

Twist-stretch profiles of DNA chains

Marco Zoli

School of Science and Technology

University of Camerino,

I-62032 Camerino, Italy

marco.zoli@unicam.it

(Dated: April 18, 2017)

Abstract

Helical molecules change their twist number under the effect of a mechanical load. We study the twist-stretch relation for a set of short DNA molecules modeled by a mesoscopic Hamiltonian. Finite temperature path integral techniques are applied to generate a large ensemble of possible configurations for the base pairs of the sequence. The model also accounts for the bending and twisting fluctuations between adjacent base pairs along the molecules stack. Simulating a broad range of twisting conformation, we compute the helix structural parameters by averaging over the ensemble of base pairs configurations. The method selects, for any applied force, the average twist angle which minimizes the molecule's free energy. It is found that the chains generally over-twist under an applied stretching and the over-twisting is physically associated to the contraction of the average helix diameter, i.e. to the damping of the base pair fluctuations. Instead, assuming that the maximum amplitude of the bending fluctuations may decrease against the external load, the DNA molecule first over-twists for weak applied forces and then untwists above a characteristic force value. Our results are discussed in relation to available experimental information albeit for kilo-base long molecules.

PACS numbers: 87.14.gk, 87.15.A-, 87.15.Zg, 05.10.-a

1. Introduction

While DNA packaging in bacteria and eukaryotic chromosomes rely on the mechanical properties of the double helix, the sequence specific chain deformability is essential as much to the formation of transient fluctuational bubbles involved in cell division and transcription as to DNA recognition by proteins which twist, bend and stretch the helix upon binding [1–3]. Proteins, typically interacting with short base pair sequences, say 10 - 20 bps, confer a manifold of functions to the bare DNA chains [4]. A remarkable example is provided by recombinases, such as RecA in *E. coli*, which promote homologous pairing between e.g., a partially single stranded DNA and a target double stranded molecule [5] thus producing displacement and exchange of strands, a vital process in the recombinational DNA repair [6]. The mechanisms by which the RecA nucleoprotein filament efficiently samples the target dsDNA and pairs homologous sequences are largely unknown: some studies [7] suggest that the RecA - dsDNA binding is favored by a double helix unwinding with opening of thermal bubbles whereas others [8] point out that the rate of homologous pairing is higher when dsDNA maintains a random coil conformation with short end-to-end distance as this enhances the local DNA concentration and allows a rapid homologous search. It is however likely that the specific twist configuration of the helix affects both the size and the binding properties of the recombinase - dsDNA complexes [9] while, in turn, the latter restructure and deform the double helix in several ways [10, 11].

Considerable knowledge of the DNA properties has been gained over the last twenty five years as a wide number of methods to micro-manipulate single molecules and sample their mechanical response to external forces has become available [12–16]. First measurements on λ -phage molecules in solution [13] have shown that, for low and intermediate forces, DNA displays entropic elastic behavior dominated by thermal bending fluctuations which constantly deform base pairs and molecule backbone so that the helix assumes different random walk configurations. Accordingly, inextensible worm-like-chain models with persistence length of ~ 50 nm fit well the force-extension data up to applied forces of ~ 10 pN [17, 18]. Above this typical force, the polymer end-to-end distance stretches linearly and takes values even larger than the B-form contour length. At ~ 65 pN, the force-extension plot shows a peculiar plateau signaling that the molecule undergoes a structural transition and over-stretches to ~ 1.7 times its B-form contour length [19–22], the precise values depending on

the solution condition [23]. Note that the RecA - dsDNA complex is stretched by a factor 1.5 with respect to the naked DNA length [24]. Importantly, assuming an experimental setup which anchors both ends of the molecule thus preventing it to twist, the over-stretching transition shifts at ~ 100 pN [25]. Moreover, regardless of the attachment geometry, optical tweezers methods combined with single molecule fluorescence imaging have visualized that over-stretched dsDNA gradually converts into single stranded DNA [26] thus suggesting a force induced melting transition.

Ten years ago Gore et al. [27] showed that DNA molecules in the kilo-base range over-twist under small to moderate applied tensions (up to of ~ 30 pN) and eventually untwist once larger applied forces begin to deform the molecule backbone. These apparently surprising findings have been interpreted in terms of a negative twist-stretch coupling which, in an elastic rod model, causes the helix both to over-twist and to extend upon constant moderate stretching. Crucial to the authors' model is that, while the rod is stretched, the helix radius can vary and, precisely, it shrinks. Similar results have been obtained by magnetic tweezers experiments [28] which find a linear *extension versus twist* behavior, independent of sequence specificities and buffer conditions, in the low force domain.

More generally, it has also been pointed out [29] that the theory of helical rods predicts different equilibrium twist conformations for a spring, subjected to a wrench. In particular, in the absence of an axial torque, the helical spring may first over-twist, then return to the initial state and eventually untwist as a function of the applied force.

However, while evidence has emerged concerning the interplay between twist and stretch [30] in the shaping of the DNA mechanics, a theoretical description of these phenomena at the molecular basis is still needed, mostly at forces for which DNA retains its double helical structure. Here, we address these issues by developing a mesoscopic Hamiltonian model for helicoidal molecules subject to external forces and compute their mechanical response to the stretching deformation. The focus is on very short DNA helices which have recently attracted much interest in view of their remarkable flexibility [31–33] despite the intrinsic inaccuracies encountered in the experimental characterization of the molecule structure at short length scales [34–36]. At such scales, the DNA contour length may become even shorter than the persistence length therefore challenging the application of usual elastic rod models which treat the molecule as a continuum [37].

Our computational method for the double helix Hamiltonian is based on a path integral

formalism [38–40] that incorporates in the base pair displacements those thermal fluctuations which are all the more relevant in short chains with large finite size effects. The same formalism has been used in some recent works [41, 42], to study persistence length and cyclization probabilities of short DNA sequences with ~ 100 base pairs. Our analysis has shown that DNA chains maintain a large bendability also at short scales and may thus have persistence lengths smaller than those predicted by the worm-like-chain-model for kilo-base long molecules. While these results point to the connection between global chain flexibility and inter-molecular interactions at the base pair level, it is true that the previously developed method assumed that the DNA sequences are torsionally constrained and therefore it cannot be applied to investigate the molecule twisting under a mechanical load. Accordingly, to pursue this task, we need to remove that constraint and formulate a more general computational program in order to evaluate the essential parameters which characterize the helix response to the external force, i.e. its diameter and the number of base pairs per helix turn.

The geometrical model for the helix is outlined in Section 2 while the Hamiltonian is presented in Section 3. The computational method is described in Section 4 and the results are shown in Section 5. Some conclusions are drawn in Section 6.

2. Model for the Helix

We depict the backbone of a DNA molecule with N base pairs, see Fig. 1(a), as a chain of $N - 1$ segments connecting the points O_i ($i = 1, \dots, N$) which are equally spaced along the helix mid-axis.

For short chains, the O_i 's can be considered as pinned to the sheet plane as the energetic cost to bend the backbone out of the plane is high. Accordingly, once the short chain closes into a loop, the supercoiling is mostly partitioned into twisting while the writhing contribution is essentially zero [43–45]. Thus, in the absence of fluctuations, the planar molecule backbone is a freely jointed polygonal chain with the line segment d being the rise distance between adjacent base pairs. For the i -th base pair, we define $x_i^{(1)}$, $x_i^{(2)}$ as the fluctuations of the two complementary mates measured from their respective strands. With respect to O_i , the positions of the two pair mates are expressed respectively by, $r_i^{(1)} = -R_0/2 + x_i^{(1)}$ and $r_i^{(2)} = R_0/2 + x_i^{(2)}$, where R_0 is the separation between the complementary

strands i.e., the equilibrium helix diameter. It follows that the relative distance between the complementary bases, with respect to the helix mid-axis, is $r_i = r_i^{(2)} - r_i^{(1)}$.

Thus, the variable r_i accounts for the radial base pair fluctuations around R_0 . Note that r_i may even become smaller than R_0 but the two bases cannot get too close to each other due to the electrostatic repulsion between negatively charged phosphate groups on complementary strands [40]. The O_i 's also provide flexible hinges allowing adjacent displacements along the molecule stack, e.g. r_i and r_{i-1} , to twist (by the angle θ_i) and to bend (by the angle ϕ_i) as shown in Fig. 1(b). Importantly, ϕ_i is also the angle between the helix diameters at adjacent sites O_i and O_{i-1} as, at any site, r_i and R_0 are parallel. While ϕ_i and θ_i are integration variables in the computation of the partition function (as discussed below), both bending and twisting fluctuations are incorporated in our three dimensional helicoidal model. Thus, both angular and radial fluctuations contribute to determine the separation between adjacent base pairs, $\overline{d_{i,i-1}}$, measured by the distance \overline{AB} in Fig. 1(b) and given by:

$$\begin{aligned}\overline{AB}^2 &= \overline{BB'}^2 + \overline{AB'}^2, \\ |\overline{BB'}| &= d + r_i \sin \phi_i, \\ \overline{AB'}^2 &= r_{i-1}^2 + (r_i \cos \phi_i)^2 - 2r_{i-1} \cdot r_i \cos \phi_i \cos \theta_i,\end{aligned}\tag{1}$$

In the following, we set the values $R_0 \sim 20\text{\AA}$ for the equilibrium helix diameter and $d \sim 3\text{\AA}$ for the bare rise distance. Small variations with respect to these values do not affect our results. Further features of the double stranded helix such as the presence of major and minor grooves and possible local distortions in helical parameters are not considered here [46, 47].

Moreover, it is important to realize that our model essentially treats the unit consisting of a phosphate, a sugar and a base, i.e., the nucleotide, as a point-like object ignoring its internal degrees of freedom. While such coarse-grained level of description seems at this stage sufficient to capture the pattern of the twist-stretch dependence in short chains, we also mention that a finer level of resolution (looking at the internal structure of the nucleotide) has been achieved in computational studies of the elastic and melting properties of nucleic acids [48].

The helicoidal chain is subject to an external force which is set along the direction of the rise distance $\overline{O_i O_{i-1}}$, as shown in Fig. 1(a), and assumed to act with the same modulus on

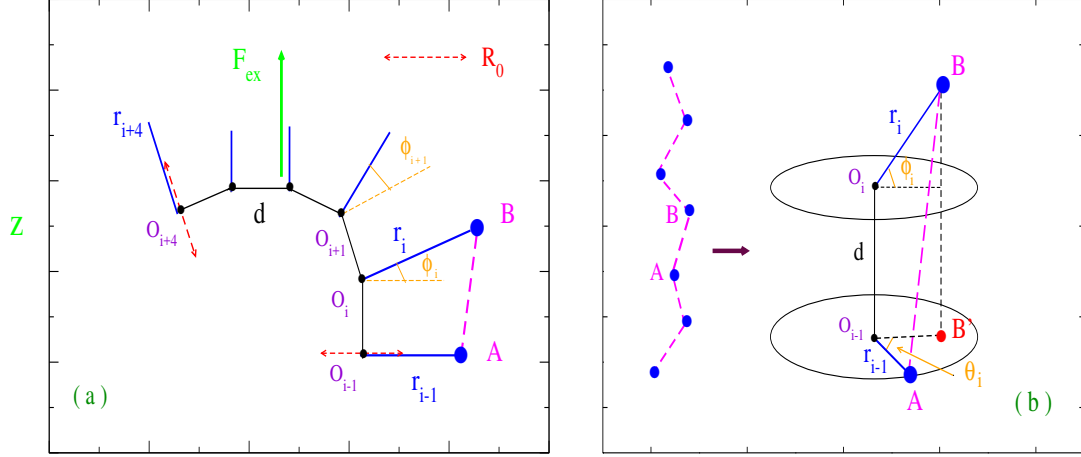


FIG. 1: (Color online) (a) Model for a chain of N base pairs with external stretching force F_{ex} applied along the z - axis. r_i is the relative distance between the two mates of the i -th base pair. The r_i 's have variable amplitude and are measured with respect to the O_i 's which are arranged along the mid-axis of the helix (lying on the sheet plane). R_0 is the equilibrium helix diameter and d is the rise distance between adjacent nucleotides along the molecule stack. In the absence of fluctuations, $r_i \equiv R_0$. Neighbor base pairs displacements, r_i and r_{i-1} , are bent by the angle ϕ_i . This is also the angle between the bond $\overline{O_{i+1}O_i}$ and F_{ex} as the latter acts along the direction of the $\overline{O_iO_{i-1}}$ segment. (b) Geometrical representation for the distance \overline{AB} between base pairs displacements. Both bending and twisting fluctuations are incorporated in the stacking potential. θ_i is the torsional angle formed by adjacent r_i and r_{i-1} . As defined in in Eq. (5), θ_i is measured with respect to the average angle $\langle \theta_{i-1} \rangle$ of the preceding base pair along the stack. In the absence of bending, the model would reduce to a fixed-plane representation as depicted by the ovals.

any molecule segment. Note that the force direction is taken constant to comply with the experimental setup while the molecule backbone remains anchored to the sheet plane also after applying the external load. In the presence of such tunable force field, we compute *i*) the average twisting fluctuations which allow to determine the number of base pairs per helix turn, i.e. the average helical repeat $\langle h_r \rangle$, *ii*) the average radial fluctuations $\langle r_i \rangle$ which measure the molecule diameter and *iii*) the average extensions $\langle \overline{d_{i,i-1}} \rangle$. All averages are

taken over an ensemble of molecule configurations according to the method explained in Section 4.

3. Mesoscopic Model

DNA in solution is stable at physiological temperatures despite the energetic molecular bumping due to the environment. In fact, as the room temperature thermal energy ($k_B T \sim 25meV$) is comparable to the hydrogen bond base pair energy [49], fluctuational bubbles can form spontaneously and transiently along the helix [50–56] whereas the molecule strands remain intact due to the overall strong covalent bonds acting between adjacent nucleotides [57]. Random coil molecule conformations are favored by entropic forces which bend the intra-strand stacking bonds although DNA becomes stiff at scales of order of the typical persistence length, i.e. $\sim 50nm$. Such length may be significantly smaller for short DNA [58–60] and sequences with ~ 100 base pairs or less could maintain an intrinsic flexibility which favors the molecule looping and the loop stability against fluctuational forces [61, 62].

To account for the main interactions that stabilize the helix we use a mesoscopic Hamiltonian containing a one-particle base pair potential $V_1[r_i]$ and a two-particle stacking term $V_2[r_i, r_{i-1}, \phi_i, \theta_i]$ also dependent on the angular degrees of freedom. Their analytic form reads:

$$\begin{aligned}
 V_1[r_i] &= V_M[r_i] + V_{Sol}[r_i], \\
 V_M[r_i] &= D_i [\exp(-b_i(|r_i| - R_0)) - 1]^2, \\
 V_{Sol}[r_i] &= -D_i f_s (\tanh((|r_i| - R_0)/l_s) - 1), \\
 V_2[r_i, r_{i-1}, \phi_i, \theta_i] &= K_S \cdot (1 + G_{i,i-1}) \cdot \overline{d_{i,i-1}}^2, \\
 G_{i,i-1} &= \rho_{i,i-1} \exp[-\alpha_{i,i-1}(|r_i| + |r_{i-1}| - 2R_0)].
 \end{aligned}
 \tag{2}$$

$V_1[r_i]$ is the sum of two terms: 1) a hydrogen bond Morse potential ($V_M[r_i]$) for the i -th base pair with dissociation energy D_i and spatial range measured by b_i . 2) A solvent term ($V_{Sol}[r_i]$) which modifies the Morse plateau through the parameters f_s and l_s . The one-particle potential has been used in the past [63–65] to simulate the denaturation of the double stranded DNA occurring for base pair separation large enough to sample the Morse

plateau. The same analytic form of $V_1[r_i]$ is suitable to model thermally stable hydrogen bonds [66] as in the current analysis provided that the base pair fluctuations remain smaller than the helix radius.

As mentioned above, the base pair separation may become shorter than R_0 but too large helix contraction are discouraged by the large energetic cost. This physical requirement is implemented in the code by discarding those paths for which the repulsive energy becomes too large, i.e. larger than the dissociation energy: $V_M[r_i] \geq D_i$. This implies that the base pair amplitudes must fulfill the condition, $|r_i| - R_0 < -\ln(2/b_i)$.

The stacking potential $V_2[r_i, r_{i-1}, \phi_i, \theta_i]$ contains an elastic force constant K_S and non-linear contributions weighed by the parameters $\rho_{i,i-1}$, $\alpha_{i,i-1}$ as originally proposed in the context of a simple ladder model for DNA [67]. V_2 depends on the angular degrees of freedom through the fluctuating distance $\overline{d_{i,i-1}}$ in Eq. (1). The physical motivations underlying the choice of these potentials have been widely discussed in previous works [40, 68] to which we refer for details. The model parameters, taken hereafter as in [41, 42], are appropriate to homogeneous sequences of *GC*- base pairs which have sizeable dissociation energies and are stable against thermal fluctuations. Certainly quantitative predictions of the twist-stretch relations may be derived for specific heterogeneous sequences since, in general, the helical properties are dependent on the base sequences, as also recently pointed out in the different context of charge transfer in B-DNA segments [69, 70].

With this caveat, our helicoidal molecule with N base pairs of reduced mass μ , stretched by a force F_{ex} as in Fig. 1(a), is described by the Hamiltonian:

$$\begin{aligned}
 H &= H_a[r_1] + \sum_{i=2}^N H_b[r_i, r_{i-1}, \phi_i, \theta_i], \\
 H_a[r_1] &= \frac{\mu}{2} \dot{r}_1^2 + V_1[r_1], \\
 H_b[r_i, r_{i-1}, \phi_i, \theta_i] &= \frac{\mu}{2} \dot{r}_i^2 + V_1[r_i] + V_2[r_i, r_{i-1}, \phi_i, \theta_i] - F_{ex} d \cos\left(\sum_{k=1}^{i-1} \phi_k\right).
 \end{aligned} \tag{3}$$

$H_a[r_1]$ is taken out of the sum as the first base pair has no preceding neighbor along the molecule backbone. Path integral techniques can be applied to Eq. (3) to derive the partition function Z_N following a well-tried method [71]. The latter assumes that the base pair distance r_i in Eq. (3) is a trajectory $r_i(\tau)$ depending on the imaginary time (τ) defined

as the analytic continuation of the real time to the imaginary axis [72]. Accordingly $r_i(\tau)$ is expanded in Fourier series whose coefficients generate an ensemble of possible base pair trajectories. As this is done for any base pair in the chain, the technique permits building an ensemble of molecule configurations which statistically contribute to the path integral with a Boltzmann weight. Technically, the size of the ensemble, i.e., the number of trajectories, is increased until numerical convergence in Z_N is achieved. This condition corresponds to the state of thermodynamic equilibrium for the system. Including also fluctuations over the bending and twisting degrees of freedom, with their respective cutoffs, the general expression for Z_N reads:

$$\begin{aligned}
Z_N &= \oint Dr_1 \exp[-A_a[r_1]] \prod_{i=2}^N \int_{-\phi_M}^{\phi_M} d\phi_i \int_{-\theta_M}^{\theta_M} d\theta_i \oint Dr_i \exp[-A_b[r_i, r_{i-1}, \phi_i, \theta_i]], \\
A_a[r_1] &= \int_0^\beta d\tau H_a[r_1(\tau)], \\
A_b[r_i, r_{i-1}, \phi_i, \theta_i] &= \int_0^\beta d\tau H_b[r_i(\tau), r_{i-1}(\tau), \phi_i, \theta_i], \tag{4}
\end{aligned}$$

where β is the inverse temperature. $\oint Dr_i$ is the measure of integration over the space of the Fourier coefficients whose temperature dependent cutoffs are consistently defined in the path integral method [73, 74]. While this peculiar feature makes the method also suitable to study thermally driven effects such as bubble formation and DNA denaturation [40, 75], the calculations hereafter presented are carried out at room temperature. From Eq. (4), the free energy of the system is computed as: $F = -\beta^{-1} \ln Z_N$.

4. Computational Method

The maximum amplitude of the bending fluctuations in Eq. (4) is set by ϕ_M which at a first stage is taken constant, i.e., $\phi_M = \pi/2$. The latter choice allows for kinks formation as a possible mechanism, first put forward long time ago [76], to reduce the bending energy between adjacent base pairs [77]. This has also been suggested more recently by molecular dynamics simulations [78] and analysis of (un)looping probabilities in short chains based on single molecule fluorescence resonance energy transfer [79]. While the polymer flexibility can be directly related to the average bending angles between any two distant monomers

[80, 81], computation of persistence length has shown that linear short sequences have an intrinsic flexibility mainly ascribed to their terminal base pairs [42, 82].

The novelty brought about here lies in the way twisting fluctuations are treated for the model drawn in Fig. 1(b). Our idea is that the twist variable θ_i in Eq. (4) should be measured with respect to the ensemble averaged $\langle \theta_{i-1} \rangle$ value obtained for the preceding base pair along the stack. Such value is incremented by $2\pi/h_r$, where the number of base pairs per helix turn is taken as an input parameter to be chosen within an appropriate range. For any assumed h_r , we further admit there may be a fluctuational range around the value $\langle \theta_{i-1} \rangle + 2\pi/h_r$, with θ_i^{fl} being the twist fluctuation integration variable. Formally our Ansatz is expressed by:

$$\begin{aligned}\theta_i &= \langle \theta_{i-1} \rangle + 2\pi/h_r + \theta_i^{fl}, \\ h_r &\in [h_r^{min}, h_r^{max}], \\ h_r^{max} - h_r^{min} &= n \cdot \Delta h_r \quad n \in \mathbb{Z},\end{aligned}\tag{5}$$

where Δh_r is the increment in the helical repeat range. The accuracy of the method grows with the number n of values sampled in such interval. For kilo-base B-DNA in solution the experimental helical repeat under physiological condition is, $h_r^{exp} = 10.4$ [83], whereas a smaller value $h_r \sim 10$ has been estimated for DNA wrapped around histone proteins [84, 85]. In general, thermal fluctuations tend to untwist the helix [86]. In short chains the helical repeat may differ from h_r^{exp} and certainly it changes under the action of applied forces. Accordingly, our strategy consists of assuming a broad range around h_r^{exp} , e.g., $h_r^{min} = 6$, $h_r^{max} = 14$, with a coarse partition ($\Delta h_r = 0.125$) and searching for the energetically most convenient helical conformations as a function of the external force. Next we take a finer partition of the range (enhancing both n and the CPU time) and repeat the search until the selected helical conformations are stable against Δh_r .

Central to this work is the calculation of the average twist angles formally given by:

$$\langle \theta_i \rangle_{(i \geq 2)} = \langle \theta_{i-1} \rangle + \frac{2\pi}{h_r} + \frac{\int_{-\theta_M}^{\theta_M} d\theta_i^{fl} \cdot (\theta_i^{fl}) \int_{-\phi_M}^{\phi_M} d\phi_i \oint Dr_i \exp[-A_b[r_i, r_{i-1}, \phi_i, \theta_i]]}{\int_{-\theta_M}^{\theta_M} d\theta_i^{fl} \int_{-\phi_M}^{\phi_M} d\phi_i \oint Dr_i \exp[-A_b[r_i, r_{i-1}, \phi_i, \theta_i]]}.\tag{6}$$

Numerical convergence on the average twist angles is achieved by setting the cutoff on the twist fluctuations, $\theta_M = \pi/4$. For the first base pair in the chain, $\theta_1 \equiv \theta_1^{fl}$ hence, $\langle \theta_1 \rangle = 0$.

Then, from Eq. (6), we derive the average helical repeat:

$$\langle h_r \rangle = \frac{2\pi N}{\langle \theta_N \rangle}. \quad (7)$$

Computing Eqs. (6), (7) for any h_r in Eq. (5), the program generates a set of values $\{\langle h_r \rangle_j, (j = 1, \dots, n)\}$ which, in general, deviate from the initially chosen h_r -values of our simulation. After obtaining such set, we select by free energy minimization the equilibrium value $\langle h_r \rangle_{j^*}$. As this procedure is iterated for any applied force, the twist-stretch profile for a specific molecule is eventually derived.

Next, integrations over the ensemble of base pair configurations are performed to get:
1) the average base pair radial fluctuations which determine the average helix diameter, $R_0 + \langle R \rangle$ [87],

$$\begin{aligned} \langle r_i \rangle &= \frac{\oint Dr_i \cdot r_i \int_{-\theta_M}^{\theta_M} d\theta_i^{fl} \int_{-\phi_M}^{\phi_M} d\phi_i \exp[-A_b[r_i, r_{i-1}, \phi_i, \theta_i]]}{\int_{-\theta_M}^{\theta_M} d\theta_i^{fl} \int_{-\phi_M}^{\phi_M} d\phi_i \oint Dr_i \exp[-A_b[r_i, r_{i-1}, \phi_i, \theta_i]]}, \\ \langle R \rangle &= \frac{1}{N} \sum_{i=1}^N \langle r_i \rangle \end{aligned} \quad (8)$$

and 2) the average distances between base pairs along the stack,

$$\begin{aligned} \langle \overline{d_{i,i-1}} \rangle &= \frac{\oint Dr_i \int_{-\theta_M}^{\theta_M} d\theta_i^{fl} \int_{-\phi_M}^{\phi_M} d\phi_i \cdot \overline{d_{i,i-1}} \exp[-A_b[r_i, r_{i-1}, \phi_i, \theta_i]]}{\int_{-\theta_M}^{\theta_M} d\theta_i^{fl} \int_{-\phi_M}^{\phi_M} d\phi_i \oint Dr_i \exp[-A_b[r_i, r_{i-1}, \phi_i, \theta_i]]}, \\ \langle d \rangle &= \frac{1}{N-1} \sum_{i=2}^N \langle \overline{d_{i,i-1}} \rangle. \end{aligned} \quad (9)$$

From Eq. (9), one finally estimates the force induced extension per base pair with respect to the un-stretched base pair separation:

$$\Delta z = \frac{\langle d(F_{ex}) \rangle}{\langle d(0) \rangle} - 1. \quad (10)$$

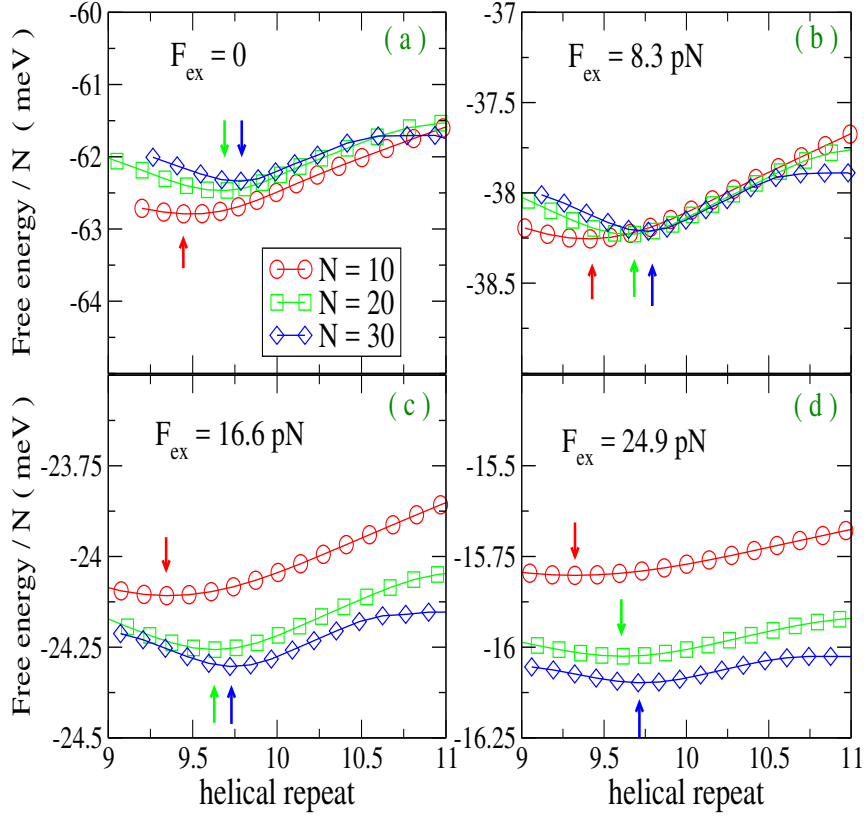


FIG. 2: (Color online) Free energies per base pair, calculated for three homogeneous sequences, as a function of the average helical repeat given by Eq. (7). The room temperature free energies are computed over a range of helical repeat values with incremental step $\Delta h_r = 0.125$ (see text). The four panels refer to the zero force case and three values of applied force. The arrows mark, for each sequence, the free energy minimum.

The $\langle R \rangle$'s and Δz 's have been calculated for any helical repeat in Eq. (5) but the values reported hereafter are those which correspond to the equilibrium conformation $\langle h_r \rangle_{j^*}$ selected by the free energy minimum.

5. Results

Our method is illustrated in Fig. 2 where the free energy per base pair (F/N) is plotted versus $\langle h_r \rangle$ for three short homogeneous sequences, both in the absence of force and for three increasing F_{ex} strengths. An increment $\Delta h_r = 0.125$ has been assumed in these calculations.

Given our choice of model parameters, the free energy scale for $F_{ex} = 0$ (panel (a)) shows that F/N is in the range of the experimental values, i.e., $1 - 1.5 \text{ kcal/mol}$ [57]. This holds for all sequences. Applied forces have the general effect to straighten the coiled molecules thus reducing the entropy. It follows that F/N should become larger by enhancing F_{ex} . Consistently the free energy scales display growing (less negative) values in the three successive panels. Also note that F/N increases more rapidly for shorter molecules and, for sizeable F_{ex} , it gets distinctly higher for the $N = 10$ molecule as shown in panels (c), (d). This indicates that external forces more effectively straighten the bonds in shorter molecules. The arrows mark, for each sequence, the free energy minima occurring at the above defined $\langle h_r \rangle_{j^*}$. Such values shift downwards (towards smaller helical repeats) by increasing F_{ex} from (a) to (d), pointing to a general helix over-twisting as a consequence of the mechanical stretching.

These findings are made more evident in Figs. 3 where the $\langle h_r \rangle_{j^*}$ are directly reported as a function of the mechanical force. For the three sequences, $N = 10, 20, 30$, in Fig. 3(a), the computation is first carried out with the same coarse partition of the helical repeat range as in Fig. 2.

It turns out that the $\langle h_r \rangle_{j^*}$'s generally decrease versus F_{ex} but, in the low forces regime, the plots display a non-monotonic behavior (empty symbols). However, reducing Δh_r , a monotonic dependence on F_{ex} is recovered for all sequences as visualized by the filled symbols. Only the $\langle h_r \rangle_{j^*}$'s calculated for $F_{ex} \sim 4, 8 \text{ pN}$ are reported in the figure as, at higher forces, the $\langle h_r \rangle_{j^*}$'s do not substantially change by taking a finer mesh. In general, a finer partition of the range in Eq. (5) should be used to achieve numerical convergence in the low force regime in which the helical repeat has a larger gradient. Instead, applying larger forces which significantly stretch the molecule, the $\langle h_r \rangle_{j^*}$'s become more stable and the transitions among twist conformations with different energy are reduced. Hence, convergence is already achieved using a coarser Δh_r mesh.

Increasing F_{ex} has the effect to dampen the thermal fluctuations and set the molecules in a more ordered conformation. This explains why the search for the equilibrium value $\langle h_r \rangle_{j^*}$ can be made with a coarser resolution at larger applied loads. It is also found that, for longer sequences, smaller Δh_r 's are necessary in order to determine with accuracy the $\langle h_r \rangle_{j^*}$'s at low external forces.

The interplay between sequence length and Δh_r is further pointed out in Fig. 3(b) for a

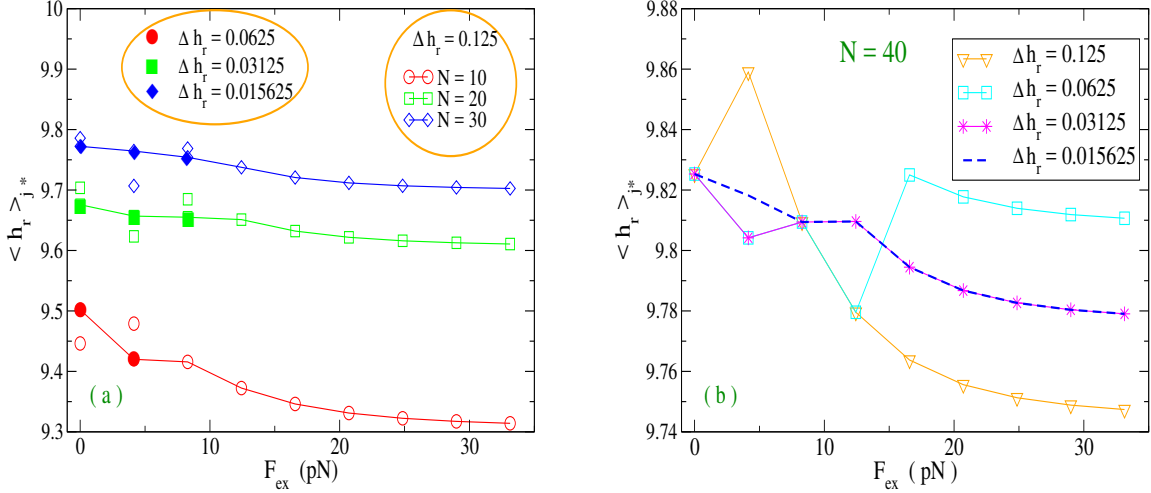


FIG. 3: (Color online) (a) Ensemble averaged helical repeats as a function of the external force, calculated via Eqs. (6), (7) for three sequences. The empty symbols refer to a coarse partition in Eq. (5). In the low forces regime, finer partitions (smaller Δh_r) are required to obtain numerical convergence. (b) As in (a) but for a longer sequence. The average helical repeats are computed by tuning Δh_r .

$N = 40$ sequence. In this case, also at large applied forces, the computation converges only with a fine partition, $\Delta h_r = 0.03125$, while a finer step, $\Delta h_r = 0.015625$, is required to get convergent and monotonic behavior at small F_{ex} . Summing up, the relevant feature put forward by our plots is that short DNA sequences over-twist under mechanical stretching and these structural deformations appear as a response dictated by energetic convenience. While this result confirms the trend suggested by the measurements of ref.[27] albeit on kilo-base long molecules, it would be interesting to test whether chains of a few tens of base pairs behave in a similar way. Besides the length, also sequence specificities may affect mechanical properties and twisting flexibility in heterogeneous chains [88].

Interestingly however, from Figs. 3, one can also estimate the force induced change in superhelical density σ . For sequences with planar molecular axis, $\sigma \equiv (Tw - Tw_0)/Tw_0$, where Tw_0 is the twist number in the absence of forces. Looking for instance at the plot for $N = 30$ and taking the computed $\langle h_r \rangle_{j^*}$ at $F_{ex} \sim 20$ pN, we obtain $\sigma \sim 0.0076$ which is in fair agreement with the experimental $\sigma \sim 0.01$ at $F_{ex} \sim 18$ pN [27] measured in long sequences.

Physically, the shrinking of the helix diameter could be associated to the shortening of the radial base pair fluctuations induced by the external force. All-atom simulations of stretched and compressed DNA structures have previously suggested that helix diameter reduction and over-twisting are concomitant effects [89]. We address this point in Figs. 4 which display for the four short chains: (a) the free energy per base pair calculated, via Eq. (4), at the helical equilibrium values $\langle h_r \rangle_{j^*}$ and (b) the average radial fluctuations, superimposed to the diameter R_0 , obtained from Eq. (8). As noticed above, the free energies rapidly increase under the ordering effect of the applied load and, on this scale, the plots for the four chains essentially overlap. However, the growth rate of F/N decreases versus F_{ex} and, markedly, above ~ 20 pN. Likewise, once F_{ex} is applied, $\langle R \rangle$ quickly drops from its value at zero force and, for all chains, tends to saturate above ~ 20 pN. Nevertheless, intrinsic base pair fluctuations persist also in the presence of large external loads. Then, we find that the DNA over-twisting and the reduction of the helix diameter are indeed correlated structural properties also at the very short length scales considered in this work. Molecular models based on energy minimization have suggested that the shrinking of the DNA diameter could be associated to a negative inclination of the base pairs towards the minor groove [28, 90].

As mentioned in the Introduction, experiments indicate that stretched kilo-base long DNA first over-twists and eventually untwists if the applied loads exceeds ~ 30 pN. Instead, the results presented so far do not account for the helix untwisting at large F_{ex} as the $\langle h_r \rangle_{j^*}$'s decrease monotonically versus F_{ex} in Figs. 3. The source of this discrepancy lies however in the way we have treated the bending fluctuations and precisely their maximum amplitude ϕ_M at the beginning of Section 4. Once the molecule is stretched, the intra-strand bonds straighten and therefore it is likely that the amplitude of the bending fluctuations between adjacent nucleotides is reduced whereas there is no physical reason to impose a similar constraint on the twisting fluctuations which may remain large also for sizeable external loads. Accordingly, a consistent model should contemplate a force dependent integration cutoff $\phi_M(F_{ex})$. Although we are not aware of experiments providing data to which we may fit such function, some plausible functional forms have been guessed to test their effects on the output of our numerical code. Taking for instance,

$\phi_M(F_{ex}) = \pi[1 - (c \cdot F_{ex})^2]/2$ with $c^{-1} = 24$ pN, we obtain the results displayed in Fig. 5 for a $N = 10$ sequence. Similar plots are obtained for longer chains. The computation follows

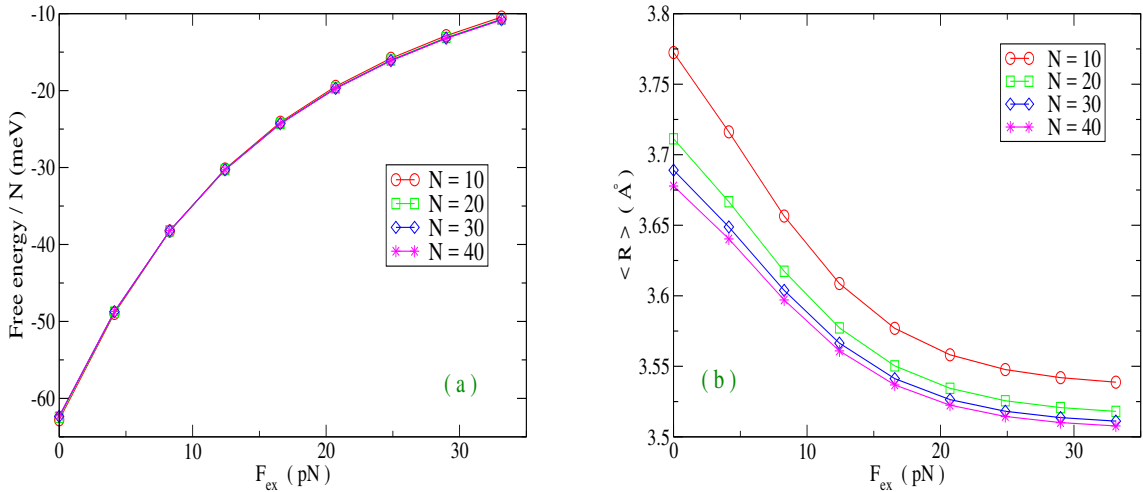


FIG. 4: (Color online) (a) Room temperature free energies per base pair, calculated for four sequences, as a function of the external force. The free energy per base pair values correspond to the minima indicated by arrows in Fig. 2. Note that such values essentially overlap for various N due to the enlarged energy scale taken here respect to Fig. 2. (b) Ensemble averaged base pair fluctuations, calculated via Eq. (8), as a function of the external force for four sequences.

the same pathway described above. The transition between over-twisting and untwisting regime is now found and it occurs at ~ 4 pN. Consistently, the average radial fluctuations (panel (a)) decrease up to ~ 4 pN and expand once the helix untwists under larger loads. While the latter force value is only indicative and clearly depends on the specific $\phi_M(F_{ex})$, there is no cogent reason why the transition at ~ 30 pN, observed for kilo-base sequences [27], should take place also for short sequences in the same force regime. In this regard, experimental research may clarify the behavior of short DNA under stretching and foster further theoretical investigation. However, the fact that the helix diameter and twisting conformations display a non-monotonic behavior versus the applied force corroborates our choice to impose a force dependent cutoff only on the bending fluctuations.

Likewise the intra-strand base pair separation (panel (c)), calculated via Eqs. (9), (10), slightly contracts with respect to the zero force conformation in the over-twisting regime and then stretches above ~ 4 pN. Thus we find that, under the action of a *variable force*, the end-to-end distance grows when the helix untwists and, vice-versa, it shortens when the helix over-twists. While this pattern seems physically intuitive, it is only apparently at

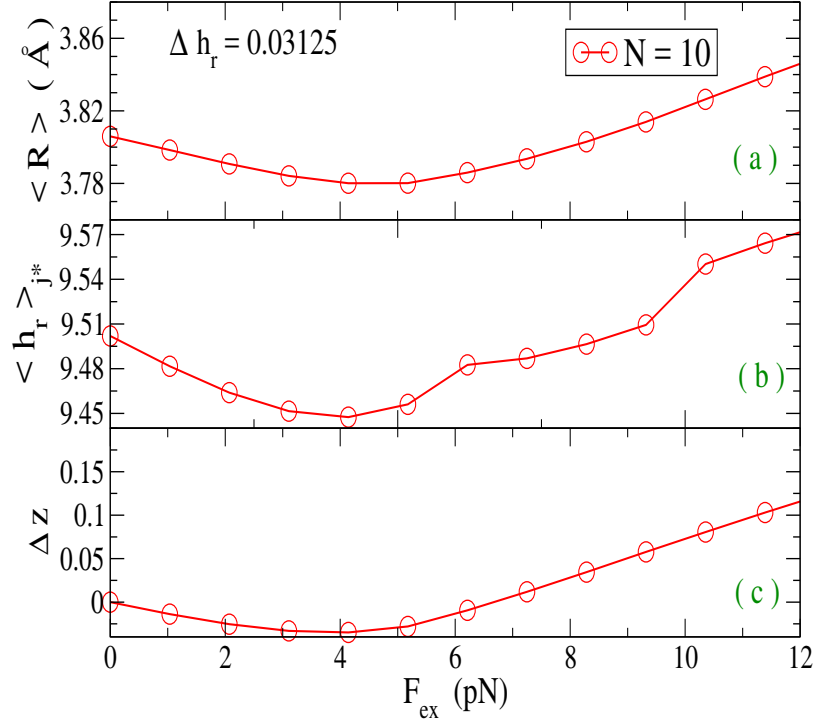


FIG. 5: (Color online) (a) Average radial fluctuations calculated from Eq. (8) versus the external force F_{ex} ; (b) Average helical repeat obtained from Eqs. (6), (7) after minimizing the free energy for any F_{ex} ; (c) Average intra-strand extension per base pair from Eqs. (9), (10). Both (a) and (c) are calculated for the equilibrium helical conformations $\langle h_r \rangle_{j^*}$ in (b). The amplitude of the bending fluctuations is assumed to be dependent on F_{ex} as described in the text.

variance with ref.[27] as in the latter the experimental setup was such that the changes in DNA extension due to an imposed over-twist had been measured at *constant force*.

To test whether our program reproduces the experimental trend also on this specific point, we should assume a constant force, say F'_{ex} , and study the average base pair distance $\langle d \rangle$ as a function of $\langle h_r \rangle$. The experimentally imposed over-twist is simulated by considering the $\langle h_r \rangle$'s which are smaller than the value $\langle h_r \rangle_{j^*}$ for the equilibrium conformation in the presence of F'_{ex} . Note that such set of $\langle h_r \rangle$'s has been computed *before* selecting $\langle h_r \rangle_{j^*}$ by free energy minimization. Clearly, any assumed $\langle h_r \rangle$ differing from $\langle h_r \rangle_{j^*}$ corresponds to a condition in which the system is away from its free energy minimum. In

this case we find (not shown here) that $\langle d \rangle$ grows by reducing $\langle h_r \rangle$ hence, the helical extension increases upon over-twisting consistently with the measurements. At the same time $\langle R \rangle$ decreases by reducing $\langle h_r \rangle$. Thus, short dsDNA is predicted to shrink and extend when over-twisted under a constant load. Further quantitative analysis of the twist-extension relation may be performed on the base of models which account for the sequence specificities of the double helical structure.

6. Conclusions

We have developed a method to evaluate the deformations of the DNA helical structure induced by an external load whose strength is such to compete on the energy scale with the base pair thermal fluctuations. Applied forces in the range of a few pico-Newton may in fact straighten the intra-strand molecule bonds and dampen the amplitudes of the base pair radial distances. Motivated by ongoing research on the properties of short DNA helices, we have analyzed the interplay between helix stretching and twisting using a mesoscopic Hamiltonian model which sees the double stranded molecule as a sequence of interacting base pairs. This discrete approach seems particularly suitable to DNA molecules whose contour length may be shorter than their persistence length. In fact, at such short scales, the applicability of continuum elastic rod models has been questioned over the last years. The statistical mechanics of the mesoscopic Hamiltonian has been formulated by a well-established finite temperature path integration method which treats the inter-strand and intra-strand interactions in terms of the ensemble of trajectories for the base pair radial configurations. Also bending and twisting fluctuations between neighboring base pairs along the molecule backbone are included in the partition function.

Introducing a computational scheme which allows the molecule to assume in principle a broad range of helical conformations, we have shown that the DNA chain generally reduces the number of base pairs per helix turn under the effect of a mechanical stretching. While the molecule over-twists in order to minimize its free energy, this behavior appears physically correlated to the shrinking of the helix diameter which is in fact consistently obtained by our method.

These results agree with the experimental behavior, observed for kilo-base long chains, in the low to intermediate force regime. Instead, for applied forces larger than 30 pN, exper-

iments show that long chains eventually untwist. To check whether the transition between over-twisting and untwisting regimes could be predicted by our model, we have considered that the amplitude of the intra-strand bending fluctuations may depend on the external force. Assuming a physically plausible force dependent cutoff on the bending fluctuation integral, we have indeed found that also short sequences may first over-twist and then untwist by enhancing the external load. Accordingly, when the molecule untwists, its average diameter expands and the average intra-strand base pair distance grows. Although, in such short chains, the transition may occur at weaker forces than those measured in kilo-base DNA molecules, these findings suggest an intrinsic correlation between bending and twisting degrees of freedom at any length scale. However, at this stage, we do not have specific experimental information to establish the possible, quantitative dependence of the bending amplitude on the applied forces.

We have here chosen a set of model parameters appropriate to homogeneous chains of GC-base pairs as discussed in previous studies. While the values presented in this paper for the average helical repeat and helix diameter versus applied force have some dependence on the specific input parameters, nevertheless the overall trend of our conclusions, summarized in Fig. 5 for a very short sequence, is not affected by the specific parameter choice. On the base of the presented results we thus believe that our computational method may offer a robust tool to investigate a large variety of molecule conformations subject to external perturbations and also model the DNA response upon protein binding.

-
- [1] Travers A A and Thompson J M T 2004 *Phil. Trans. R. Soc. Lond. A*, **362**, 1265-1279
 - [2] Apostolaki A and Kalosakas G 2011 *Phys. Biol.*, **8**, 026006
 - [3] Cherstvy A G 2011 *Phys. Chem. Chem. Phys.*, **13**, 9942-9968
 - [4] Marko J F 2015 *Physica A*, **418**, 126-153
 - [5] Radding C M 1991 *J. Biol. Chem.*, **266**, 5355-5358
 - [6] Cox M M 1991 *Mol. Microbiol.*, **5**, 1295-1299
 - [7] De Vlaminck I, van Loenhout M T J, Zweifel L, den Blanken J, Hooning K, Hage S, Kersse-makers J and Dekker C 2012 *Mol. Cell*, **46**, 616-624
 - [8] Forget A L and Kowalczykowski S C 2012 *Nature*, **482**, 423

- [9] Kiiianitsa K and Stasiak A 1997 *Proc. Natl. Acad. Sci. USA*, **94**, 7837-7840
- [10] Olson W K, Gorin A A, Lu X -J, Hock L M and Zhurkin V B 1998 *Proc. Natl. Acad. Sci. USA*, **95**, 11163-11168
- [11] Keenholtz R A, Grindley N D F, Hatfull G F and Marko J F 2016 *Nucleic Acids Res.*, **44**, 8921-8932
- [12] Chu S 1991 *Science*, **253**, 861-866
- [13] Smith S, Finzi L and Bustamante C 1992 *Science*, **258**, 1122-1126
- [14] Cluzel P, Lebrun A, Heller C, Lavery R, Viovy J L, Chatenay D and Caron F 1996 *Science*, **271**, 792-794
- [15] Wang M D, Yin H, Landick R, Gelles J and Block S M 1997 *Biophys. J.*, **72**, 1335-1346
- [16] Chou F-C, Lipfert J, Das R 2014 *PLoS Comput. Biol.* **10**, e1003756
- [17] Bustamante C, Marko J F, Siggia E D and Smith S 1994 *Science*, **265**, 1599-1601
- [18] Odijk T 1995 *Macromolecules*, **28**, 7016-7018
- [19] Smith S B, Cui Y and Bustamante C 1996 *Science*, **271**, 795-798
- [20] Marko J F 1997 *Europhys. Lett.*, **38**, 183-188
- [21] Strick T R, Allemand J F, Bensimon D and Croquette V 1998 *Biophys. J.*, **74**, 2016-2028
- [22] Storm C and Nelson P C 2003 *Phys. Rev. E*, **67**, 051906
- [23] Rouzina I and Bloomfield V A 2001 *Biophys. J.*, **80**, 882-893; *ibid.*, **80**, 894-900
- [24] Cassuto E and Howard-Flanders P 1986 *Nucleic Acids Res.*, **14**, 1149-1157
- [25] Léger J F, Romano G, Sarkar A, Robert J, Bourdieu L, Chatenay D and Marko J F 1999 *Phys. Rev. Lett.*, **83**, 1066-1069
- [26] van Mameren J, Gross P, Farge G, Hooijman P, Modesti M, Falkenberg M, Wuite G J L and Peterman E J G 2009 *Proc. Natl. Acad. Sci. USA*, **106**, 18231-18236
- [27] Gore J, Bryant Z, Nöllmann M, Le M U, Cozzarelli N R and Bustamante C 2006 *Nature*, **442**, 836-839
- [28] Lionnet T, Joubaud S, Lavery R, Bensimon D and Croquette V 2006 *Phys. Rev. Lett.*, **96**, 178102
- [29] Duričković B, Goriely A, Maddocks J H 2013 *Phys. Rev. Lett.*, **111**, 108103
- [30] Gross P, Laurens N, Oddershede L B, Bockelmann U, Peterman E J G and Wuite G J L 2011 *Nature Phys.*, **7**, 731736
- [31] Yuan C, Chen H, Lou X W and Archer L A 2008 *Phys. Rev. Lett.*, **100**, 018102

- [32] Noy A and Golestanian R 2012 *Phys. Rev. Lett.*, **109**, 228101
- [33] Wu Y Y, Bao L, Zhang X and Tan Z J 2015 *J. Chem. Phys.*, **142**, 125103
- [34] Mathew-Fenn R S, Das R and Harbury P A B 2008 *Science*, **322**, 446-449
- [35] Mastroianni A J, Sivak D A, Geissler P L and Alivisatos A P 2009 *Biophys. J.*, **97**, 1408-1417
- [36] Mazur A K and Maaloum M 2014 *Phys. Rev. Lett.*, **112**, 068104
- [37] Kim Y -J and Kim D -N 2016 *PLoS ONE*, **11**, e0153228
- [38] Zoli M 2009 *Phys.Rev. E*, **79**, 041927
- [39] Zoli M 2010 *Phys.Rev. E*, **81**, 051910
- [40] Zoli M 2011 *J. Chem. Phys.*, **135**, 115101
- [41] Zoli M 2016 *J. Chem. Phys.*, **144**, 214104
- [42] Zoli M 2016 *Phys. Chem. Chem. Phys.*, **18**, 17666
- [43] Horowitz D S, Wang J C 1984 *J. Mol. Biol.*, **173**, 75-91
- [44] Fogg J M, Kolmakova N, Rees I, Magonov S, Hansma H, Perona J J, and Zechiedrich E L 2006 *J. Phys.: Condens. Matter*, **18**, S145S159
- [45] Irobalieva R N, Fogg J M, Catanese D J, Sutthibutpong T, Chen M, Barker A K, Ludtke S J, Harris S A, Schmid M F, Chiu W and Zechiedrich L 2015 *Nat. Commun.* **6**, 8440
- [46] Bates A D and Maxwell A 2009 *DNA Topology* (Oxford University Press, Oxford)
- [47] Rossetti G, Dans P D, Gomez-Pinto I, Ivani I, Gonzalez C and Orozco M 2015 *Nucleic Acid Res.*, **43**, 43094321
- [48] Li G, Shen H, Zhang D, Li Y, and Wang H 2016 *J. Chem. Theory Comput.*, **12**, 676-693
- [49] Szatyłowicz H and Sadlej-Sosnowska N 2010 *J. Chem. Inf. Model.*, **50**, 21512161
- [50] Guéron M, Kochoyan M and Leroy J L 1987 *Nature*, **328**, 89-92
- [51] Fye R M and Benham C J 1999 *Phys. Rev. E*, **59**, 3408-3426
- [52] Altan-Bonnet G, Libchaber A and Krichevsky O 2003 *Phys. Rev. Lett.*, **90**, 138101
- [53] Chen C and Russu I M 2004 *Biophys. J.*, **87**, 2545-2551
- [54] Rapti Z, Smerzi A, Rasmussen K Ø, Bishop A R, Choi C H and Usheva A 2006 *Phys. Rev. E*, **73**, 051902
- [55] Metzler R, Ambjörnsson T, Hanke A and Fogedby H C 2009 *J. Phys.: Condens. Matter* **21** 034111
- [56] Alexandrov B, Voulgarakis N K, Rasmussen K Ø, Usheva A and Bishop A R 2009 *J. Phys.: Condens. Matter* **21**, 034107

- [57] Krueger A, Protozanova E and Frank-Kamenetskii M D 2006 *Biophys. J.*, **90**, 3091-3099
- [58] Porschke D 1991 *Biophys. Chem.*, **40**, 169
- [59] Padinhateeri R and Menon G I 2013 *Biophys. J.*, **104**, 463-471
- [60] Brunet A, Tardin C, Salomé L, Rousseau P, Destainville N and Manghi M 2015 *Macromolecules*, **48**, 3641-3652
- [61] Salari H, Eslami-Mossallam B, Naderi S and Ejtehadi M R 2015 *J. Chem. Phys.*, **143**, 104904
- [62] Waters J T and Kim H D 2016 *Phys. Rev. E*, **93**, 043315
- [63] Gao Y, Devi-Prasad K V and Prohofski E W 1984 *J. Chem. Phys.*, **80**, 6291
- [64] Joyeux M and Florescu A -M 2009 *J. Phys.: Condens. Matter* **21**, 034101
- [65] Weber G, Haslam N, Essex J W and Neylon C 2009 *J. Phys.: Condens. Matter* **21**, 034106
- [66] Baird N C 1974 *Int. J. Quantum Chem.* , **8**, (S1) 49-54
- [67] Dauxois T, Peyrard M and Bishop A R 1993 *Phys. Rev. E*, **47**, R44-47
- [68] Zoli M 2012 *J. Phys.: Condens. Matter*, **24**, 195103
- [69] Lambropoulos K, Chatzieftheriou M, Morphis A, Kaklamanis K, Theodorakou M, and Simserides C 2015 *Phys. Rev. E*, **92**, 032725
- [70] Lambropoulos K, Chatzieftheriou M, Morphis A, Kaklamanis K, Lopp R, Theodorakou M, Tassi M and Simserides C 2016 *Phys. Rev. E*, **94**, 062403
- [71] Zoli M 2014 *J. Theor. Biol.*, **354**, 95-104
- [72] Feynman R P and Hibbs A R 1965 *Quantum Mechanics and Path Integrals*, (Mc Graw-Hill, New York)
- [73] Zoli M 2011 *Eur. Phys. J. E*, **34**, 68
- [74] Zoli M 2003 *Phys. Rev. B*, **67**, 195102
- [75] Zoli M 2013 *J. Chem. Phys.* , **138**, 205103
- [76] Crick F H and Klug A 1975 *Nature*, **255**, 530-533
- [77] Sanchez D S, Qu H, Bulla D and Zocchi G 2013 *Phys. Rev. E*, **87**, 022710
- [78] Mitchell J S, Laughton C A and Harris S A 2011 *Nucleic Acids Res.*, **39**, 3928-3938
- [79] Le T T and Kim H D 2014 *Nucl. Acids Res.*, **42**, 10786-10794
- [80] Ullner M, Jönsson B, Peterson C, Sommelius O and Söderberg B 1997 *J. Chem. Phys.*, **107**, 1279
- [81] Geggier S and Vologodskii A 2010 *Proc. Natl. Acad. Sci. USA*, **107**, 15421-15426
- [82] Ferreira I, Amarante T D and Weber G 2015 *J. Chem. Phys.*, **143**, 175101

- [83] Wang J C 1976 *Proc. Natl. Acad. Sci. USA*, **76**, 200-203
- [84] Levitt M 1978 *Proc. Natl. Acad. Sci. USA*, **75**, 640-644
- [85] Hayes J J, Tullius T D and Wolffe A P 1990 *Proc. Natl. Acad. Sci. USA*, **87**, 7405-7409
- [86] Duguet M 1993 *Nucleic Acids Res.*, **21**, 463-468
- [87] For the first base pair in the chain, $\langle r_1 \rangle$ is calculated with the Boltzmann weight factor associated to the action $A_a[r_1]$ in Eq. (4) which does not depend on angular variables.
- [88] Peters J P, Yelgaonkar S P, Srivatsan S G, Tor Y and Maher L J 2013 *Nucleic Acid Res.*, **41**, 10593-10604
- [89] Kosikov K M, Gorin A A, Zhurkin V B and Olson W K 1999 *J. Mol. Biol.*, **289**, 1301-1326
- [90] Liebl K, Drsata T, Lankas F, Lipfert J and Zacharias M 2015 *Nucleic Acid Res.*, **43**, 10143-10156