CHALMERS

Extension of nanoconfined DNA in the extended de Gennes regime: comparison of theory and experiments

Erasmus Mundus Complex Systems Science M1 Project

Vitalii Iarko (930505-4711)

Department of Physics & Engineering Physics
Non-linear Dynamics & Statistical Physics
Chalmers University of Technology
Gothenburg, Sweden
October 22, 2015

Abstract

The study of DNA molecules confined to nanochannels is becoming more popular. A lack of precise theories and difficulties with model parameters have lead to an absence of quantitative analysis. A recent new theory for the extension in the extended de Gennes regime fills this gap. In this paper we evaluate the new theory by considering various theories for the model parameters of DNA, analyzing experimental data and comparing it to the predictions. We obtain qualitative agreement between the experiments and the theory with only minor issues. The remaining discrepancies might be caused by imperfections of existing theories for the model parameters or by the fact that the conditions for the extended de Gennes regime are only partially satisfied.

Acknowledgements

This project would not be possible without many great people, who I am enormously thankful to.

Firstly, my supervisors Bernhard Mehlig and Erik Werner, who not only developed the theory on which this work was based, but provided fertile soil and immeasurable support for the research.

Next, Toby St Clere Smithe, with whom I worked shoulder to shoulder and had many fruitful discussions.

Also, Mats Granath, who coordinates the program and is always ready to help.

Finally, all those who made Complex Systems Science Erasmus Mundus program, including Warwick University, Ecole Polytechnique and Chalmers University.

Vitalii Iarko, Zürich, Switzerland, August 2015

Contents

1

Introduction

Nowadays polymer nanoconfinement has many practical applications. It is a powerful tool to study single DNA molecules [1], which makes it possible to stretch the molecule and linearly unscroll its genome for analysis [2]. In addition to this, DNA separation and gene mapping rely on the change of molecular properties induced by the polymer confinement [3] and development of this area may lead to single-molecule analysis devices, which do not require cloning and molecular amplification [4]. Moreover, polymers in living systems usually are found in confined environments [5], thus, understanding of the influence of confinement is important for studying more complex biological systems [6]. For more applications one may see Ref. [2].

Such a variety of applications leads to substantial activity in this area, but at the same time many challenges remain. The main difficulty is caused by the small persistence length of DNA molecules ($\ell_{\mathrm{P}} \approx 50$ nm), which is smaller than the diffraction limit of visible light, i.e. one can not study microscopic configurations using fluorescence microscopy [6]. In addition to this, so far all proposed models use a number of interdependent parameters, for most of which we do not have well established theories and whose role in some cases is not studied at all. Notwithstanding all these difficulties, one still can connect large-scale observables such as the extension R of the DNA molecule in the channel direction with microscopic conformations and study the former, making various assumptions about the model parameters.

2

Background

A DNA molecule confined to a nanochannel exhibits different behaviors, depending on the channel size and the chemical composition of the solution in which the molecule is immersed [2]. The different regimes are classified into regimes in the literature. They are the Odijk regime [7], the extended de Gennes regime [8], and the de Gennes regime [9]. Here we will not discuss the fourth regime, the so-called backfolded Odijk regime [10] because it was only very recently developed. For all these regime there are theories for how the extension R depends on the conditions. Therefore, at first we have to define the extension R of the DNA molecule in the channel direction. One may see an illustration

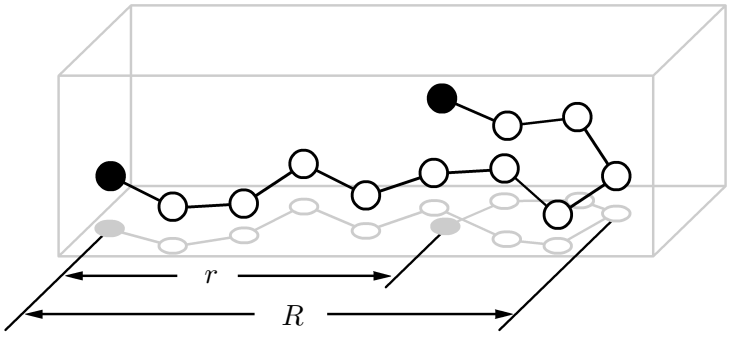

Figure 2.1: An illustration for the extension R of the DNA molecule in the channel direction and end-to-end distance r. The molecule is shown as a series of joined segments (in black). Its ends are represented by filled circles. The channel is shown in light gray as well as a projection of the molecule's segments on the channel's side.

on Fig. 2.1. It shows a simple model of a DNA molecule in a channel. The model consists of a series of joined segments and is projected on the channel's side. Then the end-to-end distance r is the distance between the polymers's ends along the channel. In contrast to the end-to-end distance, the extension R is a distance along the channel between the

leftmost and rightmost points of the polymer, which might not correspond to the two ends of the polymer.

Before considering the theories for the extension we need to discuss the physical parameters which describe the molecule. The contour length L is the length of the molecule, if it were stretched end-to-end and thermal fluctuations were ignored [2]. The persistence length ℓ_{P} quantifies the stiffness of the molecule, namely it is the contour distance over which correlations in the tangent direction decay [11]. The Kuhn length ℓ_{K} is a theoretical concept, so that one can consider the molecule as a group of freely joined rigid segments with length ℓ_{K} each [12]. In the case of DNA, $\ell_{\mathrm{K}} = 2\ell_{\mathrm{P}}$ [12]. The effective width w is the effective interaction range of the molecule. The channel with square cross-section is described by its side size D.

We describe the regimes in order of decreasing channel size D. At first the channel is so big, that the molecule is not influenced by the walls, and thus, it just forms one blob. If the channel size is much smaller than the size of this blob, but still obeys $D \gg \ell_{\mathrm{K}}^2/w$, then the molecule is in the de Gennes regime, and the average extension is expected to scale as $R \cong L\frac{(w\ell_{\mathrm{K}})^{1/3}}{D^{2/3}}$ [2]. In this regime, the molecule is described as a series of blobs with an unconfined self avoiding random walk in each of them. If we decrease D further to the region $\ell_{\mathrm{K}} \ll D \ll \ell_{\mathrm{K}}^2/w$, we encounter the extended de Gennes regime, which is similar to the de Gennes regime and has the same prediction for the extension, but at the same time differs, for instance, it has different scaling for the chain free energy [2, 8]. This regime is still described with a series of blobs, but here they become elongated [13]. Decreasing D further first leads to a transition between the extended de Gennes and the Odijk regimes and then, under $D \ll \ell_{\mathrm{K}}$, the molecule enters the Odijk regime, where $R = L\left[1 - A\left(\frac{D}{\ell_{\mathrm{K}}}\right)^{2/3}\right]$ and A is a numerical constant that depends on a channel geometry [2]. Here the molecule does not have enough space to form the blobs and the polymer just deflects from the walls.

In this thesis we focus on the the extended de Gennes regime. It was recently shown [14] that in this regime it is possible to derive rigorous expressions for the mean and variance of the extension, yielding

$$\mu_R/L = 0.9338(84)(\ell_{\mathrm{K}} w/D^2)^{1/3}, \tag{2.1}$$

$$\sigma_R/(L\ell_{\mathrm{K}})^{1/2} = 0.364(17), \tag{2.2}$$

where μ_R is an average extension and σ_R is its standard deviation. The uncertainties correspond to rigorous bounds for the prefactors, which have been derived by mapping the statistics of the polymer to a one-dimensional model known as the weakly self-avoiding random walk.

In order to compare experimental results to Eqs. (2.1) and (2.2) one needs to estimate the physical parameters of the molecule. They are highly dependent on the environment (the buffer) due to electrostatic interactions. First, the buffer is defined by the concentration of its components, which is usually denoted by a number before the buffer's name (for instance, 5×TBE). Using concentrations of components one can calculate the ionic

strength I_s of the buffer, which defines the salt level in the solution. If buffer contains ions with valences $z_1, \ldots, z_n$ and at concentrations $c_1, \ldots, c_n$ then its ionic strength is

$$I_s = \frac{1}{2} \sum_{i=1}^{n} c_i z_i^2.$$

The dependence of the model parameters on the ionic strength is not thoroughly studied, which makes it harder to do quantitative analysis.

As for ℓ_{P}, according to Dobrynin [15] "the problem of the electrostatic persistence length in solutions of charged polymers is one of the most controversial subjects of polymer physics". In Ref. [16] he obtains a new expression by fitting experimental results:

$$\ell_{\mathrm{P}} = 46.1\,\mathrm{nm} + \frac{1.9195\,\mathrm{M}^{1/2}}{\sqrt{I_s}}\,\mathrm{nm}\,. \tag{2.3}$$

Another expression for ℓ_{P} was given by Odijk [17] and Skolnick and Fixman [18]:

$$\ell_{\mathrm{P}} = 50 + \frac{0.0324\,\mathrm{M}}{I_s}\,\mathrm{nm}\,. \tag{2.4}$$

A comparison of both are given in [3].

As for effective width w, Stigter [19, 20] developed a theory based on excluded volume between two infinite cylinders. The expression is [3, 19]

$$w = \kappa^{-1}\left(0.7704 + \log\frac{2\pi v_{\mathrm{eff}}^2}{k_B T \epsilon \epsilon_0 \kappa}\right),$$

where κ is the inverse Debye length $\kappa^2 = (2N_A e^2 I_s)/(\epsilon_0 \epsilon k_B T)$; v_{eff} is an effective DNA line charge; ϵ is the dielectric constant of water; ϵ_0 is the permittivity of free space; e is the electronic charge; N_A is Avogadro's number; k_B is Boltzmann's constant and T is the absolute temperature [3].

Furthermore, the way the experiments can be conducted introduces more difficulties. Since DNA is not fluorescent, it is invisible for fluorescent microscopy. Therefore, the molecule must be labeled with a dye (such as YOYO-1 or TOTO-1 [21]), which might change the characteristics of the molecule. In addition to this, it is difficult to assure a uniform distribution of dye among the molecule [1]. Regarding the impact on contour length L there is some disagreement regarding how much one dye molecule extends the contour length (0.52 nm [22], 0.44 nm [23]). As for the effect of the dye on the persistence length the question is more controversial: Ref. [22] finds that YOYO dye does not change the persistence length, but according to Ref. [24] ℓ_{P} decreases with increasing dye concentration. The effect on effective width w is insufficiently investigated. In addition to this, experiments require nanochannels with nanoscale (~ 100 nm) sizes made without defects, which may influence the molecule.

Because of these uncertainties it is important to test the theoretical expressions for the parameters against experiments. A good overview of many of the experiments that have been performed can be found in the review by Reisner *et al.* [2]. The comparisons

between experiments and theory shown therein are complicated by the fact that the classical predictions for the extension of DNA in the extended de Gennes regime are depend on an undetermined prefactor.

In this thesis we make use of the fact that the recent theory [14] for the extension statistics in the extended de Gennes regime provide also the prefactors for the extension statistics, given in Eqs. (2.1) and (2.2). This allows us to make quantitative comparisons between theory and experimental results, and thereby to test the theoretical predictions for the parameters ℓ_{K} and w.

3

Method

The experimental measurements that we compare with theory were obtained by Nyberg *et al.* [1]. The dataset consists of 2388 time series, each corresponds to a λ-DNA molecule (48500 basepairs) and contains 200 measurements of the extension at different moments of time (see an example on Fig. 3.1). These values were obtained in the following way. First, for each molecule a series of pictures was taken with a microscope (the exposure time is 100 ms and the lag between frames is 84 ms, which overall gives 5.4 frames per second). Then the series was analyzed and numerical values of observables were extracted. The channel's cross section is $\approx 100 \times 150\,\mathrm{nm}^2$. The samples were immersed in a TBE buffer and stained with YOYO-1 dye (under various concentrations of both). Originally, the dataset was used for studying dye staining, that is why for each molecule its luminosity was measured. In addition to this some molecules were heated, namely they were kept under 50 °C for 3 or 24 hours before the measurements. For more details about the experiments and the dataset one may refer to the original paper [1].

One notable property of Eqs. (2.1) and (2.2) is that while the mean μ_R explicitly depends on the effective width w, the standard deviation σ_R does not. In the dataset there are series with 4 different buffer concentrations. Since the effective width w depends on the ionic strength I_s, which depends on the buffer concentration, the effective width w varies in the dataset. Therefore, it should be possible to notice this difference in the experimental data. But in order to do so we need to calculate the effective width w for the buffer concentrations in the dataset.

As it was discussed before, there is no perfect agreement in the literature regarding dependence of the effective width w on the ionic strength I_s, but many studies [2, 3, 10, 23] used the theory by Stigter [20]. In our analysis we will use it too, namely we use tabular data from Ref. [20] with linear interpolation in between. Lacking a better estimate, we ignore the effect of the dye on w.

As for ionic strength I_s, we calculate it by obtaining equilibrium concentrations for all ions in the solution. The technique is described in great detail in Appendix B.

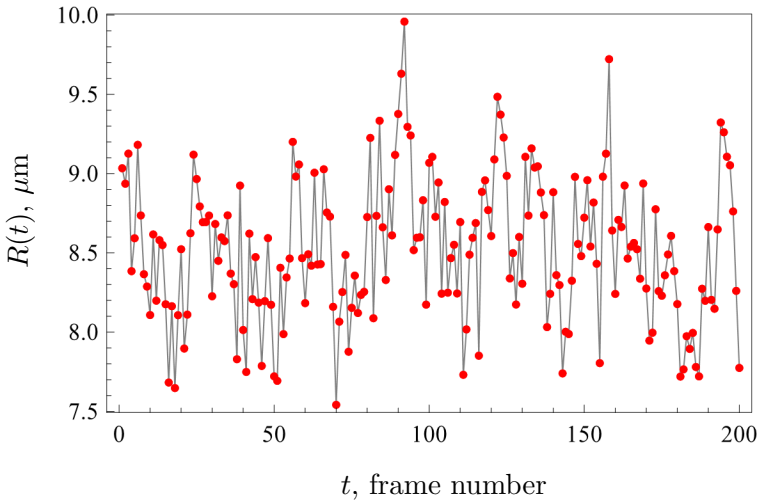

Figure 3.1: An example of a series from the dataset. Red points show the measured extension at a given frame. Consecutive points connected with light grey line for clarity. The molecule was not heated, its luminosity is 182810, the buffer is 0.05×TBE.

Except for the effective width w we also need values of the Kuhn length ℓ_{K} and contour length L, since the theory expressions for both mean μ_R and standard deviation σ_R depend on them. As for the Kuhn length ℓ_{K}, we will use that $\ell_{\mathrm{K}} = 2\ell_{\mathrm{P}}$ [12], and Dobrynin's expression for ℓ_{P} then becomes

$$\ell_{\mathrm{K}} = 2\left(46.1 + \frac{1.9195\,\mathrm{M}^{1/2}}{\sqrt{I_s}}\right)\mathrm{nm}\,.$$

As for the contour length, the height of one base pair of a DNA molecule is 0.34 nm [25]. The number of base pairs in a λ-DNA is 48502, thus $L = 48502 \times 0.34\,\mathrm{nm} = 16490.7\,\mathrm{nm}$.

One should note that in the dataset the conditions $\ell_{\mathrm{K}} \ll D \ll \ell_{\mathrm{K}}^2/w$ of the extended de Gennes regime are not well satisfied. According to the theories above, the values of ℓ_{K} in the data are between 100 and 150 nanometers and the values of w lie from 5 to 26 nanometers. Therefore, we cannot expect quantitative agreement.

In order to do the comparison, we need to estimate the mean μ_R and the standard deviation σ_R using the series from the dataset. The difficulty is that consequent measurements in each series may be correlated, that is the series may be autocorrelated. If we do not take this into account, this may lead to biased estimations.

Therefore, we need to start with checking how autocorrelated the data is. We use the initial positive sequence (IPS) estimator [26] in order to calculate an autocorrelation time τ, that is a time after which measurements become uncorrelated. The main idea of the IPS estimator is that we calculate τ as

$$\tau = 1 + 2\sum_{k=1}^{\infty}\rho_k, \tag{3.1}$$

where ρ_k is the autocorrelation function. Since we have only the series, we can estimate ρ_k only for small lags k as

$$\hat{\rho}_k = \frac{1}{ns^2} \sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X}), \tag{3.2}$$

where n is a number of measurements in the series, X_i is the i-th measurement, $\bar{X}$ is a sample mean and s^2 is a sample variance. Finally, we truncate the sum in (3.1) when $\hat{\rho}_i + \hat{\rho}_{i+1}$ becomes negative, that is why the estimator called positive.

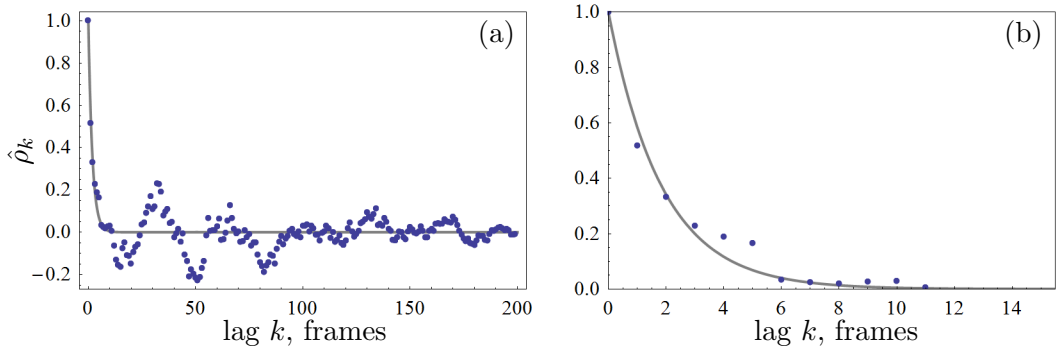

Figure 3.2: The estimation of the autocorrelation function for the molecule from Fig. 3.1. Points are the estimated values of the autocorrelation function ρ_k according to (3.2), line is a a^k fit, where $a = 0.584$. Panel (b) is a close up for lags k from 0 to 15 frames. The autocorrelation time τ is 4.11 frames (0.757 seconds). Note, that the values were normalized $\hat{\rho}_0 = 1$.

As for the mean, the series were obtained using different molecules, that is they are pairwise uncorrelated. Therefore, sample means of each series are uncorrelated as well. Since we have enough data, we can just calculate a mean of these values as an estimation of the extension mean, that is to find a sample mean of all measurements. The standard error of this estimation is a sample (over the series means) standard deviation divided by the number of series.

As for the standard deviation, this approach will not work, since an estimation for each timeseries will be biased by itself. Using [27, p.284, (4.6)] we can correct the bias:

$$E[s^2] = \sigma^2 \left(1 - \frac{2}{n-1} \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \rho_k \right),$$

where $E[s^2]$ is an expected value of the sample variance, n is a number of measurements and ρ_k are theirs autocorrelation function values. This expression was obtained using covariance-stationary model, that is under an assumption that the values are from covariance-stationary stochastic process. For such a process the mean and variance do not change over time and covariance between two observations depends only on the time between them, but not on the actual time values [27, p. 280]. A potential problem of

this assumption in our case is that the DNA molecule may damage itself during the experiment, changing the mean. We did not observe these changes in the data and, therefore, ignore it.

Thus, in order to find an unbiased estimation of the sample variance, we just need to divide the sample variance by the expression in brackets. Then we take a square root of it to obtain an estimation for the standard deviation. Unfortunately, with such an approach we cannot estimate the standard error, and are obliged to disregard it. Note that later the analysis was redone using a random-coefficient model. The results were similar to the approach described above. In addition to this, this model provided the standard error for the standard deviation. The results are in Appendix A, for the model description one may see Appendix B.

In order to calculate expression in brackets we need to know the autocorrelation function ρ_k. So far we can just estimate it using (3.2). This will add a bias, but due to the lack of better approach we will ignore it. Therefore, we assume that $\rho_k = a^k$, where a is a constant.

Now we need to estimate a. As you can see on Fig. 3.2 our assumption about the form of ρ_k disagrees with obtained values due to noise, that is the estimation for each molecule independently gives unstable results. As a solution, we will consider a group of molecules obtained under similar conditions (for instance, same buffer concentration and close luminosities). More formally, for given lag k we find $\hat{\rho}_k$ for each sequence independently and then take average as an estimate for the entire group. Using this estimate we can find a by fitting a^k.

On Fig. 3.3 you can see an example of the autocorrelation function values estimated on a group of molecules. This group includes the molecule from Fig. 3.2 and 10 molecules which are in the same buffer and are the closest by luminosity. As you can see we obtain much better agreement.

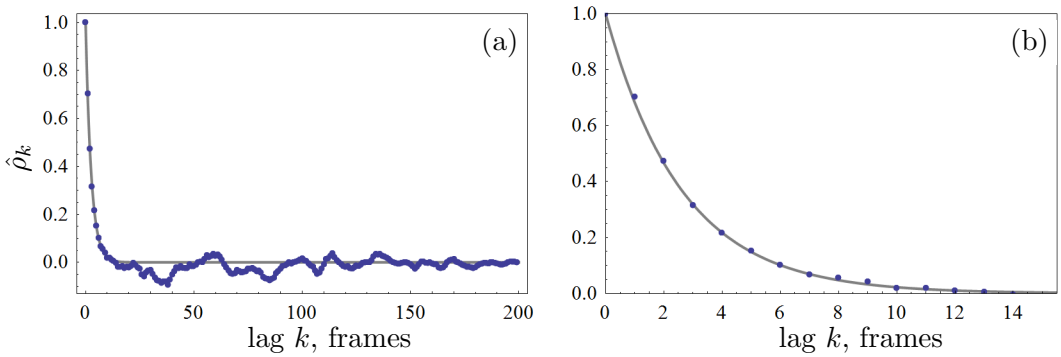

Figure 3.3: The estimation of the autocorrelation function for a group of 11 molecules: the molecule from Fig. 3.2 (its luminosity 182810) and 10 molecules which are the closest by luminosity (their luminiosities are bounded by 182160 and 183600, maximal value in the dataset is 589750), all are in 0.05×TBE buffer. Points are the estimated values, the curve is a fit in form a^k, $a = 0.682757$, the autocorrelation time is 5.35 frames (0.98 seconds).

Now we can start our analysis by checking the dependence on effective width w and

heating time. Note that both w and ℓ_{K} depend on the ionic strength I_s, but we invert this relation and consider them as functions of w. So far we ignore dye intercalation. The corresponding plots are shown on Fig. 3.4. There we can see a good agreement between experiments and the theory, but not perfect. In addition to this we see the difference in dependance on effective width w, namely the mean increases almost twice with increasing w, but the standard deviation changes less than 20 %. Note that the theory expression for standard deviation is not a constant due to dependence of ℓ_{K} on I_s. Finally, results under different heating times are similar, thus, we will ignore heating time in further analysis.

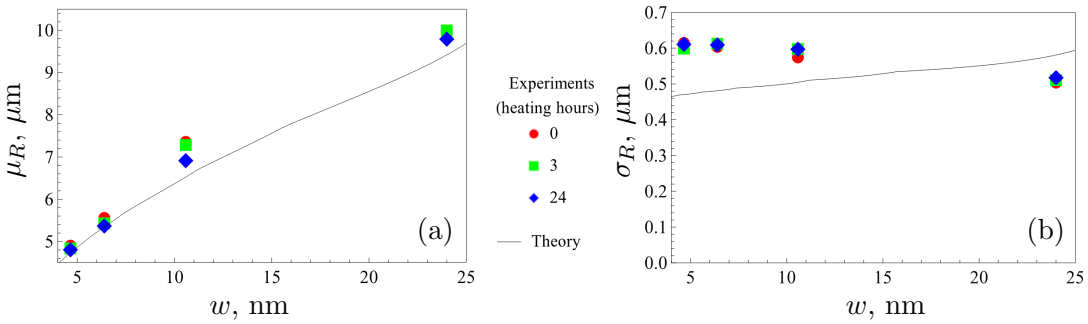

Figure 3.4: Mean and standard deviation as functions of effective width w for different heating times without taking into account dye intercalation. In case of the mean error bars are smaller then markers, for standard deviation the error is ignored (see text).

The main flaw of such an approach is that it ignores the influence of dye intercalation on contour length L. In order to take it into account, we need to know how much one dye molecule extends the contour length and the quantity of dye molecules intercalated. Estimates for dye molecule height vary from 0.4 nm to 0.5 nm with 10% uncertainty [22, 28]. We assume that it is 0.44 nm [23]. As for the number of dye molecules intercalated, we can use the measured luminosity l. First, we assume that the luminosity l depends linearly on the number of dye molecules. Then, we assume that maximal intercalation (1 dye molecule per 4 base pairs) was achieved in the dataset, that is we know corresponding luminosity $l_{\max}$. Then we consider relative luminosity $l_{\mathrm{rel}} = l/l_{\max}$. Since we know number of basepairs and contour length $L(0) \equiv L$ without dye (that is relative luminosity is 0), we can calculate contour length $L(1)$ under maximal intercalation (corresponds to relative luminosity 1):

$$L(1) = L(0) + \frac{48502 \times 0.44\,\mathrm{nm}}{4}$$

Using our assumption about linear dependence of luminosity on the number of dye molecules, we find contour length $L(l_{\mathrm{rel}})$ under relative luminosity l_{rel} as a linear interpolation between $L(0)$ and $L(1)$:

$$L(l_{\mathrm{rel}}) = (1 - l_{\mathrm{rel}})L(0) + l_{\mathrm{rel}}L(1)$$

If we check a distribution of relative luminosities l_{rel} in the dataset (see Fig. 3.5), we can see that it is very uneven. Thus, we need to use binning for the estimation of contour

length $L(l_{\mathrm{rel}})$. As we can see on Fig. 3.5 the bin width of 0.04 provides good binning. Therefore, we distribute the series into bins by their relative luminosities, then for each bin we assume that all molecules inside have one relative luminosity, which equals to the middle value of the bin. Then by using this relative luminosity value, we can estimate contour length of the molecules in the bin.

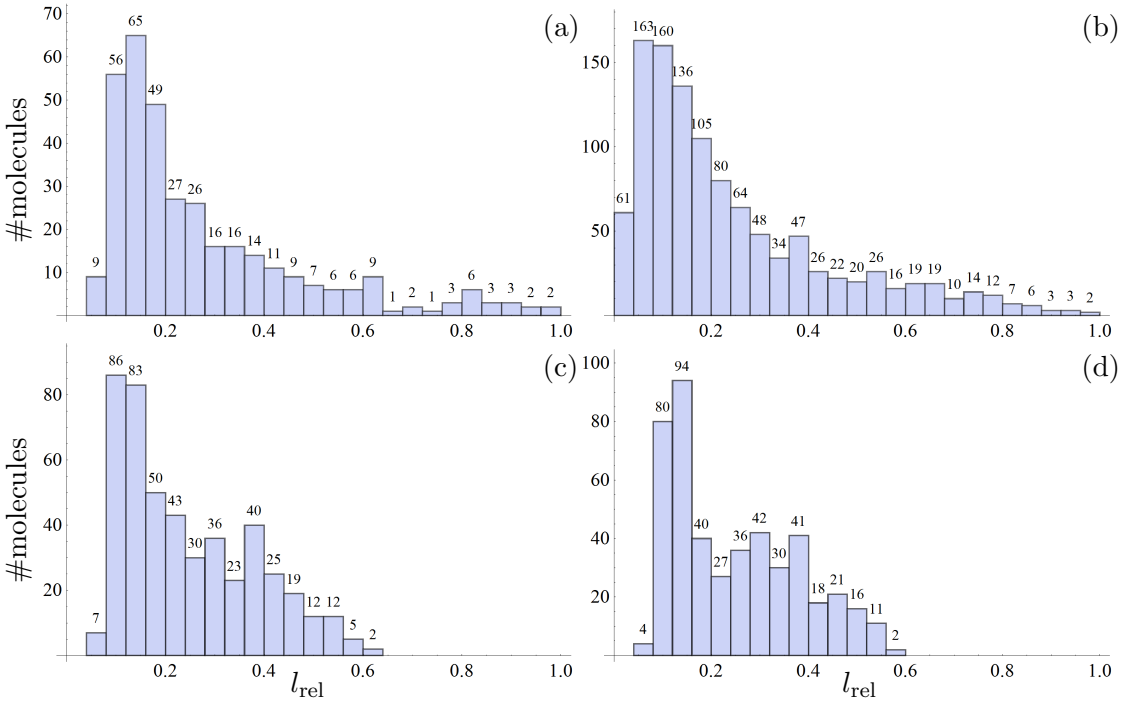

Figure 3.5: Distributions of relative luminosities in the dataset, the bin width is 0.04, binning starts from $l_{\mathrm{rel}} = 0$. Panels correspond to 0.05, 0.5, 2 and 5×TBE buffer concentrations.

After plotting the data as function of relative luminosity, we obtain Fig. 3.6. This time the difference in the dependence on effective width w is not clear, but still can be observed. Since the effective width w depends on the buffer concentration, different colors on the figure (that is the buffer concentrations) correspond to different values of effective width. We can see how the mean grows almost twice with decreasing the buffer concentration from 5 to 0.05×TBE (that is going from black points to red), on the contrary, the standard deviation mostly stays on the same level. As you can see, taking into account dye intercalation makes the agreement between the experiments and the theory better, especially under low buffer concentrations for the mean.

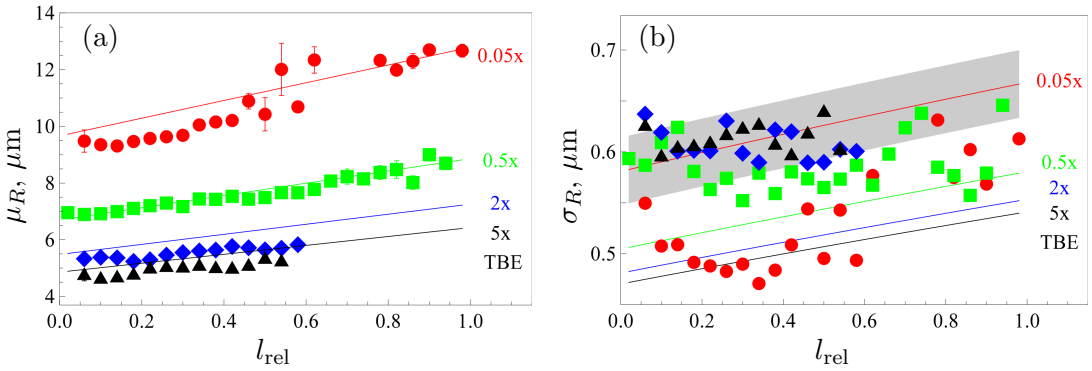

Figure 3.6: The experiments (points) and the theory (lines) as functions of relative luminosity. The panel (a) shows the mean, the panel (b) corresponds to the standard deviation. Different colors correspond to the different buffer concentrations: 0.05×TBE (red ○), 0.5×TBE (green □), 2×TBE (blue ◊) and 5×TBE (black △). The gray band in the panel (b) shows 5% uncertainty in the theory constant, all other uncertainties on this panel are bounded by this band. In panel (a) 1% the uncertainty is not shown. Standard errors for the standard deviation are not shown.

4

Conclusions

We have analyzed measurements of the extension statistics of channel-confined DNA and compared them to the new theory in the extended de Gennes regime by Werner and Mehlig [14]. We observe qualitative agreement between the experiments and the theory. That the agreement is not perfect is to be expected, since the conditions for the extended de Gennes regime are only partially fulfilled. Nevertheless, we find that the theories for the parameters of DNA together with Eqs. (2.1) and (2.2) for the extension statistics reproduce the experimental measurements surprisingly well. These results suggest that the method presented here allows, in principle, to test the theoretical predictions for how the parameters ℓ_{K} and w describing a DNA molecule depend on the ionic strength of the solution.

5

About the article

The described analysis became a basis of an article. Please find attached the article in Appendix A and its supplementary materials in Appendix B.

This analysis was improved and extended there in multiple ways:

- The estimation of the mean and standard deviation was done using a random-coefficient model, a variation of a linear-regression model adapted to correlated data. The results were similar to the approach in this report, but not exactly same. In addition to this, with the new model we estimated errors not only for the mean, but for the standard deviation as well;

- Two new datasets were obtained and analysed. This report is based on λ-DNA molecules in $\approx 100 \times 150\,\mathrm{nm}^2$ channel under 4 different buffer concentrations (0.05, 0.5, 2 and 5×TBE) and the experiments duration was 38.4 seconds. Both new datasets were obtained on T4-DNA molecules, which are longer than the λ-DNA (165647 basepairs against 48502 correspondingly). The first new data was obtained in a nanofunnel (one side was 120 nm, other side gradually changed from 92 nm to 815 nm) and 0.05 (the experiments duration was 100 seconds) and 2×TBE buffers (the duration was 400 seconds). The second new data was obtained in a square channel $300 \times 300\,\mathrm{nm}^2$ with the same 4 buffer concentrations (0.05, 0.5, 2 and 5×TBE), but the recordings were done for 200 seconds. Such settings enabled us to conclude more about the theories for the physical parameters;

- The channel was remeasured, it happen to be trapezoidal with a width at the top (bottom) of 130 nm (87 nm). The further assumption was that it could be approximated with a rectangular channel with $D_W = 108\,\mathrm{nm}$.

The article was sent to Physical Review E, so far it is on a review stage.

Bibliography

[1] L. Nyberg, F. Persson, B. Akerman, F. Westerlund, Heterogeneous staining: a tool for studies of how fluorescent dyes affect the physical properties of DNA.

[2] W. Reisner, J. N. Pedersen, R. H. Austin, DNA confinement in nanochannels: physics and biological applications, Reports on Progress in Physics 75 (10) (2012) 106601.

[3] C.-C. Hsieh, A. Balducci, P. S. Doyle, Ionic effects on the equilibrium dynamics of DNA confined in nanoslits, Nano letters 8 (6) (2008) 1683–1688.

[4] W. Reisner, N. B. Larsen, A. Silahtaroglu, A. Kristensen, N. Tommerup, J. O. Tegenfeldt, H. Flyvbjerg, Single-molecule denaturation mapping of DNA in nanofluidic channels, Proceedings of the National Academy of Sciences 107 (30) (2010) 13294–13299.

[5] E. Werner, F. Westerlund, J. O. Tegenfeldt, B. Mehlig, Monomer distributions and intrachain collisions of a polymer confined to a channel, Macromolecules 46 (16) (2013) 6644–6650.

[6] E. Werner, F. Persson, F. Westerlund, J. O. Tegenfeldt, B. Mehlig, Orientational correlations in confined DNA, Physical Review E 86 (4) (2012) 041802.

[7] T. Odijk, The statistics and dynamics of confined or entangled stiff polymers, Macromolecules 16 (8) (1983) 1340–1344.

[8] E. Werner, B. Mehlig, Scaling regimes of a semiflexible polymer in a rectangular channel, Physical Review E 91 (5) (2015) 050601.

[9] P.-G. De Gennes, Scaling concepts in polymer physics, Cornell university press, 1979.

[10] A. Muralidhar, D. R. Tree, K. D. Dorfman, Backfolding of wormlike chains confined in nanochannels, Macromolecules 47 (23) (2014) 8446–8458.

[11] M. Doi, S. F. Edwards, The theory of polymer dynamics, Vol. 73, oxford university press, 1988.

[12] G. R. Strobl, The physics of polymers, Vol. 2, Springer, 1997.

[13] T. Odijk, Scaling theory of DNA confined in nanochannels and nanoslits, Physical Review E 77 (6) (2008) 060901.

[14] E. Werner, B. Mehlig, Confined polymers in the extended de gennes regime, Physical Review E 90 (6) (2014) 062602.

[15] A. V. Dobrynin, Electrostatic persistence length of semiflexible and flexible polyelectrolytes, Macromolecules 38 (22) (2005) 9304–9314.

[16] A. V. Dobrynin, Effect of counterion condensation on rigidity of semiflexible polyelectrolytes, Macromolecules 39 (26) (2006) 9519–9527.

[17] T. Odijk, Polyelectrolytes near the rod limit, Journal of Polymer Science: Polymer Physics Edition 15 (3) (1977) 477–483.

[18] J. Skolnick, M. Fixman, Electrostatic persistence length of a wormlike polyelectrolyte, Macromolecules 10 (5) (1977) 944–948.

[19] D. Stigter, Interactions of highly charged colloidal cylinders with applications to double-stranded DNA, Biopolymers 16 (7) (1977) 1435–1448.

[20] D. Stigter, Donnan membrane equilibrium, sedimentation equilibrium, and coil expansion of DNA in salt solutions, Cell biophysics 11 (1) (1987) 139–158.

[21] H. S. Rye, S. Yue, D. E. Wemmer, M. A. Quesada, R. P. Haugland, R. A. Mathies, A. N. Glazer, Stable fluorescent complexes of double-stranded DNA with bis-intercalating asymmetric cyanine dyes: properties and applications, Nucleic acids research 20 (11) (1992) 2803–2812.

[22] K. Günther, M. Mertig, R. Seidel, Mechanical and structural properties of YOYO-1 complexed DNA, Nucleic acids research 38 (19) (2010) 6526–6532.

[23] D. Gupta, J. Sheats, A. Muralidhar, J. J. Miller, D. E. Huang, S. Mahshid, K. D. Dorfman, W. Reisner, Mixed confinement regimes during equilibrium confinement spectroscopy of DNA, The Journal of chemical physics 140 (21) (2014) 214901.

[24] C. U. Murade, V. Subramaniam, C. Otto, M. L. Bennink, Force spectroscopy and fluorescence microscopy of dsDNA–YOYO-1 complexes: implications for the structure of dsDNA in the overstretching region, Nucleic acids research 38 (10) (2010) 3423–3431.

[25] R. R. Sinden, DNA structure and function, Elsevier, 2012.

[26] M. B. Thompson, A comparison of methods for computing autocorrelation time, arXiv preprint arXiv:1011.0175.

[27] A. M. Law, W. D. Kelton, Simulation modeling and analysis, McGraw-Hill, 1991.

[28] F. Johansen, J. P. Jacobsen, 1H NMR studies of the bis-intercalation of a homodimeric oxazole yellow dye in DNA oligonucleotides, Journal of Biomolecular Structure and Dynamics 16 (2) (1998) 205–222.

A

Article

Extension of nano-confined DNA: quantitative comparison between experiment and theory

V. Iarko[1], E. Werner[1], L. K. Nyberg[2], V. Müller[2], J. Fritzsche[3], T. Ambjörnsson[4], J. P. Beech[5], J. O. Tegenfeldt[5,6], K. Mehlig[7], F. Westerlund[2], B. Mehlig[1]
[1]*Department of Physics, University of Gothenburg, Sweden*
[2] *Department of Biology and Biological Engineering, Chalmers University of Technology, Sweden*
[3] *Department of Applied Physics, Chalmers University of Technology, Sweden*
[4]*Department of Astronomy and Theoretical Physics, Lund University, Sweden*
[5]*Department of Physics, Division of Solid State Physics, Lund University, Sweden*
[6]*NanoLund, Lund University, Sweden and*
[7]*Department of Public Health and Community Medicine, University of Gothenburg, Sweden*

The extension of DNA confined to nanochannels has been studied intensively and in detail. Yet quantitative comparisons between experiments and model calculations are difficult because most theoretical predictions involve undetermined prefactors, and because the model parameters (contour length, Kuhn length, effective width) are difficult to compute reliably, leading to substantial uncertainties. Here we use a recent asymptotically exact theory for the DNA extension in the 'extended de Gennes regime' that allows us to compare experimental results with theory. For this purpose we performed new experiments, measuring the mean DNA extension and its standard deviation while varying the channel geometry, dye intercalation ratio, and ionic buffer strength. The experimental results agree very well with theory at high ionic strengths, indicating that the model parameters are reliable. At low ionic strengths the agreement is less good. We discuss possible reasons. Our approach allows, in principle, to measure the Kuhn length and effective width of a single DNA molecule and more generally of semiflexible polymers in solution.

PACS numbers: 87.15.-v, 36.20.Ey, 87.14.gk

Nano-confined DNA has recently been intensively studied [1–6] as a means of stretching the molecules in order to study local properties (e.g. DNA sequence [7, 8]). A fundamental question is how the physical properties of the DNA and the solution affect the extent to which the molecule is stretched by confinement. Experimentally this question has been investigated in detail, varying the confinement, the length of the DNA molecule, and the properties of the solution (see Ref. [1] for a review).

It is commonly assumed that DNA can be modeled as a semiflexible polymer with hard-core repulsive interactions [9–15]. Measurements [16–19] and theoretical considerations [9, 16] indicate that a worm-like chain model may be a good approximation. A recent study [3] compares experimental results for the extension of confined λ-DNA to results of computer simulations of a self-avoiding discrete worm-like chain model, indicating that it may describe the experimental results well.

Yet quantitative comparison between experiments and theoretical model calculations has remained difficult for at least two reasons. First the model parameters (contour length, Kuhn length $\ell_{\rm K}$, and effective width $w_{\rm eff}$ of the semiflexible polymer) are difficult to determine reliably: there is substantial uncertainty regarding the physical properties of DNA. Second, the model is hard to analyse theoretically. But a recent asymptotically exact theory [20, 21] for the extension of a confined self-avoiding semiflexible polymer in the so-called 'extended de Gennes regime' [11, 13, 22] overcomes the second difficulty: it makes precise predictions for the prefactors and exponents defining scaling laws as a function of the physical parameters, Eqs. (2) below. This opens the possibility to experimentally determine $\ell_{\rm K}$ and $w_{\rm eff}$ by measurements of confined DNA. In this article we report on experimental results mapping out how the extension of confined DNA in the extended de Gennes regime depends on channel geometry, ionic strength of the solution, and upon the amount of dye bound to the molecule.

At high ionic strengths we find very good agreement between experiment and theory using approximations for $\ell_{\rm K}$ and $w_{\rm eff}$ that are commonly employed [23–26], and taking into account how the contour length depends on the amount of dye molecules bound to the DNA. The comparison between experiment and theory is so precise that it enables us to detect subtle alignment effects [27] at the border of the extended de Gennes regime. At low ionic strengths the agreement is not as good. This may indicate that theoretical estimates of $\ell_{\rm K}$ and $w_{\rm eff}$ must be improved. We expect it is possible to experimentally precisely determine $\ell_{\rm K}$ and $w_{\rm eff}$ by extending the approach described in this article.

Extended de Gennes regime. Consider a semiflexible polymer of contour length L, Kuhn length $\ell_{\rm K}$ [9] and excluded volume v per Kuhn length segment. The excluded volume is often written in terms of an effective width $w_{\rm eff}$, defined by the relation $v \equiv (\pi/2)\ell_{\rm K}^2 w_{\rm eff}$. This expression for the excluded volume is based on Onsager's result [28] for the excluded volume of a cylinder of length $\ell_{\rm K}$ and diameter $w_{\rm eff}$, which in the limit $\ell_{\rm K} \gg w_{\rm eff}$ reduces to the expression above. We phrase our results in terms of

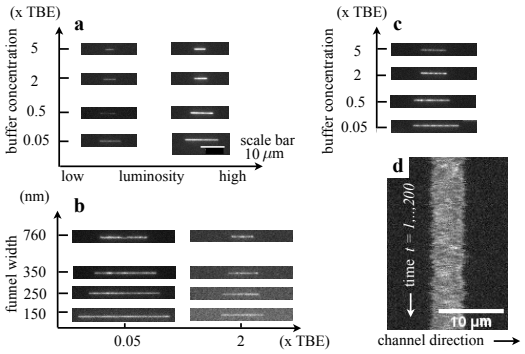

FIG. 1: **a** Experiment 1. λ-DNA in a 150 nm × 108 nm channel, different buffer concentrations, and different luminosities corresponding to different dye loadings. Shown are representative video frames (scale bar applies to panels **b**,**c** as well). **b** Experiment 2. T4-DNA in a nanofunnel in 0.05× and 2×TBE solution, varying funnel width $D_{\rm W}$ at constant $D_{\rm H} = 120$ nm. **c** Experiment 3. T4-DNA in a 302 nm × 300 nm channel, different buffer concentrations. **d** Time trace of the fluorescence intensity for λ-DNA in a 108 nm × 150 nm channel in 5×TBE solution, center-of-mass motion subtracted.

the effective width $w_{\rm eff}$. This is customary but we stress that, strictly speaking, it is the excluded volume that determines the statistics of the polymer, and that even for DNA models with only hard-core repulsion between rod-like segments, the effective width does not equal the actual width of the rods, except in the limit $\ell_{\rm K} \gg w_{\rm eff}$. This is important to consider when evaluating the results of simulations.

The polymer is confined to a channel with cross section $D_{\rm W} \times D_{\rm H}$. The polymer exhibits different confinement regimes distinguished by different laws for the polymer extension in the channel direction [21]. The extended de Gennes regime is defined by the conditions [21]

$$\ell_{\rm K} \ll D_{\rm H} \ll \ell_{\rm K}^2/w_{\rm eff} \quad \text{and} \quad D_{\rm W}^2 \ll D_{\rm H}\ell_{\rm K}^2/w_{\rm eff}\,, \quad (1)$$

where we assume that $D_{\rm W} \geq D_{\rm H}$. For a square channel the corresponding conditions were previously derived and discussed in Refs. [10, 11, 13, 20, 22]. In regime (1) exact expressions for the mean μ and the variance σ^2 of the extension in the channel direction are known [20, 21], provided that the contour length is long enough [20]:

$$\mu/L = 0.9338(84)\,[\ell_{\rm K} w_{\rm eff}/(D_{\rm W} D_{\rm H})]^{1/3}\,, \quad (2a)$$

$$\sigma/(L\ell_{\rm K})^{1/2} = 0.364(17)\,. \quad (2b)$$

The errors quoted for the coefficients reflect strict bounds [20] derived from the exact results of Ref. [29]. Provided L is known Eqs. (2) allow to infer $\ell_{\rm K}$ and $w_{\rm eff}$ from measurements of nanoconfined DNA molecules.

Calculation of parameters. We now discuss how the parameters L, $\ell_{\rm K}$ and $w_{\rm eff}$ are commonly estimated for DNA, and what the main uncertainties are. We first discuss bare DNA, before considering the effect of staining with fluorescent dye (YOYO-1).

	0.05×TBE	0.5×TBE	2×TBE	5×TBE
$I_{\rm s}$[mM]	3.81	24.9	78.4	178.0
$\ell_{\rm K}$[nm]	154.4	116.5	105.9	101.2
$w_{\rm eff}$[nm]	26	10	6.2	4.6

TABLE I: Numerical values for the ionic buffer strength $I_{\rm s}$, the Kuhn length $\ell_{\rm K}$, and the effective width $w_{\rm eff}$ (see text).

The contour length of bare DNA is 0.34 nm per base pair, with an uncertainty of about 0.01 nm, or 3% [30].

DNA is commonly modeled as a worm-like chain, for which the Kuhn length is twice the persistence length, $\ell_{\rm K} = 2\ell_{\rm P}$ [9]. The persistence length has been measured by a number of different techniques [16, 17, 19, 31] yielding $\ell_{\rm P} \approx 45$–50 nm at high ionic strength ($I_{\rm s} \gtrsim 10$ mM), and increasing at lower $I_{\rm s}$. Following Refs. [3, 25, 26, 32] we use the empirical formula suggested in Ref. [23] $\ell_{\rm P} \approx 46{\rm nm} + 1.92\ I_{\rm s}^{-1/2}\,{\rm M}^{1/2}$ nm. While this dependence of $\ell_{\rm P}$ upon $I_{\rm s}$ has a theoretical basis, the prefactors are known only from an empirical fit with an uncertainty of about 10% (Fig. 3 in Ref. [23]). The resulting values of $\ell_{\rm K}$ are given in Table I, the calculation of $I_{\rm s}$ is described in the Supplemental Material [33].

Stigter [24] computed the excluded volume between two long, strongly charged cylinders in NaCl solution, and applied this calculation to DNA to obtain an estimate of $w_{\rm eff}$. Linear interpolation on a doubly logarithmic scale of the effective widths given in Table 1 of Ref. [24] yields the values tabulated in Table I. There are many sources of uncertainty when applying this theory to our system. Stigter's calculation [24] for $w_{\rm eff}$ assumes that the Kuhn length segments can be approximated by infinitely long cylinders with an intrinsic width of 1.2 nm, and that the effective line charge of DNA is given by $0.73e^-$ per phosphate group. The approximation of infinite cylinders is problematic when the Kuhn length and the effective width are of the same order, i.e. for low ionic strengths. According to Stigter, the value 1.2 nm has an uncertainty of about 20%, leading to an uncertainty for the effective width of 5-10% [34], with a larger effect at large ionic strengths. The effective line charge estimate is based on measurements in NaCl-solutions, which do not generalise to other ions [35].

The DNA contour length is expected to increase in proportion to the amount of dye bound. We assume that dye intercalation extends the bare contour length of the DNA molecule by 0.44 nm per dye molecule [3]. Estimates of this number range from about 0.4 nm to 0.5 nm, with large uncertainties in the individual estimates [36, 37], the uncertainty is at least 10%. At a dye loading of 1 molecule per 10 base pairs, this corre-

sponds to an uncertainty in the contour length of $\approx 2\%$. There is no consensus regarding how intercalating dye molecules affect the parameters ℓ_K and $w_{\rm eff}$. Ref. [38] finds that the Kuhn length decreases with increasing dye load, whereas Ref. [37] finds no dependence. Since the dye molecules are positively charged, the effective width might decrease with dye load, but the magnitude of this effect is not known. Lacking a better estimate, it is commonly assumed that YOYO- binding does not affect these parameters [3, 26, 39]. In summary, while the contour length of DNA is known to a rather high accuracy, there is substantial uncertainty regarding the parameters ℓ_K and $w_{\rm eff}$.

Experimental method. The experimental data are obtained measuring the extension of single DNA molecules in nanochannels under different conditions (Fig. 1**a** to **c**). We use linear λ- and T4GT7-DNA (T4-DNA for short) with definite contour lengths of 48502 and 165647 base pairs, respectively (this ensures that L is large enough in our experiments, it exceeds $(D_W^2 D_H^2 \ell_K / w_{\rm eff}^2)^{1/3}$ by an order of magnitude [21]). The molecules are stained with YOYO dye and suspended inside a channel in a TBE (Tris-Borate-EDTA) buffer.

The first experiment is a re-analysis of data presented in Ref. [32]. In this experiment (Fig. 1**a**), λ-DNA is inserted into a nanochannel of height $D_H = 150$ nm and width $D_W = 108$ nm. We discuss the uncertainty in the channel dimensions in the Supplemental Material [33]. DNA extensions are measured at different buffer conditions (0.05×, 0.5×, 2× and 5×TBE), and at different dye loads. To estimate the dye load of a molecule we assume that it is proportional to the luminosity, and that the largest observed luminosity corresponds to full intercalation (one dye molecule per four base pairs). In this way we obtain an estimate for the amount of dye bound to the molecule by linear interpolation.

In experiment 2 (Fig. 1**b**), T4-DNA is inserted into a nanofunnel, with fixed height $D_H = 120$ nm and gradually changing width from $D_W = 92$ nm to $D_W = 815$ nm over a length of 500 μm. These experiments are at two different buffer concentrations (0.05× and 2×TBE).

In experiment 3 (Fig. 1**c**), T4-DNA is inserted into a channel with $D_W = 302$ nm, $D_H = 300$ nm. The buffer concentration is varied (0.05×, 0.5×, 2×, and 5×TBE). In experiments 2 and 3, the average dye load at 0.05×TBE is approximately 1 dye molecule per 10 base pairs. Assuming that the dye load is proportional to luminosity we estimate the dye load in experiment 2 at 2×TBE to 1 dye molecule per 45 base pairs, and 1 per 12, 16, 28 base pairs at 0.5×, 2×, and 5×TBE in experiment 3.

For each molecule 200 frames are recorded. Fig. 1**d** shows an example of a fluorescence-intensity trace ('kymograph') obtained in this way. Each row in the kymograph shows the fluorescence intensity in a given frame averaged over the channel cross section. Bright regions correspond to high intensity indicating where the DNA molecule is located. The extension of the molecule along the channel is simply the width of the bright region.

For a given set of parameters we estimate the mean and standard deviation of the extension by a linear mixed model that takes into account the fact that the measured extensions are correlated in time. Details concerning the experimental method and the data analysis are given in the Supplemental Material [33].

Results. Our results are shown in Fig. 2. We plot two theoretical curves. The solid curve uses the actual channel size $D_H \times D_W$. The dashed curve compensates for the repulsive interaction with the negatively charged walls [1], by using an 'effective channel size' $(D_H - \delta)\times(D_W - \delta)$. We take $\delta = w_{\rm eff}$, but it is not known how accurate this estimate is. Since the standard deviation is independent of channel size in the extended de Gennes regime, the compensation does not affect this comparison.

The results of experiment 1 are shown in panels **a** and **b**. At low relative luminosity (small dye-to-basepair ratio) the average extension is well described by Eq. (2a). For the standard deviation there are larger differences between experiment and Eq. (2b). Possible reasons are discussed below.

We turn now to the effect of increasing the dye-to-basepair ratio of the DNA. The theoretical lines are calculated under the assumption that each dye molecule increases the contour length by 0.44 nm but leaves the Kuhn length and the effective width unchanged. This yields estimates of the mean and standard deviation that overestimate the observables at high ionic strengths and high dye loads. A simple explanation would be that the persistence length decreases slightly with increasing dye load, in agreement with Ref. [38] though not with Ref. [37]. Note that since experiments 2 and 3 were performed at low dye-to-basepair ratios, such a decrease would not significantly influence the interpretations of these experiments.

The results of experiment 2 are shown in panels **c**, **d**. Again, the experimental results are in qualitative agreement with the theoretical predictions. We see that the average extension agrees well with the theoretical prediction. However, the model predictions underestimate the standard deviation for the larger ionic strength, and for the largest channel at the lower ionic strength.

It is important to note that for experiment 2 we do not expect perfect agreement with Eqs. (2), since the condition $D_H \gg \ell_K$ is not satisfied, or only weakly satisfied. However, as long as $D_W \gg \ell_K$, the violation of the condition for D_H only affects the prefactors but not the power of D_W in Eqs. (2). This follows from the fact that a mapping to a one-dimensional model is possible also when $D_H \approx \ell_K$ [21]. In accordance with this prediction, the data points satisfying $D_W \gg \ell_K$ in panel **c** obey the scaling $\mu \propto D_W^{-1/3}$ of (Eq. 2a). Similarly the data points at 2×TBE which satisfy $D_W \gg \ell_K$ show a variance that

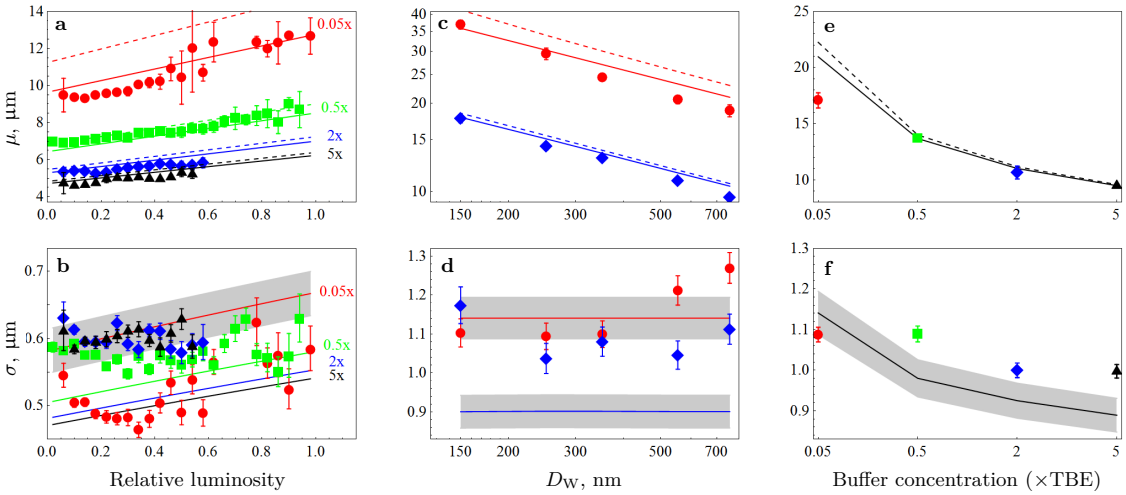

FIG. 2: Experimental results for 0.05×TBE (red ○), 0.5×TBE (green □), 2×TBE (blue ◊), and 5×TBE (black △) **a**, **b** Experiment 1. Mean and standard deviation of the extension of λ-DNA in a narrow nanochannel, as a function of relative luminosity. Theory [Eq. (2)], solid lines. The rigorous bounds on the prefactor in Eq. (2b) are indicated as a shaded region for 0.05×TBE, they are of the same order for the other cases. The corresponding uncertainty for the extension is much smaller and not shown. The dashed line shows theory corrected for wall repulsion (see text). **c**, **d** Experiment 2. Same, but for T4-DNA in a nanofunnel with varying width D_W. Note that panel **c** is a log-log plot. **e**, **f** Experiment 3. Same, but for T4-DNA in a wider square nanochannel, as a function of buffer concentration (x×TBE). Error bars correspond to 95% confidence intervals from the statistical analysis, the experimental uncertainty is not taken into account.

is approximately independent of D_W, in agreement with Eq. 2b, Fig. 2**d**. For the two rightmost data points at 0.05×TBE, the condition $D_W^2 \ll D_H \ell_K^2 / w_{eff}$ is violated. At this point the variance is expected to start increasing as D_W increases further [21], in perfect agreement with what is observed.

For square channels simulations [11, 15, 22] show that the mean extension increases more rapidly with decreasing $D = D_H = D_W$ than Eq. (2a) predicts, when $D \approx \ell_K$. The reason is that there is a tendency for the DNA molecule to align with the wall, and that the presence of the walls makes it more difficult for the molecule to change direction in the channel, forming a 'hairpin' [15, 27]. This can explain why the average extension appears to increase slightly faster with decreasing width than Eq. (2a) predicts, for the leftmost points in panel **c**. Such a trend has also been observed in previous measurements in rectangular channels [27, 40].

Now consider the standard deviation. The alignment and correlation effects mentioned above cause σ to be overestimated [27]. But when μ approaches the maximal extension L then fluctuations are suppressed [15]. These two effects could explain why, in experiments 1 and 2, σ is larger than predicted by theory at high ionic strengths but smaller for low ionic strengths and small channel sizes. It must also be noted that the standard deviation is difficult to estimate precisely, as it is not very much larger than the pixel size in the image (159 nm), and may depend on the assumptions entering into the statistical analysis (see Supplemental Material [33]). An additional source of uncertainty specific to experiment 2 is that D_W changes over the span of the molecule. For the most extended condition ($\mu = 37\,\mu$m), the channel width at either end of the molecule differs by approximately 25 nm from the stated width, measured at the center of the molecule.

The results of experiment 3 are shown in panels **e**, **f**. Here Eq. (1) is well satisfied. Simulations indicate [15, 22] that the alignment effects discussed above have little influence on μ, in square channels with $D \approx 3\ell_K$. Equally sensitive simulation results for σ have not been published, but simulations of the alignment effect [15, 27] indicate that Eq. (2b) underestimates σ by approximately 10%. We find that for the three largest ionic strengths, measurements are in excellent agreement with theory. The mean extension (panel **f**) agrees very well with the theoretical prediction of Eq. (2a), and Eq.(2b) underestimates the standard deviation (panel **f**) by about 10%, just as the measurements of alignment effects would suggest.

Intriguingly, the relation between measurements and predictions is different at 0.05×TBE than at high ionic strengths. Both mean and standard deviation are smaller than expected. This is particularly surprising considering that alignment effects should be even stronger at low ionic strength (where ℓ_K is larger). The discrepancy might indicate that the standard model does not describe

the physical parameters well at such low ionic strengths, possibly because the high relative concentration of BME significantly changes the buffer conditions, lowering the pH from ≈ 8.5 to ≈ 7.5. But we note that it may be hard to ensure uniform dye coverage under these conditions [32]. Also, BME is consumed as the experiment proceeds. This may change the ionic strength at small buffer concentrations, an effect our calculations do not include.

Conclusions. We have compared measurements of the extension of confined DNA to asymptotically exact predictions. First, we find very good agreement between experiments and theoretical predictions at high ionic strengths. A possible cause for deviations at low ionic strengths is that common estimates for w_{eff} and ℓ_{K} of stained DNA are too imprecise. Second, by measuring longer time series and more molecules in wider channels we expect to be able to precisely determine how ℓ_{K} and w_{eff} depend on the ionic strength of the solution. It may even be possible to determine ℓ_{K} and w_{eff} for single molecules. We note that w_{eff} is an effective parameter defined in terms of the excluded volume per Kuhn length, even for rigid rods w_{eff} is identical to the actual width only when $\ell_{\text{K}} \gg w_{\text{eff}}$. Finally the approach described here could be used to investigate DNA-wall interactions, a question about which little is known, and that is hard to describe theoretically.

Acknowledgements. This work was made possible through support from the Swedish Research Council (BM), the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine (BM), Chalmers Area of Advance (FW). We thank Charleston Noble and Erik Lagerstedt for helpful discussions.

After submission of this manuscript a paper [41] appeared online, which also compares experimental measurements on channel-confined DNA to the theoretical predictions of Ref. [20].

[1] W. Reisner, J. N. Pedersen, and R. H. Austin, Rep. Prog. Phys. **75**, 106601 (2012).
[2] J. J. Jones, J. R. C. van der Maarel, and P. S. Doyle, Phys. Rev. Lett. **110**, 068101 (2013).
[3] D. Gupta, J. Sheats, A. Muralidhar, J. J. Miller, D. E. Huang, S. Mahshid, K. D. Dorfman, and W. Reisner, J. Chem. Phys. **140**, 214901 (2014).
[4] A. Khorshid, P. Zimny, D. Tétreault-La Roche, G. Massarelli, T. Sakaue, and W. Reisner, Phys. Rev. Lett. **113**, 268104 (2014).
[5] M. Alizadehheidari, E. Werner, C. Noble, M. Reiter-Schad, L. K. Nyberg, J. Fritzsche, B. Mehlig, J. O. Tegenfeldt, T. Ambjörnsson, F. Persson, et al., Macromolecules **48**, 871 (2015).
[6] W. F. Reinhart, J. G. Reifenberger, D. Gupta, A. Muralidhar, J. Sheats, H. Cao, and K. D. Dorfman, J. Chem. Phys. **142**, 064902 (2015).
[7] E. Lam, A. Hastie, C. Lin, D. Ehrlich, S. K. Das, M. D. Austin, P. Deshpande, H. Cao, N. Nagarajan, M. Xiao, et al., Nature Biotechnol. **30**, 771 (2012).
[8] A. N. Nilsson, G. Emilsson, L. K. Nyberg, C. Noble, L. S. Stadler, J. Fritzsche, E. R. B. Moore, J. O. Tegenfeldt, T. Ambjörnsson, and F. Westerlund, Nucl. Acids Res. **86**, e188 (2014).
[9] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP press, 1994).
[10] T. Odijk, Phys. Rev. E **77**, 060901 (2008).
[11] Y. Wang, D. R. Tree, and K. D. Dorfman, Macromolecules **44**, 6594 (2011).
[12] D. R. Tree, Y. Wang, and K. D. Dorfman, Phys. Rev. Lett. **108**, 228105 (2012).
[13] L. Dai and P. S. Doyle, Macromolecules **46**, 6336 (2013).
[14] A. Muralidhar, D. R. Tree, Y. Wang, and K. D. Dorfman, J. Chem. Phys. **140**, 084905 (2014).
[15] A. Muralidhar, D. R. Tree, and K. D. Dorfman, Macromolecules **47**, 8446 (2014).
[16] P. J. Hagerman, Annual Review of Biophysics and Biophysical Chemistry **17**, 265 (1988).
[17] S. B. Smith, L. Finzi, and C. Bustamante, Science **258**, 1122 (1992).
[18] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, Biophysical Journal **72**, 1335 (1997), ISSN 0006-3495.
[19] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. Strick, S. M. Block, and V. Croquette, Biophysical Journal **76**, 409 (1999).
[20] E. Werner and B. Mehlig, Phys. Rev. E **90**, 062602 (2014).
[21] E. Werner and B. Mehlig, Phys. Rev. E **91**, 050601(R) (2015).
[22] L. Dai, J. van der Maarel, and P. S. Doyle, Macromolecules **47**, 2445 (2014).
[23] A. V. Dobrynin, Macromolecules **39**, 9519 (2006), ISSN 0024-9297.
[24] D. Stigter, Cell Biophys. **11**, 139 (1987).
[25] W. Reisner, J. P. Beech, N. B. Larsen, H. Flyvbjerg, A. Kristensen, and J. O. Tegenfeldt, Phys. Rev. Lett. **99**, 058302 (2007).
[26] C.-C. Hsieh, A. Balducci, and P. S. Doyle, Nano Letters **8**, 1683 (2008).
[27] E. Werner, F. Persson, F. Westerlund, J. O. Tegenfeldt, and B. Mehlig, Phys. Rev. E **86**, 041802 (2012).
[28] L. Onsager, Ann. N. Y. Acad. Sci. **51**, 627 (1949).
[29] R. van der Hofstad, F. den Hollander, and W. König, Probability Theory and Related Fields **125**, 483 (2003).
[30] R. R. Sinden, *DNA Structure and Function* (Elsevier, 2012).
[31] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, Proc. Natl. Acad. Sci. **94**, 6185 (1997).
[32] L. Nyberg, F. Persson, B. Åkerman, and F. Westerlund, Nucleic Acids Research p. gkt755 (2013).
[33] See Supplemental Material at [URL will be inserted by publisher] for brief discussions of the statistical analysis, and of the details of the experimental procedure.
[34] D. Stigter, Biopolymers **16**, 1435 (1977).
[35] J. A. Schellman and D. Stigter, Biopolymers **16**, 1415 (1977).
[36] F. Johansen and J. P. Jacobsen, Journal of Biomolecular Structure and Dynamics **16**, 205 (1998), ISSN 0739-1102.
[37] K. Günther, M. Mertig, and R. Seidel, Nucleic Acids Research **38**, 6526 (2010).

[38] C. U. Murade, V. Subramaniam, C. Otto, and M. L. Bennink, Nucleic Acids Research **38**, 3423 (2010).
[39] W. Reisner, K. Morton, R. Riehn, Y. Wang, Z. Yu, M. Rosen, J. Sturm, S. Chou, E. Frey, and R. Austin, Phys. Rev. Lett. **94**, 196101 (2005).
[40] F. Persson, P. Utko, W. Reisner, N. B. Larsen, and A. Kristensen, Nano Lett. **9**, 1382 (2009).
[41] D. Gupta, J. J. Miller, A. Muralidhar, S. Mahshid, W. Reisner, and K. D. Dorfman, ACS Macro Letters **4**, 759 (2015).

B

Supplementary materials

Supplemental material for

Extension of nano-confined DNA: quantitative comparison between experiment and theory

V. Iarko[1], E. Werner[1], L. K. Nyberg[2], V. Müller[2], J. Fritzsche[3], T. Ambjörnsson[4], J. P. Beech[5], J. O. Tegenfeldt[5,6], K. Mehlig[7], F. Westerlund[2], B. Mehlig[1]

[1] *Department of Physics, University of Gothenburg, Sweden*

[2] *Department of Biology and Biological Engineering, Chalmers University of Technology, Sweden*

[3] *Department of Applied Physics, Chalmers University of Technology, Sweden*

[4] *Department of Astronomy and Theoretical Physics, Lund University, Sweden*

[5] *Department of Physics, Division of Solid State Physics, Lund University, Sweden*

[6] *NanoLund, Lund University, Sweden and*

[7] *Department of Public Health and Community Medicine, University of Gothenburg, Sweden*

EXPERIMENTAL PROCEDURE

Channel manufacture. The channels were manufactured as described in Ref. [S1]. The nanochannels used in experiment 1 have a depth of $D_{\rm H} = 150$ nm. The channel cross-section is not perfectly rectangular, but rather trapezoidal, with a width at the top (bottom) of the channel of 130 nm (87 nm). We assume that these channels can be approximated by rectangular channels with a width $D_{\rm W} = 108$ nm, which is the arithmetic mean of the measurements at the top and bottom of the channel. The funnels used in experiment 2 have a depth of approximately $D_{\rm H} = 120$ nm, and $D_{\rm W}$ increases from 92 nm (top: 111 nm, bottom: 73 nm) at the narrow end to 815 nm (top: 830 nm, bottom: 800 nm) at the wide end – over a length of 500 µm. Finally, the channels used in experiment 3 have a depth $D_{\rm H} = 300$ nm, and a width $D_{\rm W} = 302$ nm (top: 330 nm, bottom: 275 nm). The dimensions are measured before the channel is closed by the lid. Since the bonding process changes the depth of the channel, our values for the depth of the channels have a relatively large uncertainty of about 10 nm.

Buffer preparation. The buffers were obtained by diluting 10×TBE tablets from Medicago to the desired concentration. 1×TBE contains 0.089 M Tris, 0.089 M Borate and 0.0020 M EDTA. Right before the experiments, 3% BME (3 μL of BME to 97 μL of sample solution) was added to prevent photo-nicking of the DNA. In addition to this the micro- and nanochannels in the chip were flushed with the right buffer concentration, containing 3% BME as well, to keep the same environment in the chip as in the sample solution.

Dye intercalation. Since intercalation of YOYO-dye molecules extends the contour length of the DNA it is important to estimate the dye load as accurately as possible. In experiment 1, the dye load was not equilibrated between molecules, resulting in heterogeneous staining [S2]. In experiments 2 and 3 the aim was instead to achieve homogeneous staining. To this end we first mixed the sample in high ionic strength (5×TBE), let it rest for at least 20 min, and then diluted to the desired ionic strength [S2].

DNA insertion. The nanofluidic chips consist of four loading wells connected two and two by microchannels. The microchannels are in turn connected by many nanochannels. Sample solution was inserted into one of the loading wells. Then the DNA was carried to the nanochannels by pressure-driven flow (N2 gas for T4 experiments and air for lambda experiments). The DNA was forced into the nanochannels by applying pressure over two wells that are connected by a microchannel. Once inserted in the channel, the pressure was switched off. Before imaging the DNA was allowed to relax for about 30 seconds to reach its equilibrium extension.

Video recording. Videos were recorded using a Photometrics Evolve$^{\rm TM}$ EMCCD camera. The image pixel size is 159.2 nm. For all measurements, the exposure time was 100 ms per photograph, but the delay between frames differed between experiments. For experiment 1, the delay between frames was 84 ms, yielding a frame-rate of 184 ms/frame. For experiments 2-3 with T4-DNA, the correlation time is significantly longer, so to minimise problems with photo bleaching and nicking, the frame-rate was decreased. For experiment 2, the frame-rate was 0.5 s/frame for the measurements at 0.05× TBE, and 2 s/frame for the ones at 2× TBE. For experiment 3, the frame-rate was 1 s/frame for all measurements.

Number of molecules. In experiment 1, 2388 molecules were analysed in total. Since the relative intensities are distributed unevenly, so are the number of molecules in each bin. This leads to very different error estimates for different bins in the estimation of the mean and standard deviation. We excluded bins with 2 molecules or less from the analysis. Further 3 molecules were excluded because the extension could not be succesfully extracted from all frames. In experiment 2, 9-10 molecules were analysed per data point under 0.05×TBE, and 7 molecules under 2×TBE. In experiment 3, 30 to 35 molecules were analysed per measured data point. One outlier at 0.5x TBE was removed, as its extension changed abruptly halfway through the data series.

DATA ANALYSIS

Extraction of DNA extension from kymographs. Consider a raw kymograph of the form provided in the main text (Fig. 1d). Dark regions in this image correspond to background, the bright regions correspond to fluorescence from DNA. The fluorescence intensities measured for both regions are subject to noise. This makes it difficult to reliably identify the end points of the bright regions in an automated fashion. In previous studies this was achieved by fitting the difference of two sigmoidal functions [S3] to each time frame. This method was used for analysing experiment 1 [S2]. In experiments 2-3 we used a computationally faster method that yields very similar results for the mean, and agrees to within about 5% for the standard deviation of the extension. The new method for detecting end positions relies on the assumption that the fluorescence intensities from DNA and from the background assume consistent and easily differentiated values, and proceeds as follows. First, each frame in the raw kymograph is smoothed using a moving average (we use an averaging window that is five pixels wide). Second, each frame is then segmented into binary low- and high-intensity regions using Otsu's method [S4, S5]. Third, we even out gaps between high-intensity regions that are equal to or shorter than five pixels. Fourth, the largest connected high-intensity component is found, and its edges are identified. The distance between these edges is the extension of the DNA molecule.

Statistical analysis. The experimental data consist of repeated measurements of the extension of individual DNA molecules at given values of parameters and are modeled using a random-coefficient model, a variation of a linear-regression model adapted to correlated data [S6]. Let y_{ij} denote the extension of molecule i ($i = 1, \ldots, n$) at time t_j ($j = 1, \ldots, 200$). The extension is modeled as a linear function of time as $y_{ij} = \mu + \beta t_j + u_i + e_{ij}$, where μ and β denote the overall mean value and slope. To account for the within-molecule correlation of data points the error term is split into two uncorrelated components, a molecule-specific correction to the mean, u_i, and an error term e_{ij} that measures the deviation of a data point from the molecule-specific regression line, $\mu + \beta t_j + u_i$. Both u_i and e_{ij} are assumed to be independently normally distributed with mean zero and variance r^2 and σ^2, respectively. The model describes the mean extension of molecules at a particular set of external parameters and the linear time dependence allows to include small unmeasured changes of external conditions, for instance drift into wider or narrower channel regions, or change of buffer concentration. The random intercept adjusts for the initial conditions at the start of measurement of each time series. A random intercept model is adequate as each set of molecules can be viewed as a random sample representative for all molecules at this particular set of external parameters. On average, the covariance of two arbitrary data points within molecules is equal to the variance of intercepts between molecules. The ratio of variances, $r^2/(r^2 + \sigma^2)$ is both a measure for the proportion of variation that is explained by differences between molecules, and the correlation of data points within a typical time series. By isolating the molecule-specific random variation due to unknown fluctuations of external parameters we are able to obtain an improved value for the residual variance of the extension (σ^2) that can be compared with the theory. Although the time trend was weak in most cases, we evaluated the mean value μ at $t = 100$, i.e. we replaced t_j with $(t_j - 100)$ in the above model. Results are given in terms of point estimates together with a 95% confidence interval (95% CI). The statistical analyses were performed by using SAS (version 9.4; SAS Institute, Cary, NC).

CALCULATION OF IONIC STRENGTH

The parameters w and ℓ_{K} of the DNA molecule depend on the ionic strength of the surrounding solution. The ionic strength is defined as

$$I_s = \frac{1}{2} \sum_i c_i z_i^2. \tag{S1}$$

Here, the sum runs over all ions in the solution, c_i is the concentration of ion species i, and z_i its valence. To compute the ionic strength, we must calculate the equilibrium concentrations c_i of all ions. Since there is some confusion in the literature (discussed below) about the ionic strength of TBE buffer, we document our calculation in some detail.

We follow the method outlined in Ref. [S7, Section 12-2]. Denote by $C[X]$ the total amount of all species of a substance X, in neutral and ionic form. At $N\times$TBE, the solution contains $C[\mathrm{T}] = N \times 0.089$ M Tris, $C[\mathrm{Bo}] = N \times 0.089$ M Borate, $C[\mathrm{E}] = N \times 0.0020$ M EDTA, and $C[\beta] = 0.429$ M BME. The dissociation and association constants that determine the chemical equilibrium are Tris: $\mathrm{p}K_b = 5.94$; Borate: $\mathrm{p}K_a = 9.24$, EDTA: $\mathrm{p}K_a = \{1.99, 2.67, 6.16, 10.26\}$; BME: $\mathrm{p}K_a = 9.6$. At the resulting pH-value of approximately 8.5, most of the EDTA is triply ionised, and we can safely ignore the minute concentrations of neutral and singly ionised EDTA. We denote the molar concentration of a species by [X], and its activity coefficient by γ_{X}. The activity coefficients are assumed to be given by the Davies equation, as stated in Ref. [S7] (the empirical prefactors of this equation differ between sources. For our ionic strength calculations, the difference between different formulations makes a difference of two percent at most). The system

of equations that must be solved is

$$\frac{[\mathrm{TH^+}]\gamma_{\mathrm{TH^+}}[\mathrm{OH^-}]\gamma_{\mathrm{OH^-}}}{[\mathrm{T}]\gamma_{\mathrm{T}}} = 10^{-5.94} \quad \text{(S2)}$$

$$\frac{[\mathrm{Bo^-}]\gamma_{\mathrm{Bo^-}}[\mathrm{H^+}]\gamma_{\mathrm{H^+}}}{[\mathrm{HBo}]\gamma_{\mathrm{HBo}}} = 10^{-9.24} \quad \text{(S3)}$$

$$\frac{[\mathrm{HE^{3-}}]\gamma_{\mathrm{HE^{3-}}}[\mathrm{H^+}]\gamma_{\mathrm{H^+}}}{[\mathrm{H_2E^{2-}}]\gamma_{\mathrm{H_2E^{2-}}}} = 10^{-6.16} \quad \text{(S4)}$$

$$\frac{[\mathrm{E^{4-}}]\gamma_{\mathrm{E^{4-}}}[\mathrm{H^+}]\gamma_{\mathrm{H^+}}}{[\mathrm{HE^{3-}}]\gamma_{\mathrm{HE^{3-}}}} = 10^{-10.26} \quad \text{(S5)}$$

$$\frac{[\beta^-]\gamma_{\beta^-}[\mathrm{H^+}]\gamma_{\mathrm{H^+}}}{[\mathrm{H}\beta]\gamma_{\mathrm{H}\beta}} = 10^{-9.6} \quad \text{(S6)}$$

$$[\mathrm{H^+}]\gamma_{\mathrm{H^+}}[\mathrm{OH^-}]\gamma_{\mathrm{OH^-}} = 10^{-14.0} \quad \text{(S7)}$$

$$[\mathrm{T}] + [\mathrm{TH^+}] = C[\mathrm{T}] \quad \text{(S8)}$$

$$[\mathrm{HBo}] + [\mathrm{Bo^-}] = C[\mathrm{Bo}] \quad \text{(S9)}$$

$$[\mathrm{H_2E^{2-}}] + [\mathrm{HE^{3-}}] + [\mathrm{E^{4-}}] = C[\mathrm{E}] \quad \text{(S10)}$$

$$[\mathrm{H}\beta] + [\beta^-] = C[\beta] \quad \text{(S11)}$$

$$[\mathrm{TH^+}] + [\mathrm{H^+}] = \quad \text{(S12)}$$
$$[\mathrm{OH^-}] + [\mathrm{Bo^-}] + 2[\mathrm{H_2E^{2-}}] + 3[\mathrm{HE^{3-}}] + 4[\mathrm{E^{4-}}] + [\beta^-]$$

Eqs. (S2)-(S6) are the equilibrium conditions from the law of mass action, Eqs. (S7)-(S11) ensure that the total concentration of a substance is conserved, and Eq. (S12) ensures charge neutrality. The activity coefficients γ_{X} depend on the ionic strength I_s according to the Davies equation [S7],

$$\log_{10}\gamma_{\mathrm{X}} = -0.51 z_{\mathrm{X}}^2 \left(\frac{\sqrt{I_s}}{1+\sqrt{I_s}} - 0.3 I_s \right). \quad \text{(S13)}$$

Here, z_{X} is the valence of species X, and I_s is measured in units of M. Finally, the ionic strength itself is given according to Eq. S1 by

$$I_s = 1/2\Big([\mathrm{TH^+}] + [\mathrm{H^+}] + [\mathrm{OH^-}] + [\mathrm{Bo^-}] + \quad \text{(S14)}$$
$$4[\mathrm{H_2E^{2-}}] + 9[\mathrm{HE^{3-}}] + 16[\mathrm{E^{4-}}] + [\beta^-] \Big).$$

We solve the system of equations (S2)-(S14) iteratively. Starting from the initial guess $\gamma_{\mathrm{X}} = 1$ for all species, Eqs. (S2)-(S12) are solved with the Mathematica routine `FindInstance`. Plugging the resulting concentrations into Eqs. (S13)-(S14) yields new values for γ_{X}, which are then used in Eqs. (S2)-(S12). This process was repeated until the ionic strength converged.

We tested our calculations by comparing with Table 1 of Ref. [S8]. We reproduce the reported ionic strengths for 0% BME and 0.5% BME to within one percent.

[S1] F. Persson and J. O. Tegenfeldt, Chemical Society Reviews **39**, 985 (2010), ISSN 1460-4744.

[S2] L. Nyberg, F. Persson, B. Åkerman, and F. Westerlund, Nucleic Acids Research p. gkt755 (2013).

[S3] W. Reisner, J. N. Pedersen, and R. H. Austin, Rep. Prog. Phys. **75**, 106601 (2012).

[S4] N. Otsu, IEEE Transactions on Systems, Man and Cybernetics **9**, 62 (1979), ISSN 0018-9472.

[S5] P.-S. Liao, T.-S. Chen, and P.-C. Chung, J. Inf. Sci. Eng. **17**, 713 (2001).

[S6] T. Snijders and R. Bosker, *Multilevel Analysis* (Sage Publications Ltd. London, 1999).

[S7] D. C. Harris, *Quantitative chemical analysis* (W.H. Freeman and Co, New York, 2010), 8th ed., ISBN 9781429218153.

[S8] C.-C. Hsieh, A. Balducci, and P. S. Doyle, Nano Letters **8**, 1683 (2008).