}}), \quad (10)$$

where  $\alpha$  is a prefactor of order unity. We estimated the parameter  $\ell_{\text{SA}}$  from our computer simulations ([Appendix B](#)). The results are shown in Table II. This parameter determines the ratio of the coefficients  $P_1$  and  $P_2$  in Eq. (5). We show in [Appendix A](#) that



$$\frac{P_1}{P_0} = \frac{\ell_{SA}}{g} \left[ 1 - \exp\left(-\frac{L}{2\ell_{SA}}\right) \right]. \quad (11)$$

As expected, the limit  $\ell_{SA} \gg L$  reproduces the ideal result  $(P_1/P_0)_{ideal} = L/(2g)$  (Appendix A). The opposite limit yields a ratio that is independent of  $L$ .

In summary, Eqs. (5), (6), (9), and (11) yield the desired approximation for the distribution  $\rho(X)$  of the span  $X$ . The result is shown in Fig. 3 as dashed lines. We observe good qualitative agreement with both experiments and simulations, without any fitting parameters. For wider channels, the theory underestimates the width of the peak at large span, as well as its skewness. This is a result of the assumption that the Odijk theory [Eqs. (7) and (8)] describes the statistics of each hairpin, but this assumption starts to fail when  $D_W \approx \ell_P$ . We also note that the simple theory neglects the effect of the curved part of the hairpin on the extension (span) of the filament in the channel. This results in an error of the order of the channel width, so that the theory may overestimate the span by about 20%. Comparing the simulation and theory in Fig. 3, we see that the left tail does extend to smaller values of  $X$  and is heavier in the simulations, compared with the theory.

#### D. Diffusion model for hairpin size

We mentioned above that a possible explanation for the difference between experiments and simulations in Fig. 3 is that the experimental conformations have been insufficiently sampled. To investigate whether this is the case, we analyzed the experimental time series of the filament spans. A representative result is shown in Fig. 5. It demonstrates that the hairpin length first shrinks and then grows, demonstrating that the dynamics are stochastic, unlike the unfolding process described in Refs. 24 and 25. To qualitatively explain this observation, and to estimate the unfolding time scale, we formulated a diffusion model for the hairpin dynamics.

A hairpin configuration can be represented by two polymer strands, connected at one end. If one of the strands is much longer than the other, it is practically immobile. The shorter strand diffuses like a polymer in the Odijk regime, with diffusion constant  $k_B T / \zeta(\ell)$ .<sup>16</sup> Here,  $\zeta$  is a friction coefficient. It is defined in the following way: imagine that the polymer is dragged slowly along the channel with a velocity  $v$ . The opposing force acting on the polymer is then proportional to  $v$ , with proportionality constant  $\zeta$ . Since a change in the span requires that a large fraction of the polymer segments move, we can assume that  $v \approx \dot{X}$ . It follows that a change in the span leads to an average friction force  $F \approx -\zeta \dot{X}$ . The length of the hairpin changes as it diffuses. In addition to the diffusive motion, the strand experiences an unfolding force  $f$ , due to the collisions between the two strands. The simultaneous action of diffusion and the constant unfolding force can be described by the generalized diffusion equation

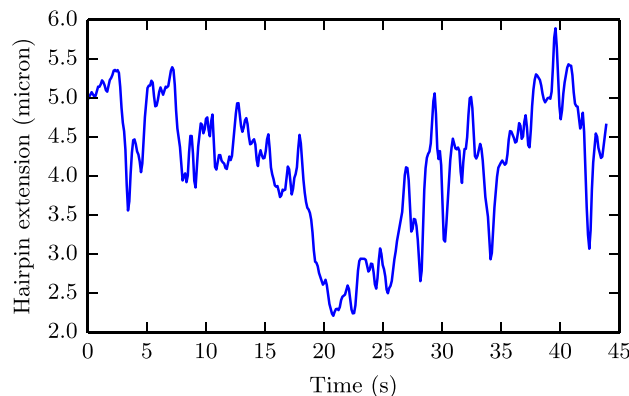


FIG. 5. Span of the hairpin shown in Fig. 2 as a function of time. Note that the span decreases significantly before increasing again, showing that the hairpin extension dynamics are non-deterministic.

$$\frac{\partial \rho(\ell)}{\partial t} = -\frac{\partial}{\partial \ell} \left[ \frac{f}{2\zeta(\ell)} \rho(\ell) \right] + k_B T \frac{\partial}{\partial \ell} \left[ \frac{1}{2\zeta(\ell)} \frac{\partial \rho(\ell)}{\partial \ell} \right]. \quad (12)$$

Note that Eq. (12) has the equilibrium distribution  $\rho_{\text{eq}}(\ell) \propto \exp[\int^\ell f/(k_B T) d\ell']$ , in agreement with the Boltzmann distribution if the force is given by the derivative of an energy,  $f = -\partial E/\partial \ell$ . In the deterministic limit, where the rightmost term is negligible, Eq. (12) is consistent with the deterministic description of an unfolding process given in Refs. 24 and 25, as well as with the description of ejection of a polymer out of a nanochannel.<sup>26</sup>

The force  $f$  in Eq. (12) is due to the self-avoiding interaction between the two strands in the folded configuration that tends to unfold the hairpin. This can be modeled by an entropic force given as the slope of the free energy.<sup>27</sup> Apart from terms that are independent of the macroscopic configuration, the free energy is given by the sum of the elastic energy stored in the hairpin bend and an entropic free-energy contribution due to collisions between the strands,  $-k_B T \log P_{\text{no overlap}}$ . The first term does not depend on the hairpin length, so Eq. (10) shows that the entropic force is given by

$$f = k_B T \frac{\partial \log P_{\text{no overlap}}}{\partial \ell} = \frac{-k_B T}{\ell_{\text{SA}}} = -\alpha k_B T \frac{w}{D_H \ell_P^{1/3} D_W^{2/3}}. \quad (13)$$

While it is possible to provide a more accurate estimate of the entropic force, for example by using the methods of Ref. 28, we cannot directly apply the results of the latter reference to our analysis. The confinement free energy of a high aspect ratio rectangular channel, which arises in part due to excluded volume, is different than in a circular tube.<sup>3,21</sup>

Equation (12) allows one to define two time scales for a hairpin of length  $\ell$ , one diffusive time scale  $\tau_{\text{diff}} = \zeta(\ell)\ell^2/(kT)$  and one deterministic time scale  $\tau_{\text{drift}} = \zeta(\ell)\ell/f$ . In drift-diffusion problems, the relative effects of drift to diffusion are captured by the ratio of these time-scales, namely the Péclet number  $\text{Pe} = f\ell/kT$ . For a hairpin of length  $3 \mu\text{m}$  in a  $1300 \text{ nm}$  channel, we find  $\text{Pe} \approx 0.2$ , suggesting that drift is a small effect. The prediction that the unfolding dynamics of a DNA-RecA hairpin is diffusive contrasts with the predominantly deterministic dynamics seen in experiments on bare DNA.<sup>24,29</sup> This prediction is validated by the experimental data shown in Fig. 5. Since only a small fraction of our experiments show hairpins, we do not have enough data for a quantitative test of Eq. (12), or to estimate  $\tau_{\text{diff}}$ , but the simulation results of Ref. 16 yield an estimate  $\tau_{\text{diff}} \approx 10 \text{ s}$  for a hairpin of length  $3 \mu\text{m}$ . This is of the same order as the length of an experimental video ( $\approx 40 \text{ s}$ ). As discussed in Sec. III A, this could indicate that the small- $X$  tails of the experimental distributions in Fig. 3 are insufficiently sampled.

#### IV. CONCLUSIONS

In this article, we investigated the extension distribution of nanoconfined RecA-coated DNA. We compared the experimental distribution of the filament span with the results of equilibrium Monte Carlo simulations of strongly confined wormlike chains that are shorter than the global persistence length. In the wider channels, the experimental distributions are strongly non-Gaussian, in good qualitative agreement with our results from the Monte Carlo simulations. We found the best fit between simulations and experiments if we assumed that the DNA-RecA filaments have a persistence length of  $\ell_P \approx 2 \mu\text{m}$ . This value is larger than  $\ell_P \approx 1 \mu\text{m}$  found in Ref. 9, obtained using an entirely different technique. The reason for this discrepancy is not known.

Apart from the central peak, the most distinct feature of the non-Gaussian distributions is a heavy tail at short spans. We derived a simple model for the extension distribution that explains that this tail is due to conformations that contain at least one hairpin. This also furnishes a possible explanation for the quantitative differences that we observed between the experimental and simulation data. It could be partly due to the fact that the small- $X$  tails of the experimental distributions are insufficiently sampled, as mentioned above. In addition to the time scale for unfolding that discussed in Sec. III D, our ability to sample the tail is also affected by the time scale for folding, since folding involves a high energy penalty. We expect that the sampling of the tails of

the distribution is the largest source of discrepancy between the experiment and simulations in our system. At any rate, our results show that the equilibrium theory works quite well even though the actual experiment may not sample sufficiently. A recent analysis of unfolding of DNA in a nanochannel<sup>31</sup> allows a similar conclusion, although for quite different parameters. Despite the fact that the dynamics is deterministic on the time scales of that experiment, Odijk's equilibrium theory works very well. We formulated a stochastic model for the stochastic dynamics of a hairpin extension to estimate the time scale associated with this dynamics, indicating the time scale for hairpin length fluctuations is indeed similar to the length of an experimental video.

More generally, the hairpin dynamics of the DNA-RecA filaments are very different than the corresponding dynamics of bare DNA. While the latter are almost deterministic,<sup>24,29</sup> the former are not. We therefore proposed a stochastic model for how the size of a hairpin fluctuates over time. This model includes both the effect of self-avoidance (which tends to reduce the length of the hairpin), and the diffusion caused by thermal motion. Since self-avoidance is relatively weak for our system, the dynamics is dominated by the diffusive term. Our experimental data are in qualitative agreement with the stochastic model, but further experiments and modeling work are necessary to establish a quantitative description of the hairpin dynamics.

## SUPPLEMENTARY MATERIAL

See [supplementary material](#) for a summary of the experimental data in the form of kymographs.

## ACKNOWLEDGMENTS

The labeled RecA protein was a kind gift from Edwige B. Garcin and Mauro Modesti. E.W. and B.M. wish to thank Stefano Bo and Tobias Ambjörnsson for discussions concerning Eq. (12). Financial support from Vetenskapsrådet [Grant Nos. 2013-3992, 2017-3865 (BM) and 2011-4324 (FW)] and from the National Institutes of Health [R01-HG006851(KDD)] is gratefully acknowledged. The computational work was carried out in part using computing resources at the University of Minnesota Supercomputing Institute, and in part at the Chalmers Centre for Computational Science and Engineering (C3SE) provided by the Swedish National Infrastructure for Computing (SNIC).

## APPENDIX A: DERIVATION OF $\rho_1(X)$

The approximation (9) for  $\rho_1(X)$  is derived in several steps. First, let  $\ell$  denote the length of the shortest hairpin strand, as illustrated in Fig. 1. In this configuration, the span of the filament in the channel is, to a good approximation, given by the extension of the longest hairpin strand, which obeys Odijk statistics with a contour length of  $L - \ell$ . For ideal polymers, i.e., without interactions between the two strands of the hairpin, the hairpin length is drawn uniformly between 0 and  $L/2$ . Thus

$$\rho_1^{\text{id}}(X) \approx \frac{2}{L} \int_0^{L/2} d\ell \rho_0(X; L - \ell) \approx \frac{1}{\mu_{\text{Od}}} \left[ \text{erf} \left( \frac{2X - \mu_{\text{Od}}}{2\sigma_{\text{Od}}} \right) - \text{erf} \left( \frac{X - \mu_{\text{Od}}}{\sqrt{2}\sigma_{\text{Od}}} \right) \right]. \quad (\text{A1})$$

Here,  $\rho_0(X; L)$  is the distribution of the span of a polymer segment of length  $L$  that is free of hairpins [Eq. (6)].

Second, we must modify Eq. (A1) to account for the excluded volume. If the polymer is self-avoiding, then configurations with large  $\ell$  are penalized. The probability of overlaps between these two strands can be estimated by a mean-field argument,<sup>3,21</sup> yielding

$$P_{\text{no overlap}}(\ell) = \exp(-\alpha \xi \ell / g) \equiv \exp(-\ell / \ell_{\text{SA}}), \quad (\text{A2})$$

where  $\alpha$  is a prefactor of order unity. This is Eq. (10) in the main text. The second equality defines the self-avoidance length  $\ell_{\text{SA}}$ . Equation (A2) allows us to modify Eq. (A1) to take into account

self-avoidance: the integrand in Eq. (A1) must include an extra factor proportional to  $[1 - P_{\text{no overlap}}(\ell)]$ . Performing the integration then yields

$$\rho_1(X) = \mathcal{N} \left[ \operatorname{erf} \left( \frac{2X - \mu_{\text{Od}}}{2\sigma_{\text{Od}}} \right) - \operatorname{erf} \left( \frac{X - \mu_{\text{Od}}}{\sqrt{2}\sigma_{\text{Od}}} \right) \right] \exp \left( \frac{X - L/2}{\ell_{\text{SA}}} \right). \quad (\text{A3})$$

This is Eq. (9) in the main text. The prefactor is given by the condition that  $\rho_1(X)$  is normalized on  $[0, L]$ .

Third, to determine the first two terms in Eq. (5) in the main text we also require  $P_0$  and  $P_1$ . For ideal polymers, the formation of a hairpin bend at a given location is independent of how many other hairpins that have already formed. It follows that the number of hairpins is Poisson distributed, with a rate constant  $\mu \propto L/g$ . A precise calculation yields  $\mu = L/(2g)$ . It follows that  $P_1^{\text{id}}/P_0^{\text{id}} = L/(2g)$ . We also note that  $P_2^{\text{id}}/P_0^{\text{id}} = L^2/(8g^2)$ , so the contribution of two hairpins to  $\rho(X)$  is small when  $L \ll g$ . Now consider a self-avoiding polymer. It is less likely than an ideal polymer to exhibit hairpins, as only a fraction of all ideal configurations with a hairpin are free of overlaps. We denote this fraction by

$$\langle P_{\text{no overlap}} \rangle \equiv \frac{2}{L} \int_0^{L/2} d\ell P_{\text{no overlap}}(\ell) = \frac{2\ell_{\text{SA}}}{L} \left[ 1 - \exp \left( -\frac{L}{2\ell_{\text{SA}}} \right) \right]. \quad (\text{A4})$$

This implies that

$$\frac{P_1}{P_0} = \frac{L}{2g} \langle P_{\text{no overlap}} \rangle = \frac{\ell_{\text{SA}}}{g} \left[ 1 - \exp \left( -\frac{L}{2\ell_{\text{SA}}} \right) \right]. \quad (\text{A5})$$

This is Eq. (11) in the main text.

## APPENDIX B: ESTIMATION OF $\ell_{\text{SA}}$

We determined the self-avoidance length scale  $\ell_{\text{SA}}$  from the configurations produced by the PERM simulations using a polymer with a persistence length of  $\ell_p = 1150$  nm confined to a channel of width  $D_w = 600$  nm. To compute these data, we used  $3 \times 10^5$  tours, corresponding to  $2.7 \times 10^6$  configurations. For each tour that reached the maximum contour length, we output the configuration of the chain and its corresponding statistical weights for post-processing. The configuration of a polymer chain allows us to calculate the hairpin-contour lengths. To do so, we considered a moving window of 10 beads and computed the span and end-to-end distance for the subchain in a given window. If these two sizes were found to be equal, we concluded that there was no hairpin within the window; otherwise, the window was flagged as belonging to a hairpin. Typically, a given hairpin involves multiple contiguous windows, which were grouped into a single hairpin event by a clustering algorithm. Within each cluster, we then identified the bead with the minimum value of position  $x_i$  along the channel axis and assigned the hairpin location to that bead. If the algorithm identified two or more hairpin strands, we used the shortest strand length as  $\ell$ . A naive application of this algorithm produces many small hairpins that would be undetectable by the optical resolution in the experiments. To provide a result that is experimentally relevant, we only included values of  $\ell$  if the chain within that cluster spans wider than 500 nm.

We extracted the self-avoidance length scale  $\ell_{\text{SA}}$  from the distribution of  $\ell$  using Eq. (10). Figure 6 shows the probability distribution of hairpin lengths for a self-avoiding polymer in a channel. The value of  $\ell_{\text{SA}}$  is obtained by fitting the distribution to Eq. (10); as seen in Fig. 6. Since the hairpin length of an ideal polymer is uniformly distributed between 0 and  $L/2$ , it follows that  $P(\ell)$  equals  $(2/L)P_{\text{no overlap}}(\ell)$ . The parameter  $\alpha$  must be independent of the channel size, so we used the result from Fig. 6,  $\alpha = 1.7$ , to compute the values of  $\ell_{\text{SA}}$  in Table II.

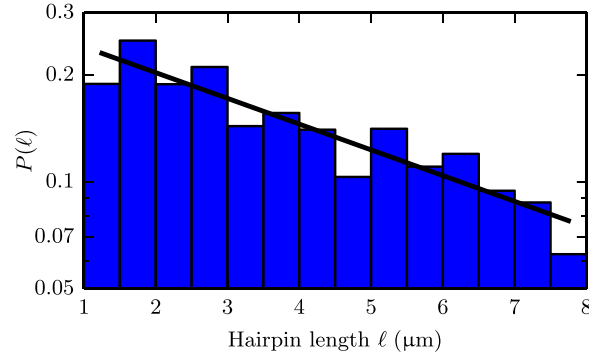


FIG. 6. Semi-logarithmic plot of the probability distribution of the hairpin length for a polymer with persistence length  $\ell_p = 1150$  nm in a channel of width  $D_W = 600$  nm. The solid line indicates an exponential fit to Eq. (A2), leading to  $\ell_{SA} = 6.0$   $\mu\text{m}$ . Together with the numerical values for  $g$  and  $\xi$  (see Table II), this gives us a dimensionless prefactor of  $\alpha = 1.7$  after comparing against Eq. (10).

### APPENDIX C: ESTIMATION OF $\ell_p$

Figure 7 shows the same experimental data for the median of the spans of all DNA-RecA filaments as Fig. 4 in the main text. In addition, also the experimental data from Fig. 2A in Ref. 8 are shown, corresponding to measurements on a single DNA-RecA filament. Figure 7 indicates that the data are consistent with each other. Also shown is Odijk's theory, Eq. (7)

$$\langle X \rangle = L \left\{ 1 - 0.091 \left[ \left( \frac{D_H}{\ell_p} \right)^{2/3} + \left( \frac{D_W}{\ell_p} \right)^{2/3} \right] \right\} \quad (\text{C1})$$

using  $L = 13.78$   $\mu\text{m}$  and  $\ell_p = 2$   $\mu\text{m}$ . In Ref. 8, the same data were fitted with  $L \approx 13.6$   $\mu\text{m}$ ,  $\ell_p \approx 1.39$   $\mu\text{m}$ , and using a numerical prefactor 0.085 that applies to Gaussian confinement rather than the value 0.091 appropriate for rectangular channels. This yields a better fit at larger values of  $D_W$ , but one must bear in mind that Eq. (C1) is only asymptotically valid at small  $D_W/\ell_p$ . We can conclude that the fit of Eq. (C1) to the experimental data is relatively insensitive to  $\ell_p$ , in particular, if  $L$  must be fitted at the same time. A further source of uncertainty is that the variance between estimates of  $\ell_p$  obtained in Ref. 8 from different filaments is large, and that the average value  $\ell_p = 1.15$   $\mu\text{m}$  obtained in Ref. 8 over many filaments does not include filaments with persistence lengths larger than 2  $\mu\text{m}$ . We also note that outliers in the data were discarded, possibly

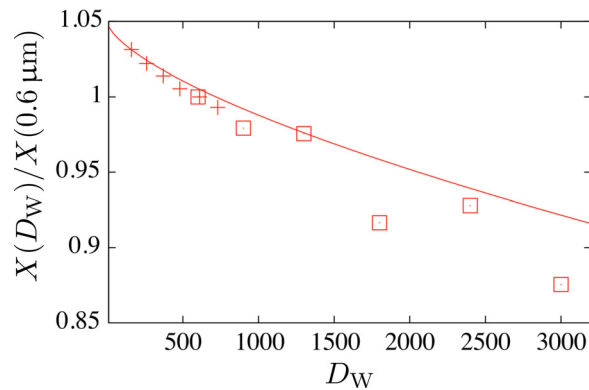


FIG. 7. Median span taken over all DNA-RecA filaments from our experiments ( $\square$ ), normalized by  $X_{\text{median}}(600\text{ nm})$ . Experimental data (+) read off from Fig. 2A in Ref. 8 corresponding to measurements on a single DNA-RecA filament, normalized by the span at  $D_W = 610$  nm. Also shown is Eq. (C1) for  $\ell_p = 2$   $\mu\text{m}$  and  $L = 13.78$   $\mu\text{m}$  (solid line).

leading to a bias to smaller persistence lengths (the analysis is described in detail in Ref. 8 and its supplementary material). For these reasons, we have chosen to determine  $\ell_p$  by comparison with simulation result in the range of (larger)  $D_w$  directly relevant for our study. This results in  $\ell_p = 2 \mu\text{m}$ , but also with a large uncertainty because the minimum in Fig. 4(b) is so shallow.

- <sup>1</sup>T. Odijk, "The statistics and dynamics of confined or entangled stiff polymers," *Macromolecules* **16**, 1340–1344 (1983).
- <sup>2</sup>T. W. Burkhardt, Y. Yang, and G. Gompper, "Fluctuations of a long, semiflexible polymer in a narrow channel," *Phys. Rev. E* **82**, 041801 (2010).
- <sup>3</sup>T. Odijk, "Scaling theory of DNA confined in nanochannels and nanoslits," *Phys. Rev. E* **77**, 060901 (2008).
- <sup>4</sup>A. Muralidhar, D. R. Tree, and K. D. Dorfman, "Backfolding of wormlike chains confined in nanochannels," *Macromolecules* **47**, 8446–8458 (2014).
- <sup>5</sup>E. Werner and B. Mehlig, "Scaling regimes of a semi-flexible polymer in a rectangular channel," *Phys. Rev. E* **91**, 050601(R) (2015).
- <sup>6</sup>T. Odijk, "DNA confined in nanochannels: Hairpin tightening by entropic depletion," *J. Chem. Phys.* **125**, 204904 (2006).
- <sup>7</sup>J. Z. Y. Chen, "Conformational properties of a back-folding wormlike chain confined in a cylindrical tube," *Phys. Rev. Lett.* **118**, 247802 (2017).
- <sup>8</sup>K. Frykholm, M. Alizadehheidari, J. Fritzsche, J. Wigenius, M. Modesti, F. Persson, and F. Westerlund, "Probing physical properties of a DNA-protein complex using nanofluidic channels," *Small* **10**, 884–887 (2014).
- <sup>9</sup>M. Hegner, S. B. Smith, and C. Bustamante, "Polymerization and mechanical properties of single RecA–DNA filaments," *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10109–10114 (1999).
- <sup>10</sup>F. Persson and J. O. Tegenfeldt, "Dna in nanochannels—directly visualizing genomic information," *Chem. Soc. Rev.* **39**, 985–999 (2010).
- <sup>11</sup>F. Persson, J. Fritzsche, K. U. Mir, M. Modesti, F. Westerlund, and J. O. Tegenfeldt, "Lipid-based passivation in nanofluidics," *Nano Lett.* **12**, 2260–2265 (2012).
- <sup>12</sup>J. O. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W. W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, and P. Silberzan, "The dynamics of genomic-length DNA molecules in 100-nm channels," *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10979–10983 (2004).
- <sup>13</sup>J. Wang and H. Gao, "A generalized bead-rod model for Brownian dynamics simulations of wormlike chains under strong confinement," *J. Chem. Phys.* **123**, 084906 (2005).
- <sup>14</sup>P. C. Heimenz and T. P. Lodge, *Polymer Chemistry* (CRC Press, Boca Raton, FL, 2007).
- <sup>15</sup>K. D. Dorfman, "The statistical segment length of DNA: Opportunities for biomechanical modeling in polymer physics and next-generation genomics," *J. Biomech. Eng.* **140**, 020801 (2018).
- <sup>16</sup>A. Muralidhar and K. D. Dorfman, "Kirkwood diffusivity of long semiflexible chains in nanochannel confinement," *Macromolecules* **48**, 2829–2839 (2015).
- <sup>17</sup>P. Grassberger, "Pruned-enriched Rosenbluth method: Simulations of  $\theta$  polymers of chain length up to 1 000 000," *Phys. Rev. E* **56**, 3682–3693 (1997).
- <sup>18</sup>T. Prellberg and J. Krawczyk, "Flat histogram version of the pruned and enriched Rosenbluth method," *Phys. Rev. Lett.* **92**, 120602 (2004).
- <sup>19</sup>D. R. Tree, A. Muralidhar, P. S. Doyle, and K. D. Dorfman, "Is DNA a good model polymer?," *Macromolecules* **46**, 8369–8382 (2013).
- <sup>20</sup>D. R. Tree, Y. Wang, and K. D. Dorfman, "Extension of DNA in a nanochannel as a rod-to-coil transition," *Phys. Rev. Lett.* **110**, 208103 (2013).
- <sup>21</sup>A. Muralidhar, M. Quevillon, and K. Dorfman, "The backfolded Odijk regime for wormlike chains confined in rectangular nanochannels," *Polymers* **8**, 79 (2016).
- <sup>22</sup>C. T. A. Wong and M. Murugappan, "Scaling theory of polymer translocation into confined regions," *Biophys. J.* **95**, 3619–3627 (2008).
- <sup>23</sup>A. Muralidhar and K. D. Dorfman, "Backfolding of DNA confined in nanotubes: Flory theory versus the two-state cooperativity model," *Macromolecules* **49**, 1120–1126 (2016).
- <sup>24</sup>S. L. Levy, J. T. Mannion, J. Cheng, C. H. Reccius, and H. G. Craighead, "Entropic unfolding of DNA molecules in nanofluidic channels," *Nano Lett.* **8**, 3839–3844 (2008).
- <sup>25</sup>G. O. Ibáñez-García, P. Goldstein, and A. Zarzosa-Pérez, "Hairpin polymer unfolding in square nanochannels," *J. Polym. Sci. B Polym. Phys.* **51**, 1411–1418 (2013).
- <sup>26</sup>A. Milchev, L. Klushin, A. Skvortsov, and K. Binder, "Ejection of a polymer chain from a nanopore: Theory and computer experiment," *Macromolecules* **43**, 6877–6885 (2010).
- <sup>27</sup>E. Werner and B. Mehlig, "Confined polymers in the extended de Gennes regime," *Phys. Rev. E* **90**, 062602 (2014).
- <sup>28</sup>J. M. Polson, A. F. Tremblett, and Z. R. N. McLure, "Free energy of a folded polymer under cylindrical confinement," *Macromolecules* **50**, 9515–9524 (2017).
- <sup>29</sup>M. Alizadehheidari, E. Werner, C. Noble, M. Reiter-Schad, L. K. Nyberg, J. Fritzsche, B. Mehlig, J. O. Tegenfeldt, T. Ambjörnsson, F. Persson, and F. Westerlund, "Nanoconfined circular and linear DNA: Equilibrium conformations and unfolding kinetics," *Macromolecules* **48**, 871–878 (2015).
- <sup>30</sup>M. T. J. van Loenhout, T. van der Heijden, R. Kanaar, C. Wyman, and C. Dekker, "Dynamics of RecA filaments on single-stranded DNA," *Nucl. Acids Res.* **37**, 4089–4099 (2009).
- <sup>31</sup>J. G. Reifemberger, H. Cao, and K. D. Dorfman, "Odijk excluded volume interactions during the unfolding of DNA confined in a nanochannel," *Macromolecules* **51**, 1172–1180 (2018).