Detection of chain backfolding in simulation of DNA in nanofluidic channels

DATASET *in* SOFT MATTER · AUGUST 2012

Impact Factor: 4.03 · DOI: 10.1039/c2sm26128f

2 AUTHORS, INCLUDING:

Peter Cifra
Slovak Academy of Sciences
96 PUBLICATIONS **1,027** CITATIONS

Soft Matter

Dynamic Article Links

Cite this: *Soft Matter*, 2012, **8**, 9022

www.rsc.org/softmatter

PAPER

Detection of chain backfolding in simulation of DNA in nanofluidic channels

Peter Cifra* and Tomáš Bleha

Received 16th May 2012, Accepted 5th July 2012
DOI: 10.1039/c2sm26128f

The DNA extension in cylindrical nanochannels under moderately strong confinements in the "transition" region, where experiments are conducted, is examined by Monte Carlo simulations. The focal point is estimation of the amount of backfolding structures such as hairpins and loops in extended (linearized) DNA molecules. At the chain-ends of DNA an extensive folding (mainly as the J-type hairpins) is detected under all confinements, covering the Odijk and de Gennes regimes and the transition region between them. In contrast, in the DNA chain interior, the backfolding into Z-type hairpins is abundant in the de Gennes regime, significantly reduced in the transition region, and practically absent in the Odijk regime. The linear relationship between the DNA extension and its contour length is validated also in the transition region where explicit theories are absent.

I. Introduction

With recent nanotechnology advances single DNA molecules can be manipulated and analyzed in new ways. Stretching (linearizing) of double-stranded (ds)DNA in nanofabricated channels has emerged as an innovative technique for fragment length analysis and genome mapping.[1–8] Confinement-induced stretching of dsDNA is accomplished in fluidic channels of the dimensions much smaller than the radius of gyration R_g of the DNA molecule in free solution. Combination of nanofluidic devices fabricated by chip lithography with fluorescence microscopy allows direct experimental visualization of single DNA molecules.

It can be expected that in nanochannels the chain dimensions of DNA molecules such as the mean span R_{sp} or the end-to-end distance R will depart significantly from their values in bulk. Confining and stretching of DNA molecules inside nanochannels have provided evidence that DNA molecules can be extended to a substantial fraction of their contour length L.[9–13] In experiments, especially DNA extension as a function of the channel width and ionic strength has been explored. Interpretation of experimental data on confined DNA molecules is assisted by the polymer physics theories and coarse-grained computer simulations. The extension of nanoconfined DNA molecules is described by two current theories of polymer confinement: by the Odijk theory[14,15] at strong confinement and by the scaling (blob) theory at moderate confinement.[16–18] The Odijk deflection theory[14,15] is based on the interplay of confinement and chain elasticity in narrow channels. The undulated chain extended along the channel axis can be viewed as a sequence of deflection segments of a characteristic length λ (Fig. 1). In the scaling theory of de Gennes a moderately confined polymer is represented by a one-dimensional sequence of spherical blobs of the size of the channel diameter[16,17] (Fig. 1). It is assumed that within each blob the chain segments have the same structure as in the bulk solution and the monomer density along the channel is uniform. The blob theory was generalized[17,18] for semiflexible polymers such as DNA on condition that the channel width D is much greater than the persistence length P of the polymer.

Computer simulations using the coarse-grained mesoscopic models offer a powerful option to experimental and theoretical studies of confined DNA. Monte Carlo (MC) simulations based on a non-charged worm-like chain model have been employed to compute the chain extension R of DNA as a function of channel dimension D at high salt concentrations.[19,20] The chain extension profiles $R(D)$ computed for a rectangular channel, tube and slit displayed three sectors, in a qualitative agreement with nanofluidic measurements of DNA. Another simulation[21] using a slightly different model confirmed the existence of transitional regimes in between the Odijk and de Gennes (blob) limits predicted earlier[22] by the scaling analysis. For the wormlike chain model the details of polymer behavior in the Odijk regime were determined[23] by a special simulation algorithm. Recently, simulations of nanoconfined DNA involving explicit electrostatic interactions between charged segments were reported.[8] The ionic strength dependence of DNA extension from simulations[8] showed a reasonable agreement with experimental findings and Odijk theory.

Nanochannel-induced DNA elongation is now being developed along the route to robust analytical devices for biological bioassays. In the ideal case of narrow nanochannels, of the diameter slightly larger than the diameter of DNA molecule, macromolecules would stretch up to their full contour length L. However, the very difficult loading of DNA into such narrow channels presents a major handicap for application in high-

Polymer Institute, Slovak Academy of Sciences, 845 41 Bratislava, Slovakia. E-mail: cifra@savba.sk

throughput techniques. Hence, the channel dimensions much exceeding the DNA diameter have to be used in practice. As a result, instead of a fully stretched molecule the confined DNA chains may involve various folds of chain on themselves, loops, knots and other kinks.

It is thus imperative to explore the extent of backfolding in the confined DNA molecule as a function of the channel dimension. In this paper we present the simulation data on the DNA extension in a cylindrical channel, pertaining to the high salt concentration where electrostatic interactions are routinely neglected. Three regions, deflection, transition and blob regime, were identified at the DNA extension profiles as a function of the channel width. For all the three regions the extent of backfolding structures such as hairpins and loops were estimated using a number of quantities from the simulations. A striking difference in the DNA folding at the chain ends and in the molecule interior was found.

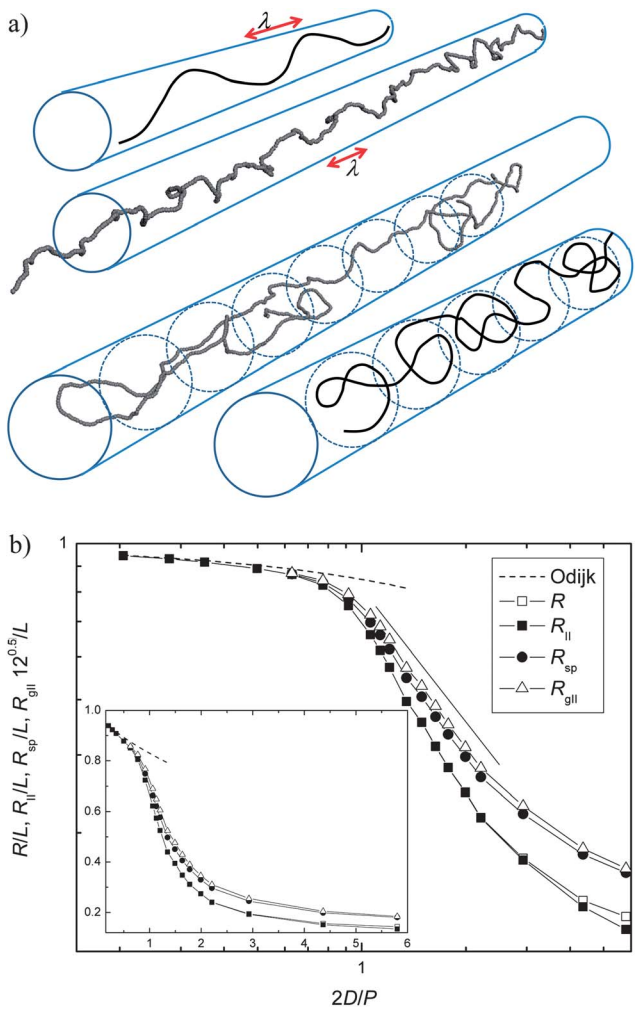

Fig. 1 (a) Customary sketches of chains in the Odijk and de Gennes regimes together with the respective simulation snapshots (in the middle) of a confined DNA molecule of $N = 800$ for $2D/P = 0.77$ and 1.49. (b) Logarithmic plot of various measures (see text) of the relative extension of DNA of the length $N = 1000$ as a function of the channel diameter $2D$ normalized by the DNA persistence length P. The straight line indicates the slope -1 and the dashed line represents eqn (1). A corresponding linear plot is shown in the inset.

II. Model and method

The chain extension of dsDNA in channel was computed by a bead–spring model of wormlike chains described previously.[19,20,24,25] The chain consists of N beads of diameter d connected by $N - 1$ effective bonds of the average length $\langle l \rangle$ with a corresponding contour length $L = (N - 1)\langle l \rangle$. The potential energy of the model consists of three contributions: the bond stretching potential U_{FENE}, the bending energy potential U_{b} and the repulsive non-bonded potential U_{nb}. The finitely extensible, nonlinear elastic potential U_{FENE} between the consecutive bonded beads is given as

$$\frac{U_{\mathrm{FENE}}(l)}{k_{\mathrm{B}}T} = -\kappa(l_{\max} - l_{\mathrm{o}})^2 \ln\left[1 - \left(\frac{l - l_{\mathrm{o}}}{l_{\max} - l_{\mathrm{o}}}\right)^2\right]$$

The bond vibrations were considered to be temperature independent with the spring constant $\kappa = 20$. The bond length was restricted to the preferred value $l_{\mathrm{o}} = 0.7$ and allowed to oscillate by the increment $l_{\max} - l_{\mathrm{o}} = 0.3$. The maximum allowable bond distance $l_{\max} = 1$ was identified with the length unit. The fully repulsive cut and shifted Morse potential U_{nb} accounts for excluded volume interaction between nonbonded beads and is given as

$$U_{\mathrm{nb}}(r)/\varepsilon = \exp[-2\alpha(r - r_{\min})] - 2\exp[-\alpha(r - r_{\min})] + 1$$

for $r < r_{\min}$ and zero otherwise. The energy at minimum ε, the distance at minimum $r_{\min}$ and the steepness of the potential energy function α were set to 1 $k_{\mathrm{B}}T$, 0.8 and 24, respectively. This combination corresponds to good solvent conditions for the chain. Since the bead core diameter exceeds the bond length, the approach used resembles the model of partially fused spheres. The U_{b} term gives the penalty for the deviation of a chain from the straight rod. The bending energy parameter b is proportional[19] to the persistence length P, $b = P/\langle l \rangle$, of the free dsDNA molecule.

$$U_{\mathrm{b}}(\theta)/k_{\mathrm{B}}T = b(1 + \cos\theta)$$

Here θ is the angle between two consecutive bonds in chain. The bending stiffening of a free uncharged chain is characterized by the intrinsic ("bare") persistence length P. In polyelectrolytes such as DNA an additional, electrostatic term P_{elst} contributes to the total bending persistence length. The influence of electrostatic interactions on the persistence length of a polyelectrolyte diminishes with increasing ionic strength. To eliminate the electrostatic term P_{elst} an environment of high salt concentrations similar to physiological conditions was assumed. Hence, a consensus value of DNA intrinsic stiffness $P = 50$ nm was used in simulations.

The average bond length $\langle l \rangle$ was set equal to the effective diameter $d = 2.5$ nm of the dsDNA helix. With this choice one bead in the model represents 7.4 base pairs (bp). Accordingly, the reduced persistence length $P/d = 20$, *i.e.* the correlation of bond vectors persists on the distance of about 20 beads. The selected value $d = 2.5$ represents a geometric measure of DNA chain thickness. Alternatively, the larger value of the helix diameter

d can be considered to account for the electric charge on the phosphate group. However, as shown[26] for the model of thick polymer tube, the variables d and P are interconnected; any increase in the DNA diameter implicitly enhances the persistence length.

Hard-wall confinement of DNA chains in cylindrical channels is considered in simulations. The radius of cylinder D is an effective dimension involving the correction to the accessible volume of the bead centers.[20] Periodic boundary conditions along the channel axis were applied. The DNA chain lengths N up to 2000 beads were considered. The chain length $N = 1000$ (7400 bp) in our model corresponds to the contour length L of dsDNA of about 2.5 μm. The Metropolis MC method with reptation chain updates of up to 6×10^7 MC cycles of trial reptation moves was used in the chain generation. In initial tests we included also bead displacement moves but since this did not change the observed properties we used further only the reptation moves. The error bars of data in figures showing ensemble averages are in the range of the size of symbols.

III. Results and discussion

Several quantities are commonly used to assess the DNA extension in a channel. The mean span R_{sp} of DNA molecules along the channel axis is of the primary importance since it corresponds to fluorescence micrograph images captured in DNA nanochannel experiments. The chain span R_{sp} is defined as the average of the difference between the maximum and minimum distances of chain segments along the channel axis in all accessible conformations. Alternatively, the chain size expressed by the end-to-end distance R or its longitudinal component $R_{\|}$ parallel to the centerline of the channel is used in the theory and simulations of DNA nanoconfinement.

Simulations of DNA stretching in cylindrical channels of the diameter $2D$ reduced by the DNA persistence length P are shown in Fig. 1. It is seen that plots of the chain size normalized by the contour length L almost coincide for the end-to-end parameter R and the axial component $R_{\|}$; only at $2D/P > 3$ their slight divergence can be noted. Importantly, except for very narrow channels, the mean span of confined DNA molecules is larger than the respective values of R and $R_{\|}$. The difference grows as the channel dimension increases; for example, R_{sp}/L is by about 20% larger than R/L at $2D/P = 3$. The data for the longitudinal component $R_{g\|}$ of the radius of gyration R_g of confined DNA are also included in Fig. 1. In this case the curve is reduced by the maximum achievable extension $L/\sqrt{12}$. In bulk the mean span dimension should be close to the diameter of an equivalent sphere, $2R_g$.

The experimental and simulation data on DNA extension in nanochannels as a function of the channel width are usually analyzed by two theoretical concepts valid at strong and moderate confinements, the deflection chain model and blob model, respectively. In the Odijk deflection model[14,15] the chain is extended along the channel axis with undulations having a characteristic deflection length $\lambda \approx (2D)^{2/3}(P)^{1/3}$ between successive deflections from the opposite channel walls (Fig. 1). The DNA longitudinal extension in the Odijk regime is given as

$$R_{\|} = L[1 - A(2D/P)^{2/3}] \quad (1)$$

where A is a prefactor depending on the channel profile; $A = 0.170$ was derived[23] for the cylindrical channels. Since the channel width is smaller than the persistence length of the DNA chain, any backfolding (hairpin formation) of DNA in the deflection regime should be energetically very unfavorable.

At moderate confinement, where channel width is a multiple of the persistence length, the longitudinal extension $R_{\|}$ of DNA is described by the blob theory.[16–18] In this theory a "cigar-like" conformation of the chain in a cylinder is visualized as a linear sequence of mutually excluding blobs (Fig. 1). The chains inside blobs follow the excluded-volume statistics. When the effective width d of a DNA molecule is explicitly considered, the scaling relation for DNA extension reads[18]

$$R_{\|} \cong L(dP)^{1/3}D^{-2/3} \quad (2)$$

In stiff chains such as DNA, long chains are needed to reach the excluded-volume statistics required by the blob model. The relevant scaling-theory condition[17] $N \gg (P/d)^3$ gives for our selection of parameters $N \gg 8000$. Hence, for short DNA molecules below this limit, a sub-regime has to exist where chains behave more or less ideal at moderate confinement. For such a model of ideal-chain blobs the longitudinal extension can be written as

$$R_{\|} \cong NPd/D = LP/D \quad (3)$$

The coil size for an ideal semiflexible chain is $R^2 = NdP$. One blob of the size of the channel width D consisting of g segments thus becomes $D^2 = gdP$ and the size of the linear chain of N/g blobs in the channel yields $R_{\|} = (N/g)D = NPd/D$.

The sectors relevant to two outlined theories of confined DNA can be identified at the stretching profiles in Fig. 1. The first few points at strong confinements, up to $2D/P = 0.7$, are located in the Odijk regime. Here the mean span dimension R_{sp} along the channel direction is nearly the same as the end-to-end distance R. The central sector of the stretching profiles in Fig. 1 corresponds to the ideal-chain blob model. The data for the reduced chain span R_{sp}/L and R_g/L agree very well with the power-law dependence, eqn (3), with the exponent -1 and are also in a fair accord with the exponent -0.85 determined in the single-molecule experiments.[3] On the other hand, the data for R and $R_{\|}$ yield in Fig. 1 a slope slightly more negative than -1. Apparently, the concerted behavior of R_{sp} and R_g quantities stems from their averaging through all segments over the chain, while the end-to-end distance quantities refer to the averaging of just the chain ends.

Between the deflection and blob regimes on the DNA stretching profiles a transition region occurs (Fig. 1) in the range of $2D/P$ from about 0.7 up to about 1.3. This is the region of the major interest for DNA extension experiments. However, the appropriate theoretical framework for this region is still largely missing. On going from the deflection to the blob regime a substantial transformation of the shape of DNA molecules takes place: from a smooth sinusoidal structure of a high orientation order imposed by the nearby channel walls to an elongated structure of sizeable degree of disorder in individual random-coil sub-elements (see sketches and snapshots in Fig. 1). At the crossover to the transition region (*i.e.* for the channel widths

above 35 nm) various backfoldings and loops become feasible in the linearized DNA structure. Some extra scaling regimes of DNA extension, additional to the Odijk and blob ones, were envisioned in the transition region from theory[22] and simulations.[21] Such regimes may arise for example when the chain backfolding (formation of hairpins) is allowed in the deflection model of the confined DNA molecule.

The distribution functions of the end-to-end distance $W(R)$ of the DNA molecule in cylindrical channels (Fig. 2) provide evidence of the shape change of DNA molecules in the transition region. The dominant peaks in the $W(R)$ functions correspond to the extended DNA molecules. By increasing the channel width the distribution curves widen and their maxima move to the lower values of R/L. Two rightmost narrow curves depict the distribution in the deflection regime where a vast majority of confined DNA is highly extended, up to about 90% of the contour length. In these two curves shorter DNA conformations are by about two orders of magnitude less abundant and most probably arise by folding of the chain ends into the J-type hairpins.

By channel widening to the transition region the extension of DNA exclusively by the deflection mode is replaced by some mixed mechanism. Correspondingly, the intermediate chain extensions R/L become widely populated in distribution functions $W(R)$ in Fig. 2. For channel widths slightly above $2D/P = 0.7$ the intermediate extensions of DNA molecules are presumably realized by the conformations folded preferably at the chain ends, such as in the J-type hairpins and loops. Occasionally, the local Z-type hairpins may arise in the chain interior. These comprise the two coupled U-type hairpins with the sharp turns and prolonged stem alignment.

By further opening of the channel lumen the chains are getting shorter due to hairpins and loops and the amount of DNA molecules of intermediate extensions increases (Fig. 2). Alongside, by channel widening, the hairpins with sharp turns are progressively replaced by smooth loops of large curvatures. Finally, two distribution functions $W(R)$ with the uppermost values of $2D/P$ in Fig. 2 cover the blob regime of DNA extension. Here, the high elongation peak indicative of the deflecting structure is completely missing and a broad distribution of DNA extension, typical for coiled structures, is present. Interestingly, a minor peak can be seen at small values of R/L in Fig. 2. This peak most likely corresponds to some average of many quasi-symmetrical U-form hairpin structures for which $R \sim 0$.

In shorter DNA molecules ($N \sim 200$) the hairpin-related bimodality of the $W(R)$ function[19] as well as of the related distribution functions of the radius of gyration $W(R_g)$[20] turned out to be very pronounced. One can note that the $W(R)$ function determined from the measurements of the thermal fluctuation of the internal segments of DNA in channels about 50–100 nm wide also showed two peaks.[13] From such a bimodality the existence of the local folded structures in confined DNA, analogous to those discussed above, was inferred[13] in the transition region. The folding states of DNA were also detected[27] in nanocapillaries of diameters down to 45 nm from multiple peaks in the resistive-pulse measurements.

Simulations suggest that the chain-end regions of confined DNA, due to their enhanced flexibility, are most prone to the local folding events such as the J-type or U-type hairpins and loops. Such behavior is well established from intensity profile analysis of DNA images.[4,6,8,10] It also allows us to explain the higher values of the DNA span R_{sp} relative to the end-to-end distance R or its longitudinal component $R_{\|}$ in Fig. 1. At $2D/P > 1$ the latter two values are significantly reduced by local end-folding. However, in narrow channels the end-folding is suppressed and all the three quantities are identical in the Odijk regime (Fig. 1).

The relatively stable structures involving backfolding and loops in nanoconfined DNA chains can be directly observed in the simulation traces during the time evolution of the chain extension. Fig. 3 illustrates the typical courses of the fractional extension of DNA molecules as a function of time at $2D/P = 0.92$ in the transition regime. The extension fluctuates most of the time around the value R/L at roughly 0.8, symptomatic of a fairly extended and undulated DNA molecule. However, the distinctive events of transitions to the shorter folded structures are clearly seen on the traces. Simulation snapshots for individual events provide an intuitive identification of changes of DNA shape at the transitions. Notably, even the formation of the U-type hairpins and loops with R close to zero can be detected.

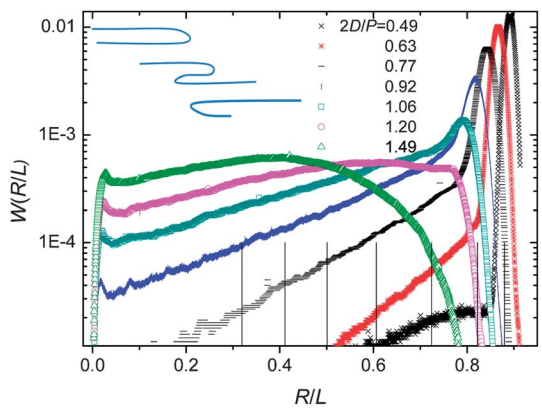

Fig. 2 Semilogarithmic plot of the distribution function $W(R/L)$ of the normalized end-to-end distance of a DNA chain of $N = 800$ confined in the cylinder of the reduced diameter $2D/P$ shown in the legend together with the sketch of U, Z and J type hairpin structures.

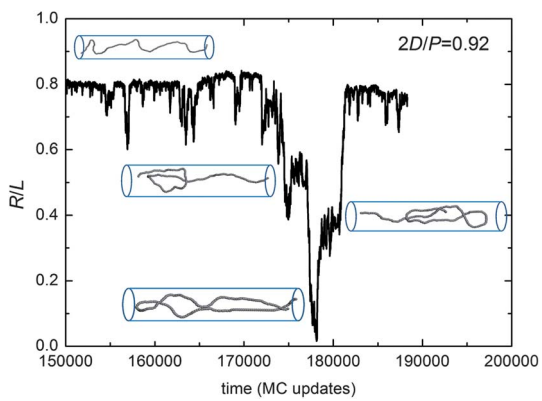

Fig. 3 Monte Carlo traces of the relative extension of DNA of $N = 800$ in a channel of the width $2D/P = 0.92$ as a function of time. Simulation snapshots assigned to individual transition events are presented.

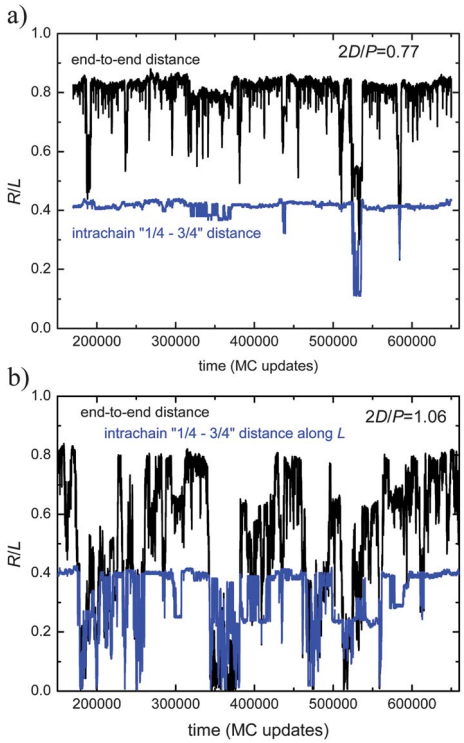

Fig. 4 Monte Carlo traces of the relative extension of DNA of $N = 800$ in a channel of the width (a) $2D/P = 0.77$ and (b) $2D/P = 1.06$ as a function of time for the whole DNA molecule (upper traces) and for the internal segments only (bottom traces).

To separate the influence of the chain end segments, the fluctuation of the interior portion of confined DNA located between 0.25 and 0.75 fractions of the contour length L was sampled separately in simulation (Fig. 4). For the channel width $2D/P = 0.77$ near to the onset of the transition region the fluctuation of interior segments is radically suppressed in comparison with fluctuations of the whole chain (Fig. 4a). The local transformations of the extended (in essence deflecting) chains to the folded structures in the chain interior such as Z-hairpins are just sporadic. The difference between two traces in Fig. 4a provides an additional confirmation that the folding events in the DNA molecule are located preferentially at the chain ends.

In contrast, in slightly wider channels of $2D/P = 1.06$, about in the middle of the transition region, the interior fluctuations are still less pronounced than the whole chain fluctuations, nevertheless both kinds of fluctuations are considerably enhanced (Fig. 4b). Evidently, going from the Odijk regime, the onset of the transition region is signaled by an appearance of the folded events in chain interior. The MC traces in Fig. 4b suggest that in channels of dimensions matching the transition region the amount of folded structures in DNA interior can be substantial. An analogous quantity from DNA nanochannel experiments, the time traces of fluorescent intensities, revealed as well the significant DNA folding in the chain interior[13] or at chain ends.[10] By modeling the folded structures by the J-type hairpins and loops it was found[10] that the folded structures in channels are about 30% more extended (per unit length) than the unfolded region of DNA. Such an enhancement was ascribed to the excluded volume interactions since the monomer density of the folded portions increased relative to the unfolded ones. A similar increase of excluded volume was reported recently for rings confined in cylinders.[28]

The chain backfolding in individual DNA conformations can be detected in simulations from the evaluation of the associated bond vectors. The abundance of various types of folded structures can be estimated from quantities related to the projection persistence length.[29] One of the principal definitions[29] of the persistence length employs the projection of the end-to-end vector $\boldsymbol{R}$ in the direction of the reference bond vector $\boldsymbol{l}_{\mathrm{ref}}$

$$P = P_{\mathrm{proj}} = \langle \boldsymbol{R}\cdot\boldsymbol{l}_{\mathrm{ref}}\rangle / \langle l\rangle = \sum_{i=1}^{N-1}\langle \boldsymbol{l}_i\boldsymbol{l}_{\mathrm{ref}}\rangle \Big/ \langle l\rangle = \langle l\rangle \sum_{i=1}^{N-1}\langle \cos\,\theta_{\mathrm{ref},i}\rangle \tag{4}$$

where $\boldsymbol{l}_i$ is the i^{th} bond vector and $\theta_{\mathrm{ref},i}$ stands for an angle between the reference and i^{th} bond vectors. The first or the middle unit bond vectors in the chain, denoted as $\boldsymbol{e}$ and $\boldsymbol{m}$, respectively, are most often taken as the reference $\boldsymbol{l}_{\mathrm{ref}}/\langle l\rangle$ (Fig. 5a). It was shown recently[30] that the middle chain option gives a better measure of the intrinsic persistence length since such a choice is not affected by an enhanced flexibility of chain ends.

The persistence length according to eqn (4) is an *average* of the scalar product of the vector $\boldsymbol{R}$ with either of the reference bond vectors $\boldsymbol{e}$ or $\boldsymbol{m}$ over all sampled conformations. The values of such a scalar product for individual conformations ("instantaneous" values) give the distribution functions $Q(\boldsymbol{R}\cdot\boldsymbol{e})$ or $Q(\boldsymbol{R}\cdot\boldsymbol{m})$. These functions will be used to assess the amount of backfolding and loops in nanoconfined DNA molecules.

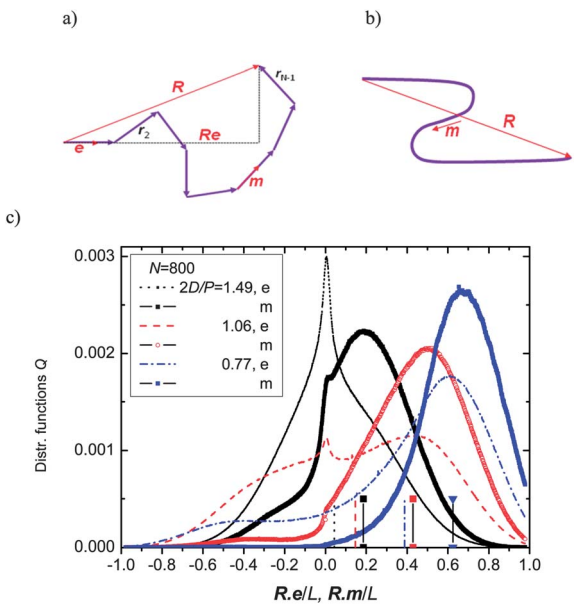

Fig. 5 (a) Sketch of a chain with the reference bond vectors $\boldsymbol{e}$ (chain end) and $\boldsymbol{m}$ (middle of chain), (b) orientation of vectors in the Z-hairpin structure, and (c) the distribution functions of the scalar products $Q(\boldsymbol{R}\cdot\boldsymbol{e})$ and $Q(\boldsymbol{R}\cdot\boldsymbol{m})$, dashed and full lines respectively, calculated for DNA molecules of $N = 800$ in nanochannels of the width $2D/P$ given in the legend. The mean values of the distribution functions (the projection persistence lengths) are shown by the vertical lines and respective symbols.

The distribution functions of the scalar products $Q(\boldsymbol{R}\cdot\boldsymbol{e})$ and $Q(\boldsymbol{R}\cdot\boldsymbol{m})$ calculated for DNA molecules in nanochannels are shown in Fig. 5c. Two channel widths considered in Fig. 5c cover the transition region and the third one at $2D/P = 1.49$ (about 75 nm) is situated in the blob regime. The function $Q(\boldsymbol{R}\cdot\boldsymbol{e})$ allows us to monitor the orientation of the terminal bond with respect to the vector $\boldsymbol{R}$. It is sensible to assume that in narrow channels the orientation of the vector $\boldsymbol{R}$ generally follows the centerline of the cylindrical channel. Then, the negative values of the product $\boldsymbol{R}\cdot\boldsymbol{e}$ in Fig. 5c suggest that the end portion of the DNA molecule is oriented opposite ($\Theta < 0$) to the direction of the remaining part of the molecule specified by the vector $\boldsymbol{R}$. Such a situation is realized for example in the case of J-hairpins. In moderately wide channels of $2D/P = 1.49$ the abundance of structures with negative and positive scalar products $\boldsymbol{R}\cdot\boldsymbol{e}$ is roughly the same. However, by reducing the channel diameter, the zone of the negative product $\boldsymbol{R}\cdot\boldsymbol{e}$ and the fraction of the folded structures rapidly diminish.

Analogous reasoning can be applied to the function $Q(\boldsymbol{R}\cdot\boldsymbol{m})$ referring to the interior bond in the middle of the chain. In this case the variable $\boldsymbol{R}\cdot\boldsymbol{m}$ is negative ($\Theta < 0$) for example in conformations resembling the Z-type hairpins (Fig. 5b). As expected, in the blob regime at $2D/P = 1.49$ a considerable amount of DNA chain loops satisfying the condition $\Theta < 0$ can be deduced from Fig. 5c. However, for channels in the transition region, just a slight indication of interior folding can be seen at $2D/P = 1.06$ and practically no interior backfolding is detected in the narrow channel of $2D/P = 0.77$.

An intriguing feature in Fig. 5c is the sharp peaks at the distribution function $Q(\boldsymbol{R}\cdot\boldsymbol{e})$ when the abscissa variable $\boldsymbol{R}\cdot\boldsymbol{e}$ is close to zero. Likewise, there are discontinuities at the distribution function $Q(\boldsymbol{R}\cdot\boldsymbol{m})$ in this area. Such a situation can be realized by two types of conformations: either the vectors $\boldsymbol{R}$ and $\boldsymbol{e}$ (or $\boldsymbol{m}$) are in an approximately perpendicular position, or solely the vector $\boldsymbol{R}$ is about zero, as in the case of the U-type of hairpins. The population of such conformations increases by raising the channel diameter and even prevails at $2D/P = 1.49$ in the blob regime.

The fractions of the folded structures in nanoconfined DNA, estimated from the area below the functions $Q(\boldsymbol{R}\cdot\boldsymbol{m})$ and $Q(\boldsymbol{R}\cdot\boldsymbol{e})$ in Fig. 5c in the negative zone of the abscissa relative to the total integral of the respective functions are listed in Table 1. The second column, extracted from the function $Q(\boldsymbol{R}\cdot\boldsymbol{m})$, should represent the population of folded structures in the chain interior of confined DNA. The third column, extracted from the function $Q(\boldsymbol{R}\cdot\boldsymbol{e})$, presumably involves both the chain-end folding as well as the interior folding. The difference of these two columns shows that the folding events at the chain ends are by far the dominant ones.

The mean values of the distribution functions $Q(\boldsymbol{R}\cdot\boldsymbol{e})$ or $Q(\boldsymbol{R}\cdot\boldsymbol{m})$ give the projection persistence lengths $P_{app}(e)$ and $P_{app}(m)$ defined by eqn (4) and represent the global stiffness of the DNA molecule in a channel of a diameter $2D$. They are apparent quantities involving both the intrinsic DNA stiffness and the confinement-induced external term.[29] It is seen (Fig. 5c and Table 1) that extra stiffening generated by the channel walls can increase the conventional value of the DNA persistence length by more than one order of magnitude.

The persistence length in a channel P_{app} is loosely connected to the parameter g ("global persistence length") introduced in ref. 15 as the distance between two consecutive U-hairpins in confined DNA. The theory of hairpin formation in nanochannels alleges that walls of a channel impose an entropic force that tightens hairpins. An exponential relation was deduced between the hairpin separation g and the bulk persistence length P. For the cylindrical channel of dimension $2D/P = 1$ the distance $g = 22$ μm between two consecutive hairpins was predicted.[15] The apparent persistence lengths $P_{app}(m)$ from Table 1 for $2D/P = 1.06$ is much smaller, 0.87 μm. The proposed theory[15] seems to underestimate the actual abundance of interior and chain-end hairpins in nanoconfined DNA.

The application of the technique of DNA linearization through confinement critically depends on the conjecture that extension of confined DNA scales linearly with the contour length L. The deflection and blob regime theories corroborate this assumption: the linearity of the function $R_{\|}(L)$ is postulated by eqn (1) and (3). However, in the transition region around $2D/P \sim 1$ the linear relation $R_{\|} \sim L$ for DNA extension is not authenticated since appropriate theories are not available. The

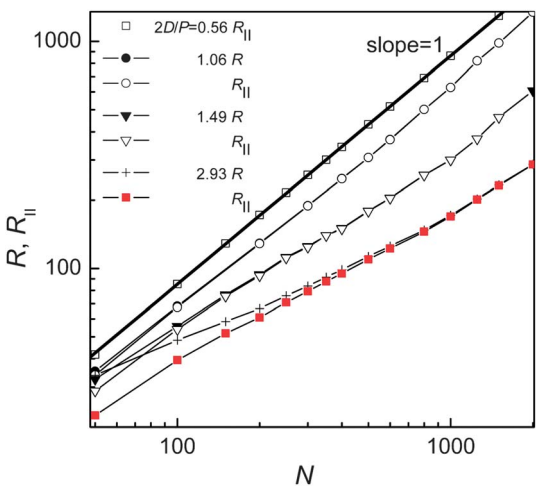

Fig. 6 DNA extension as a function of the chain length N in a cylinder for the four values of $2D/P$ channel dimensions given in the legend, representing the various confinement regimes.

Table 1 The abundance (in %) of folded DNA chains of $N = 800$ estimated from the distribution functions $Q(\boldsymbol{R}\cdot\boldsymbol{m})$ and $Q(\boldsymbol{R}\cdot\boldsymbol{e})$ shown in Fig. 5c in a range of channel widths $2D/P$. The apparent persistence lengths and $P_{app}(m)$ and $P_{app}(e)$ of DNA under channel confinement are also listed

$2D/P$	From $Q(\boldsymbol{R}\cdot\boldsymbol{m})$ (chain interior)	From $Q(\boldsymbol{R}\cdot\boldsymbol{e})$ (ends + interior)	$P_{app}(m)$ [nm]	$P_{app}(e)$ [nm]
1.49 (blob regime)	18.1	43.7	380	87
1.06 (transition region)	4.8	36.1	870	297
0.77 (~Odijk regime)	0.4	16.1	1260	782

channel widths corresponding to the transition region are within the dimension scale used in experimental studies.

To examine the scaling relation $R_{\|} \sim N^{z}$ the simulations in the range of N from 50 to 2000 were performed for four selected channel dimensions (Fig. 6). It is seen that for a narrow channel of $2D/P = 0.56$ in the Odijk regime the simulations confirmed the exponent $z = 1$ in the whole range of DNA lengths. Importantly, the linear scaling of the DNA extension with its length is also confirmed in Fig. 6 in the transition region. Evidently, a significant DNA backfolding in the transition region, particularly at the chain ends (Table 1), does not affect the exponent $z = 1$.

On the other hand, for moderately wide channels $2D/P = 1.49$ in the blob regime, the extension of confined DNA scales linearly with the chain length only above some critical chain length, roughly for $N > 1000$. For chains shorter than the critical length the exponent $z < 1$ in Fig. 6. This particular finding from simulation may be of relevance to analysis of the nanofluidic channel experiments of the short fragments of DNA.[5] Finally, Fig. 6 presents the results for the wide channel $2D/P = 2.93$, well below the de Gennes regime. Obviously, the extension of DNA molecules under weak confinement does not follow the linear scaling with the contour length L.

IV. Conclusions

The role of the chain backfolding in nanoconfinement-based techniques of dsDNA linearization was examined by MC simulation. Three regions, the deflection, transition and ideal-chain blob regime, were identified on the DNA extension profiles as a function of the channel diameter. From several quantities used to assess the DNA extension in a channel, the chain span was found to best concur with the blob scaling relation. A deflection peak due to the population of undulated molecules was observed on the distribution functions of the end-to-end distance $W(R)$ in narrow channels. The amount of DNA molecules of intermediate extensions on the function $W(R)$ increased by channel widening as the chains were getting shorter mainly due to the end-chain J-hairpins and loops. The distinctive events of the transition to shorter folded structures of DNA were detected on the simulation traces of chain extension as a function of time. The calculation of the scalar products related to the projection persistence length allowed for the first time the evaluation of the abundance of DNA chains folded at the chain ends and in the chain interior as a function of the channel width. It was found that the chain-end regions, due to their enhanced flexibility, are most prone to the local folding exemplified by the J-type or U-type hairpins and loops. The backfolding in the DNA chain interior into the Z-hairpins is relatively abundant in the de Gennes regime, significantly restricted in the transition region and practically absent in the Odijk regime. The linear proportionality between the chain extension $R_{\|}$ and the contour length L was examined by simulations in the whole scale of channel diameters and validated for the deflection and transition regimes. The importance of accounting for the local folding in interpretation of the DNA experiments in nanochannels was pointed out.

Acknowledgements

This work was supported by the grant APVV-0451-11 and VEGA grants 2/0093/12 and 2/0079/12.

References

1 N. Douville, D. Huh and S. Takayama, *Anal. Bioanal. Chem.*, 2008, **391**, 2395–2409.
2 J. O. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W. W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, P. Silberzan and R. H. Austin, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 10979–10983.
3 W. Reisner, K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, M. Rosen, J. C. Sturm, S. Y. Chou, E. Frey and R. H. Austin, *Phys. Rev. Lett.*, 2005, **94**, 196101.
4 K. Jo, D. M. Dhingra, T. Odijk, J. J. de Pablo, M. D. Graham, R. Runnheim, D. Forrest and D. C. Schwartz, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 2673–2678.
5 D. Stein, F. H. J. van der Heyden, W. J. A. Koopmans and C. Dekker, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15853–15858.
6 J. T. Mannion, C. H. Reccius, J. D. Cross and H. G. Craighead, *Biophys. J.*, 2006, **90**, 4538–4545.
7 J. Tang, S. L. Levy, D. W. Traham, J. J. Jones, H. G. Craighead and P. S. Doyle, *Macromolecules*, 2010, **43**, 7368–7377.
8 Y. Kim, K. S. Kim, K. L. Kounovsky, R. Chang, G. Y. Jung, J. J. dePablo, K. Jo and D. C. Schwartz, *Lab Chip*, 2011, **11**, 1721–1729.
9 W. Reisner, J. P. Beech, N. B. Larsen, H. Flyvbjerg, A. Kristensen and J. O. Tegenfeldt, *Phys. Rev. Lett.*, 2007, **99**, 058302.
10 S. L. Levy, J. T. Mannion, J. Cheng, C. H. Reccius and H. G. Craighead, *Nano Lett.*, 2008, **8**, 3839–3844.
11 F. Persson, P. Utko, W. Reisner, N. B. Larsen and A. Kristensen, *Nano Lett.*, 2009, **9**, 1382–1385.
12 J. J. Jones, J. R. C. van der Maarel and P. S. Doyle, *Nano Lett.*, 2011, **11**, 5047–5053.
13 T. Su, S. K. Das, M. Xiao and P. K. Purohit, *PLoS One*, 2011, **6**, e16890.
14 T. J. Odijk, *Macromolecules*, 1983, **16**, 1340–1344.
15 T. J. Odijk, *Chem. Phys.*, 2006, **125**, 204904.
16 P.-G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University, Ithaka, NY, 1979.
17 F. Brochard-Wyart, T. Tanaka, N. Borghi and P.-G. de Gennes, *Langmuir*, 2005, **21**, 4144–4148.
18 D. W. Schaefer, J. F. Joanny and P. Pincus, *Macromolecules*, 2010, **43**, 3094–3102.
19 P. Cifra, Z. Benková and T. Bleha, *Faraday Discuss.*, 2008, **139**, 377–392.
20 P. Cifra, Z. Benková and T. Bleha, *J. Phys. Chem. B*, 2009, **113**, 1843–1851.
21 Y. Wang, D. R. Tree and K. D. Dorfman, *Macromolecules*, 2011, **44**, 6594–6604.
22 T. Odijk, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2008, **77**, 060901(R).
23 Y. Yang, T. W. Burkhardt and G. Gompper, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2007, **76**, 011804.
24 P. Cifra, Z. Benková and T. Bleha, *J. Phys. Chem. B*, 2008, **112**, 1367–1375.
25 P. Cifra, *J. Chem. Phys.*, 2009, **131**, 224903.
26 D. Marenduzzo and C. E. Micheletti, *J. Mol. Biol.*, 2003, **330**, 485–492.
27 L. J. Steinbock, O. Otto, C. Chimerel, J. Gornall and U. F. Keyser, *Nano Lett.*, 2010, **10**, 2493–2497.
28 Y. Jung, C. Jeon, J. Kim, H. Jeong, S. Jun and B.-Y. Ha, *Soft Matter*, 2012, **8**, 2095–2102.
29 P. Cifra, Z. Benková and T. Bleha, *Phys. Chem. Chem. Phys.*, 2010, **12**, 8934–8942.
30 H.-P. Hsu, W. Paul and K. Binder, *Macromolecules*, 2010, **43**, 3094–3102.

