

Force responses of strongly intrinsically curved DNA helices deviate from worm-like chain predictions

M. TOMPITAK¹, H. SCHIESSEL¹ and G. T. BARKEMA²

¹ Lorentz Institute, Leiden University - Niels Bohrweg 2, 2333CA Leiden, The Netherlands

² Department of Information and Computing Sciences, Utrecht University - Princetonplein 5, 3584CC Utrecht, The Netherlands

received 1 November 2016; accepted in final form 20 January 2017

published online 14 February 2017

PACS 87.14.gk – DNA

PACS 87.15.La – Mechanical properties

PACS 87.15.Qt – Sequence analysis

Abstract – DNA sequences with nontrivial intrinsic curvature are of interest for a range of biological and artificial DNA systems. We design both intrinsically strongly curved and intrinsically straight sequences. We find that such sequences with opposing curvatures can be designed even under constraints that would naively lead one to assume that those sequences would be highly similar in their mechanical properties. We then characterize the force response of those sequences and find that their force-extension curves deviate significantly in the low-force regime, and that the standard worm-like chain description is inadequate to describe the low-force response of the strongly bent sequences. We propose a modified description that takes the intrinsic curvature into account, making the DNA act, in the low-force regime, like a nanoscale helical spring. We find strongly improved agreement between the model and the simulated force-extension curves.

Copyright © EPLA, 2016

DNA with intrinsic curvature plays an important role in biological systems, influencing, *e.g.*, the positions of nucleosomes [1–3] and plectonemes [4–6] and the propensity for stretches of DNA for loop formation [7–9]. DNA molecules can intrinsically encode their preferential spatial organisation through their underlying base pair sequence. This is the first step of many layers of spatial DNA organisation on larger and larger scales, the latter steps only now beginning to be accessible to a qualitative and quantitative understanding [10].

The standard description of DNA as a Worm-Like Chain (WLC) assumes the mechanical properties of DNA to be isotropic, and hence does not account for intrinsic curvature. Previous work has shown that, in general, this assumption is reasonable as long as the intrinsic curvature does not build up [11]. Here, we turn our attention to DNA molecules in which this assumption does not hold. Since strongly bent sequences occur in real organisms [12], it is not out of the question that long stretches of DNA with coherent curvature naturally occur, and we wish to know how this may cause predictions based on the WLC model to go wrong.

Artificial sequences with specific patterns of intrinsic curvature are also of interest. Tandem repeats of the

Widom 601 sequence [13] have been used as templates for reconstituting chromatin fibers [14–16]. This sequence is intrinsically curved in one specific direction, and tandem repeats of this sequence have the potential to form superhelical structures, see fig. 1(a).

Here, we design sequences with such superhelical intrinsic curvature *in silico* and characterize them by simulating their force response. Such structures have been studied theoretically (*e.g.*, [17–19]) and it is known that the superhelical geometry influences their force response, if the curvature of the helix is strong enough so that the persistence length of the polymer is at least comparable to the contour length of a superhelical turn [19].

The question is then whether there exist DNA sequences with an intrinsic curvature strong enough to influence their force response. It is not *a priori* clear whether we should expect this to be the case. The tandem repeats of curved sequences like 601, are not generally assumed to feature such effects, but the 601 sequence is unlikely to be the most strongly bent sequence, because it is experimentally difficult to access the entirety of the space of sequences of substantial length. To do this, and find the limits of how strongly curved a DNA sequence can be, we need computational methods

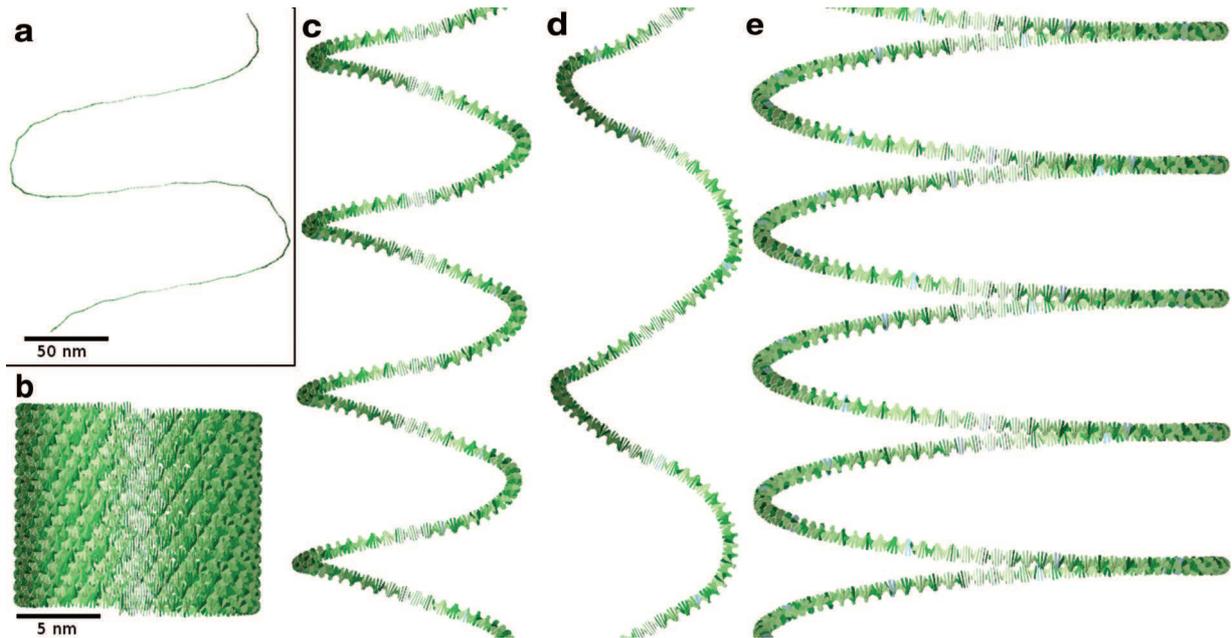


Fig. 1: (Colour online) The ground-state structures of tandem repeats of (a) the Widom 601 sequence, (b)–(d) our artificial, highly bent DNA sequence, as predicted by the rigid base pair model, and (e) the natural (in its singular form) kinetoplast sequence. Helices with various shapes can be produced by varying the number of additional nucleotides between copies of the sequence. The numbers of additional base pairs in these figures are (a) 0; (b)–(d) 0, 1 and 2, (e) 2. For the geometric parameters corresponding to these structures, see table 1.

like those presented in [2], as described in the next section.

It turns out that strong, coherent intrinsic curvature can indeed be achieved for DNA molecules. Molecules that feature such curvature act like nanoscale helical springs, which intrinsically resist stretching. The force response predicted by the WLC model provides a poor fit in the low-force regime for these sequences. We suggest an alternative description, in which we take the superhelical structure into account.

Some study has been made of the finite-temperature force response of intrinsically curved polymers, but the problem has only been fully solved in two dimensions [20]. Here we employ a discrete description [21] of a flexible helix, coarse-graining the DNA to the same level as done by the Rigid Base Pair model [22], which we use to numerically assess the force response. For low forces, the force response of this discrete superhelix can be described by an extensible WLC model [23,24] with effective values for the bending and stretching moduli. This model leads to a significantly improved prediction for the low-force regime of the force response over the naive WLC model.

Designing sequences. – To design sequences with specific properties, we modeled the DNA using the Rigid Base Pair (RBP) model parameterized by crystallography data [22], and we employed the Mutation Monte Carlo (MMC) method to search the space of all possible sequences. The original method [2] introduces mutation moves to a standard Monte Carlo simulation of a DNA

molecule. We employed a simplified version, which performed only mutation moves and searched (through simulated annealing) for the most strongly bent sequence. The strength of the intrinsic curvature of the sequence was measured by calculating the ground-state configuration of the sequence and taking the inner product between the first and last tangent vectors to the DNA backbone.

We ran this algorithm on a DNA molecule consisting of 84 base pairs, and created a tandem repeat of the most bent sequence thus found. In order to create sequences with different superhelical properties, we created repeats with additional homogeneous (sequence-averaged) DNA between the repeats, to interfere with the alignment of the direction of curvature between successive copies of the sequence. This led to DNA sequences with various values for the superhelical radius and pitch angle, see fig. 1(b)–(d).

We also designed a sequence with low intrinsic curvature by taking our tandem repeat sequence and applying an MMC algorithm similar to that described above, but which maximizes the sum of the inner products between the first tangent vector, and a number of tangent vectors along the rest of the DNA molecule, at intervals of 50 base pairs (roughly 1/3 of the persistence length of DNA). This algorithm ensures that the resulting sequence has low intrinsic curvature both locally and globally.

A major difference between this MMC simulation and the previously described one, however, is a constraint on the allowed mutations. Starting with the strongly curved sequence, we mutated it such that the distribution of dinucleotides in the sequence remained the same, and only

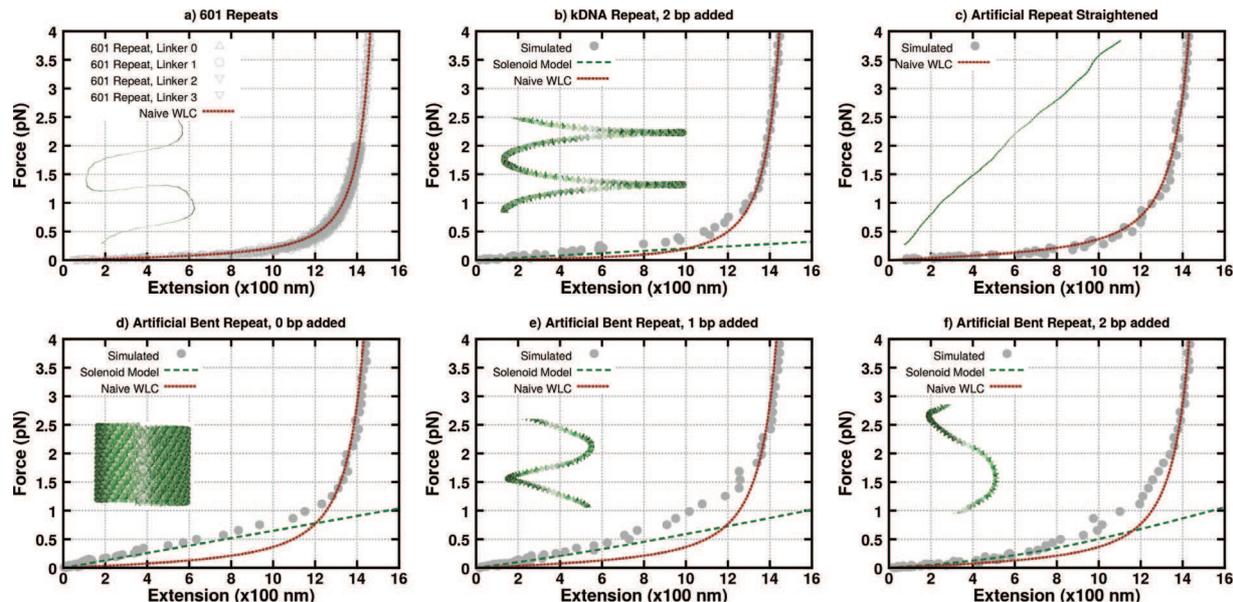


Fig. 2: (Colour online) Simulated force-extension curves, WLC fits and predictions of the two-angle model for the low-force regime, for (a) repeats of the Widom 601 sequence with various numbers of additional base pairs; (b) the kinetoplast sequence (kDNA); (c) the straightened artificial sequence; (d)–(f) repeats of the artificial curved sequence with various numbers of additional base pairs. The inset pictures show the intrinsic shapes of these sequences. For the geometric parameters corresponding to these sequences, see table 1.

their order was allowed to change. We achieved this by performing mutation moves that swap pieces of the sequence that both start and end with the same nucleotides. For example, ATA might be swapped with ATTA, but not with ATT. By imposing this constraint, both the nucleotide and dinucleotide contents are kept identical. This means, for example, that the GC content of the sequence, which is generally thought to be a good indicator of the mechanical properties of a sequence, is invariant. Even more strongly, keeping the dinucleotide content constant means that (in the RBP model) the resulting DNA molecules contain the exact same elastic components, only in a different order. Even under this constraint, we find that we can design sequences that seem similar, but in fact have vastly different intrinsic curvature properties, and, therefore, different force responses.

Force responses. – In order to measure the force response, we ran a Monte Carlo simulation of a DNA molecule with the given sequence under a number of fixed forces and in each case sampled the extension. This led to the force-extension curves presented in fig. 2. In the case of the intrinsically straight sequence, the data can be fitted with the prediction from a simple WLC model [23] to good agreement, see fig. 2(c).

For an example of sequences with superhelical intrinsic curvature, we first looked to the Widom 601 sequence. A repeat of this sequence indeed forms superhelices, as can be seen in fig. 1(a). Pulling on the helix displayed there as well as three other variants with various numbers of additional base pairs added between repeats, we obtain the force responses depicted in fig. 2(a). The different

Table 1: Geometrical parameters (superhelical radius R and distance along the backbone between successive base pairs s_0) and number of times the persistence length of DNA fits into a single helical turn (l being the contour length of a single helical turn and l_p the persistence length) for the superhelices depicted in fig. 1.

Sequence	R (nm)	s_0 (nm)	l/l_p
Artificial + 0 bp	14.60	0.00537	1.96
Artificial + 1 bp	13.37	0.0826	1.86
Artificial + 2 bp	11.09	0.164	1.72
kDNA	26.02	0.0309	3.52
601	113.34	0.152	17.1

responses strongly overlap and do not deviate appreciably from the WLC fit. The intrinsic curvature of the 601 sequence is not strong enough to have a significant effect on the force response. This has also been observed experimentally, see supplementary fig. 2 in ref. [15], and can be understood from the geometry of the superhelix: the contour length of a superhelical turn is larger than the persistence length of the DNA at room temperature (see table 1) meaning that the intrinsic curvature is lost to thermal fluctuations over distances at which its magnitude becomes significant [17,19]. The result is that at this temperature, the relatively weak superhelical nature of the 601 repeat is not distinguishable from a straight molecule.

In order to see the effect of the intrinsic curvature, we need a more strongly bent sequence. Nature in fact provides such a sequence. A strongly curved section of DNA

has been discovered in kinetoplast DNA (kDNA) [12,25]. Taking the sequence depicted in fig. 2 in [12] and repeating it with 2 additional base pairs in between, this DNA sequence forms a much tighter superhelix than the 601 sequence, see fig. 1(e). For this structure, with a single-turn contour length only 3.5 times the persistence length (see table 1), the force response turns out not to fit the WLC prediction well. In fig. 2(b), we see a clear discrepancy for low forces. At high extension, the force response is similar, but at low extension, the tendency of the DNA to intrinsically curl up means that it acts like a spring. Hence more force is needed to stretch it, and the slope of the force-*vs.*-extension curve is correspondingly higher.

To get a better grasp on the importance of the DNA sequence to the force response, we turn to the artificial sequences described in the previous section and compare the force responses of the helical sequences with that of the straightened one. Since the average elastic properties of the straight and the curved versions of this sequence are identical, any difference must be due to the build-up of intrinsic curvature. In figs. 1(b)–(d) and table 1 we see that the artificial curved sequence forms even tighter helices, with single-helix contour lengths not much larger than the persistence length, so we expect an even larger effect.

In each of the figs. 2(d)–(f), the red dotted line is identical to that in fig. 2(c), plotted for reference. In each case there is indeed significant deviation from the WLC prediction; the more compact the helix, the stronger the deviation.

Modeling superhelical DNA molecules. – The failure of the WLC model shown above is due to the assumption of no significant intrinsic curvature. The WLC force-extension curve assumes that the force response is dominated by the ironing out of thermal fluctuations around an intrinsically straight ground state. To correctly describe the situation under consideration, we must include the response due to the intrinsic resistance to stretching by the DNA molecule.

Some studies of flexible rods with intrinsic curvature exist [17–20,26] but no analytical description at finite temperature in three dimensions is known. Since the discrepancy for DNA lies mostly in the low-force regime, we propose a partial solution that describes this regime well.

In order to account for the intrinsic curvature of the superhelical DNA molecules, we turn to a discrete model for such structures, consisting of a series of flexible, straight rods, each of which is connected to the next at fixed angles [21,27,28]. Two such angles are necessary to describe the orientation of one segment with respect to the next, so that this model is generally known as the two-angle model. These two angles, together with the length of the connecting rods, fix the shape of the entire superhelix.

For low forces, the backbone of the superhelix in this description behaves as an extensible worm-like chain [23,24], with effective values for the stretching ($\tilde{\gamma}$), bending (\tilde{A})

and twisting (\tilde{C}) moduli, as well as a coupling between stretching and twisting (\tilde{g}) [21,27]. With these parameters, the linear response of the superhelix is given by

$$\begin{pmatrix} F \\ M_t \\ M_b \end{pmatrix} = \begin{pmatrix} k_B T \tilde{\gamma} & k_B T \tilde{g} & 0 \\ k_B T \tilde{g} & \tilde{C} & 0 \\ 0 & 0 & \tilde{A} \end{pmatrix} \begin{pmatrix} x \\ \Omega \\ R^{-1} \end{pmatrix}, \quad (1)$$

where F , M_t and M_b are the force, torsional torque and flexural torque applied to the system, and x , Ω and R^{-1} the extension, twist and curvature. We are interested in the dependence of F on x . If we assume no torques on the system, eq. (1) reduces to

$$F = \left(k_B T \tilde{\gamma} - (k_B T)^2 \frac{\tilde{g}^2}{\tilde{C}} \right) x. \quad (2)$$

The effective values of the mechanical properties of the superhelical backbone depend on the elastic properties of the flexible rods that make up the superhelix, as well as the geometry of the structure, *i.e.*, the two angles mentioned above, as described by eqs. (96) and further in ref. [27]. In applying this model to our DNA structures, the flexible rods represent the connections between consecutive base pairs, so their elastic properties are those of the DNA dinucleotide steps in the RBP model.

We will characterize a superhelix not by the two local angles between successive rods, but by the superhelical radius R and the superhelical rise (*i.e.*, the distance travelled along the backbone) per base pair, s_0 , as these are simpler to determine numerically. The expressions for the elastic parameters in eq. (1) simplify considerably if we assume that $R \gg b$ (for our helices, R is generally larger by at least a factor of 30), in which case we find

$$\tilde{\gamma} = \frac{r}{k_B T} \frac{C + (A - C)r^2}{R^2}, \quad (3)$$

$$\tilde{g} = \frac{r^2}{k_B T} \frac{(A - C)(b^2 - s_0^2)}{bR}, \quad (4)$$

$$\tilde{C} = r(A - (A - C)r^2), \quad (5)$$

$$\tilde{A} = \frac{2rAC}{A + C - (A - C)r^2}, \quad (6)$$

$$r = \frac{s_0}{b}, \quad (7)$$

where A and C are the bending and twisting moduli of our DNA model, R and s_0 are as defined above, and b is the length of the flexible rods, *i.e.*, the distance between successive base pairs, taken to be 0.34 nm. In this continuum limit, the effective spring constant of the system, *i.e.*, the constant coefficient in eq. (2), reduces to the standard classical form [19,26].

The bending modulus is easily estimated by sampling the tangent-tangent correlations in a standard Monte Carlo simulation of a homogeneous DNA molecule with sequence-average properties at zero force. The twist modulus can be estimated (neglecting the cross-terms in the

RBP Hamiltonian) from the twist stiffnesses of the RBP model [29]. The geometrical parameters of the helices were estimated directly from the ground-state structure of the sequences as follows.

We created ground-state structures for long sequences consisting of 500–1000 concatenated copies of the sequences of interest. This gave us DNA states like those depicted in fig. 1 but where the superhelical backbone was much larger than the radius, reducing edge effects. We approximated the backbones of these structures by calculating the three-dimensional straight line that best fitted (in the least-squares sense) the positions of all base pairs in the superhelix.

We estimated the radius of each superhelix by taking the average distance of the base pairs to their closest point on this backbone. We also used the backbone to calculate for each base pair the (signed) angle between the connecting line between the base pair and the backbone, and the same for a reference base pair. These angles ran cyclically from $-\pi$ to π , and allowed us to determine the base pairs which were an integer number of helical turns away from the first base pair. This allowed us to calculate the average distance along the backbone between two successive such base pairs, and the average number of base pairs per superhelical turn. Hence we calculated s_0 , the superhelical rise (*i.e.*, the distance travelled along the backbone) per base pair. The resulting parameters are shown in table 1.

Using these values to inform eqs. (2)–(7), the force-extension curves for these structures are given by that of the extensible worm-like chain with our effective parameters, which can be approximated by [24]

$$F = \left(\frac{(k_B T)^2}{\tilde{A}} \right) \left[\frac{1}{4(1 - x/L + F/K)^2} - \frac{1}{4} + \frac{x}{L} - \frac{F}{K} \right], \quad (8)$$

where L is the contour length of the DNA and K the effective spring constant of the backbone, written out explicitly in eq. (2).

This equation gives us the force-extension curves represented by the dashed green lines in figs. 2(b), (d)–(f). The agreement with the data at low forces is significantly better than that of the naive WLC prediction.

The force-extension curves of these superhelical DNA molecules have three regimes. At low forces, entropy dominates and the force response follows from the random walk statistics of the effective superhelical axis. Increasing the force, the superhelical backbone is straightened out. When the extension is approximately equal to the intrinsic contour length of the superhelical backbone, we reach a regime where the superhelix acts as a spring (the energetic regime of our effective extensible WLC) [30]. The slope of the force-extension curve response becomes higher because of this additional restoring force and approaches the classical result for the linear regime of a flexible helix [19,26]. Finally, the superhelix becomes so distorted that the effective model breaks down, and the normal high-force regime of DNA takes over.

Conclusion. – We find that, while in most cases intrinsic curvature may not be an important factor to the force response of a DNA molecule, it is easy to design sequences for which the effects of intrinsic curvature are visible. Short sequences with such strong curvature also occur in nature, *e.g.*, in kinetoplast DNA. In such cases, fitting the naive WLC force-extension curve does not give satisfactory results in the low-force regime. The intrinsic curvature needs to be taken into account, and we have provided a model to do so, at finite temperature in three dimensions, and valid for low forces.

The commonly used tandem repeats of the 601 sequence do not show appreciable effects in their force response and experimental setups using such sequences should not be affected. However, should more strongly bent sequences become desirable for such experiments in the future, care should be taken to ensure that the effects we note here are accounted for.

We also note that the effects to the force response of a DNA molecule can be tuned through strongly constrained mutations of the sequence. Surprisingly, both very straight and strongly bent DNA molecules, with markedly different force responses, can be obtained without altering the overall GC content, or even the overall numbers of different dinucleotides.

We would like to thank JOHN VAN NOORT for useful discussions. This work was supported by the Netherlands Organisation for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience program.

REFERENCES

- [1] SEGAL E., FONDUFE-MITTENDORF Y., CHEN L., THÅSTRÖM A., FIELD Y., MOORE I. K., WANG J.-P. Z. and WIDOM J., *Nature*, **442** (2006) 772.
- [2] ESLAMI-MOSSALLAM B., SCHRAM R. D., TOMPITAK M., VAN NOORT J. and SCHIESSSEL H., *PLoS ONE*, **11** (2016) e0156905.
- [3] FREEMAN G. S., LEQUIEU J. P., HINCKLEY D. M., WHITMER J. K. and DE PABLO J. J., *Phys. Rev. Lett.*, **113** (2014) 1.
- [4] LAUNDON C. H. and GRIFFITH J. D., *Cell*, **52** (1988) 545.
- [5] VAN LOENHOUT M. T. J., DE GRUNT M. V. and DEKKER C., *Science*, **338** (2012) 94.
- [6] BUSSIEK M., KLENIN K. and LANGOWSKI J., *J. Mol. Biol.*, **322** (2002) 707.
- [7] BOEDICKER J. Q., GARCIA H. G., JOHNSON S. and PHILLIPS R., *Phys. Biol.*, **10** (2013) 066005.
- [8] HAEUSLER A. R., GOODSON K. A., LILLIAN T. D., WANG X., GOYAL S., PERKINS N. C. and KAHN J. D., *Nucleic Acids Res.*, **40** (2012) 4432.
- [9] ROSANIO G., WIDOM J. and UHLENBECK O. C., *Biopolymers*, **103** (2015) 303.
- [10] DEKKER J. and MIRNY L., *Cell*, **164** (2016) 1110.
- [11] VOLOGODSKAIA M. and VOLOGODSKII A., *J. Mol. Biol.*, **317** (2002) 205.
- [12] WIDOM J., *BioEssays*, **2** (1985) 11.

- [13] LOWARY P. T. and WIDOM J., *J. Mol. Biol.*, **276** (1998) 19.
- [14] ROBINSON P. J. J., FAIRALL L., HUYNH V. A. T. and RHODES D., *Proc. Natl. Acad. Sci. U.S.A.*, **103** (2006) 6506.
- [15] KRUTHOF M., CHIEN F.-T., ROUTH A., LOGIE C., RHODES D. and VAN NOORT J., *Nat. Struct. Mol. Biol.*, **16** (2009) 534.
- [16] SCHRAM R. D., KLINKER H., BECKER P. B. and SCHIESSEL H., *Eur. Phys. J. E*, **38** (2015) 85.
- [17] PANYUKOV S. and RABIN Y., *Phys. Rev. E*, **62** (2000) 7135.
- [18] KESSLER D. A. and RABIN Y., *Phys. Rev. Lett.*, **90** (2003) 024301.
- [19] WADA H. and NETZ R. R., *EPL*, **77** (2007) 68001.
- [20] BENETATOS P. and TARENTJEV E. M., *Phys. Rev. E*, **81** (2010) 031802.
- [21] BEN-HAÏM E., LESNE A. and VICTOR J. M., *Phys. Rev. E*, **64** (2001) 51921.
- [22] OLSON W. K., GORIN A. A., LU X. J., HOCK L. M. and ZHURKIN V. B., *Proc. Natl. Acad. Sci. U.S.A.*, **95** (1998) 11163.
- [23] MARKO J. F. and SIGGIA E. D., *Macromolecules*, **28** (1995) 8759.
- [24] WANG M. D., YIN H., LANDICK R., GELLES J. and BLOCK S. M., *Biophys. J.*, **72** (1997) 1335.
- [25] WU H. M. and CROTHERS D. M., *Nature*, **308** (1984) 509.
- [26] LOVE A. E. H., *Treatise on Mathematical Theory of Elasticity*, 2nd edition (Cambridge University Press) 1906.
- [27] SCHIESSEL H., *J. Phys.: Condens. Matter*, **15** (2003) R699.
- [28] SCHIESSEL H., *Biophysics for Beginners: A Journey through the Cell Nucleus* (Pan Stanford) 2014.
- [29] MERGEL B., EJTEHADI M. and EVERAERS R., *Phys. Rev. E*, **68** (2003) 15.
- [30] ODIJK T., *Macromolecules*, **28** (1995) 7016.