Definition of the persistence length in the coarse-grained models of DNA elasticity

Fathizadeh, A.; Eslami Mossallam, B.; Ejtehadi, M.R.

Definition of the persistence length in the coarse-grained models of DNA elasticity

A. Fathizadeh[*]

Sharif University of Technology, Institue for Nanoscience and Nanotechnology, Post Office Box 14588-89694, Tehran, Iran

B. Eslami-Mossallam[*]

Instituut-Lorentz for Theoretical Physics, Post Office Box 9506, 2300 RA Leiden, The Netherlands

M. R. Ejtehadi[†]

Sharif University of Technology, Department of Physics, Post Office Box 11155-9161, Tehran, Iran

By considering the detailed structure of DNA in the base pair level, two possible definitions of the persistence length are compared. One definition is related to the orientation of the terminal base pairs, and the other is based on the vectors which connect two adjacent base pairs at each end of the molecule. It is shown that although these definitions approach each other for long DNA molecules, they are dramatically different on short length scales. We show analytically that the difference mostly comes from the shear flexibility of the molecule and can be used to measure the shear modulus of DNA.

I. INTRODUCTION

The DNA molecule is involved in many fundamental life processes and is usually found in highly compact configurations in the cell, where it is significantly bent at short length scales. Therefore, it is expected that the nanomechanical behavior of DNA contributes strongly to its biological functions. As a result, elasticity of the DNA molecule at short length scales, and especially its bending rigidity, has been an important subject of study in recent decades and still can be regarded as a highly progressive field of research [1,2].

Like that of any other polymer, the bending rigidity of DNA is quantified by its persistence length. For every pair of points along the DNA chain with a distance of L, the persistence length is defined by the equation $\langle \cos \theta(L) \rangle = \exp(-L/l_p)$, where $\theta(L)$ is the bending angle between tangents along the corresponding pair of points and l_p is the persistence length. In the well-known wormlike chain (WLC) model [3], where DNA is modeled as a continuous and inextensible rod, the bending angle is defined by the tangent unit vectors at the ends of the rod. This leads to a constant, length-independent persistence length [4].

Persistence length of DNA can be determined in DNA stretching experiments by fitting an elasticity model to the experimental data [5] or directly by measuring the bending angle along the tangent vectors of DNA in atomic force microscopy (AFM) images [6–8]. While the former method has been applied only to micron-sized DNA molecules, with high-resolution AFM it is possible to study the bending elasticity of DNA on a nanoscale level. Although the WLC model successfully explains the elasticity of long DNA molecules, coarse-grained models of DNA at the base pair level, in which each base pair is regarded as a rigid body [9], are more suitable for short DNA molecules at the nanoscale level. In these coarse-grained models, the bending angle can be defined

as the angle between two unit vectors at the two ends of the DNA molecule. Two different ways to define the unit tangent vectors are depicted in Fig. 1(a). One option is to consider the normal vectors to the terminal base pairs [dashed red (gray) vectors in Fig. 1(a)], and the other is to use the center-to-center connective vectors of two adjacent base pairs at the two ends of DNA to define the unit vectors [solid blue (gray) vectors in Fig. 1(a)]. The former definition is not applicable in AFM studies, where the direction of base pairs is not detectable. However, a generalization of the latter definition can be related to the AFM experiments. In this paper we show that the first definition leads to a length-independent persistence length, which can be found from large-scale experiments, while the persistence length in the second definition depends on DNA length. We show analytically that this difference comes from contributions of shear elasticity to the latter definition, as well as the resolution of the coarse-grained model. Our findings point out some important considerations for analyzing AFM experiments on DNA elasticity.

II. MATERIALS AND METHOD

In the coarse-grained model at the base pair level, a local coordinate system with an orthonormal basis is attached to each rigid base pair. For the kth base pair, this basis is denoted by $\{\hat{d}_1(k), \hat{d}_2(k), \hat{d}_3(k)\}$, where $\hat{d}_3$ is perpendicular to the base pair, while $\hat{d}_1$ and $\hat{d}_2$ lie in the base pair plane (see Fig. 2). Three translational degrees of freedom, as well as three rotational degrees of freedom, are assigned to each base pair [9]. It is convenient to define the degrees of freedom of the base pairs from their relative position and orientation. The relative rotational degrees of freedom are called tilt, roll, and twist, and for the kth base pair step, we denote them by $\Theta_1(k)$, $\Theta_2(k)$, and $\Theta_3(k)$. The parameters $\rho_1(k)$, $\rho_2(k)$, and $\rho_3(k)$ describe the translational degrees of freedom of the k-base pair step and are called shift, slide, and rise respectively. They are the projections of the relative translation vector $\vec{r}(k)$ on the midframe coordinate system $\{\hat{d}_1^M(k), \hat{d}_2^M(k), \hat{d}_3^M(k)\}$ [i.e., $\rho_i(k) = \vec{r}(k) \cdot \hat{d}_i^M(k)$, $i = 1, 2, 3$] [9].

---

[*]These authors contributed equally to this work.
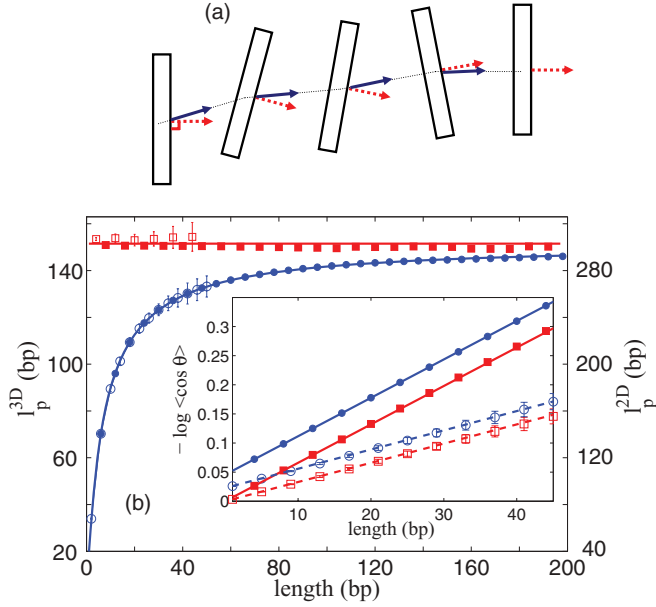[†]ejtehadi@sharif.edu

FIG. 1. (Color online) (a) Schematic representation of the two choices for defining the unit vectors. Rectangles represent the base pairs, dashed vectors (red [gray]) are normal to the base pairs, and solid vectors (blue [gray]) are along the lines that connect two adjacent base pairs. (b) Simulation data for l_p^n (squares, red [gray]) and l_p^c (circles, blue [gray]) as a function of DNA length, along with the theoretical results (solid and dashed curves). Filled symbols and solid curves show the simulation data and theoretical results in three dimensions (3D), while empty symbols and dashed curves correspond to two-dimensional (2D) DNA. The inset shows $-\ln\langle\cos\theta(L)\rangle$ vs L for the same data and with the same markers.

For a DNA with N base pair steps, the elastic energy of molecule in the simplest form is given by

$$\frac{E}{k_B T_r} = l_0 \sum_{k=1}^{N} \sum_{i=3}^{3} \left[\frac{A_i}{2l_0^2} \Delta\Theta_i^2(k) + \frac{K_i}{2l_0^2} \Delta\rho_i^2(k) \right], \quad (1)$$

with $\Delta\Theta_i = \Theta_i - \hat{\Theta}_i$ and $\Delta\rho_i = \rho_i - \hat{\rho}_i$, where k_B is the Boltzmann constant, T_r is the room temperature, and $\hat{\Theta}_i = \hat{\rho}_i = 0$ for $i = 1, 2$. In Eq. (1) we have $\hat{\rho}_3 = l_0$, where l_0 is the intrinsic rise, and $\hat{\Theta}_3 = l_0 \omega_0$, where ω_0 is the intrinsic twist of DNA per unit length. A_is and K_is are the rotational and
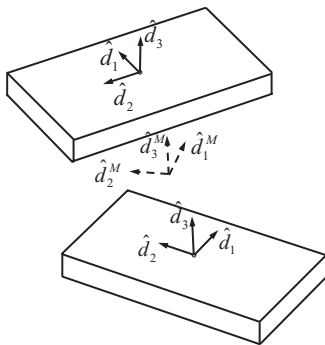

FIG. 2. Schematic representation of the local coordinates and midframe. Cuboids represent two adjacent base pairs. The local coordinates corresponding to each base pair and the midframe are shown with solid and dashed arrows respectively.

translational elastic constants respectively. In the limit where l_0 goes to zero with finite $L = Nl_0$, and the translational elastic constants go to infinity, the elastic energy in Eq. (1) reduces to the usual wormlike chain model [3].

The two definitions of the persistence length mentioned above can be written in terms of the components of the model as follows. If the normal vectors to the terminal base pairs are used to define the bending angle, we have

$$l_p^n = -Nl_0[\ln(\langle\hat{d}_3(1) \cdot \hat{d}_3(N+1)\rangle)]^{-1}, \quad (2)$$

where the superscript n stands for *normal*. Alternatively, if we connect adjacent base pairs at the two ends of DNA to define the terminal unit vectors, we can write the persistence length-denoted by l_p^c-as

$$l_p^c = -Nl_0[\ln(\langle\hat{u}(1) \cdot \hat{u}(N+1)\rangle)]^{-1}, \quad (3)$$

where $\hat{u}(k)$ is the unit vector along the line which connects the base pair k to $k+1$:

$$\hat{u}(k) = \frac{1}{|\vec{\rho}(k)|} \sum_{i=1}^{3} \rho_i(k) \, \hat{d}_i^m(k). \quad (4)$$

To demonstrate the difference, we performed simulations of a free DNA in the 3D space and a confined DNA on a 2D surface. In both simulations DNA was coarse grained at the base pair level, and the interactions between the base pairs are given by the elastic energy in Eq. (1). The elastic parameters of DNA were obtained by averaging the elastic parameters of all 10 possible base pair steps introduced in Becker *et al.* [10]. Equation (1) implies that, for a free DNA in 3D, the distribution functions of all local base pair step parameters are Gaussian, which can be generated numerically using a standard algorithm [11]. Using this method, we generated an ensemble of 1 million DNA molecules in the 3D space, each consisting of 300 base pairs, which obeys Maxwell-Boltzmann statistics at room temperature. Although this method is highly efficient in spanning the phase space of a free DNA in 3D, it is not applicable to a confined DNA in 2D space, since the distribution functions of the degrees of freedom will no longer be Gaussian when DNA is confined to a plane. Therefore, a coarse-grained molecular dynamics has been used to simulate 2D DNA molecules. Every base pair in these molecular dynamics (MD) simulations is considered as a rigid object with a geometry given in Ref. [12]. In addition to the elastic potential energy, two rigid walls at $z = \pm 0.1\text{Å}$ with hardcore potential are placed around the DNA to ensure that it is confined in the xy plane. This potential has a simple form of $\sum_{i=1,2} 1/r_i^{12}$ with r_i as the distance between the center of mass of each base pair and the ith rigid wall. The equations of motion for each base pair are integrated via a symplectic algorithm [13] and the Nose-Hoover chain thermostat is used to achieve the canonical ensemble at room temperature. After giving enough time for relaxation, we generated an ensemble of two-dimensionally confined DNA molecules. To calculate the persistence length in 2D, we projected every DNA chain in the ensemble on the 2D plane to mimic AFM experiments. Since we needed to generate a large number of chains, and the MD simulation was time-consuming, the simulation was only performed for DNA with 50 base pairs and the persistence length measurment only applied for distances up to this length.

III. RESULTS AND DISCUSSION

Figure 1(b) shows l_p^n (squares) and l_p^c (circles) as a function of DNA length obtained from the 2D and 3D simulations (empty and filled data points respectively). It is seen that l_p^n is constant, independent of DNA length, just as for the WLC model, while l_p^c is length dependent. The difference between the two definitions vanishes at large length scales. The figure also shows that $l_p^{2D} \approx 2l_p^{3D}$ for both definitions. The inset in Fig. 1(b) shows that $-\ln\langle\cos\theta(L)\rangle$ is a linear function of DNA length for both definitions. The lines have equal slopes, but the line corresponding to the latter definition does not pass the origin.

The two different behaviors shown in Fig. 1(b) can be understood in terms of general properties of the elastic energy. The elastic energy in Eq. (1) is invariant under the reflection transformation $\mathcal{R}$, which changes the signs of tilt and roll as well as shift and slide for all base pairs. It is straightforward to find the symmetry properties of the unit vectors $\{\hat{d}_i(k)\}$ and $\hat{u}(k)$ under the inversion transformation (see Appendix A). It follows that the basis vectors $\{\hat{d}_i(k)\}$ transform in such a way that $\hat{d}_i(k) \cdot \hat{d}_3(k+1)$ changes sign for $i = 1, 2$. Thus the invariance of the elastic energy under the reflection transformation implies that $\langle\hat{d}_1(k) \cdot \hat{d}_3(k+1)\rangle = \langle\hat{d}_2(k) \cdot \hat{d}_3(k+1)\rangle = 0$. Using this result as well as the locality property of the elastic energy, one can show that [4] $\langle\hat{d}_3(k_1) \cdot \hat{d}_3(k_2)\rangle = \langle\hat{d}_3(k_1) \cdot \hat{d}_3(k_3)\rangle\langle\hat{d}_3(k_3) \cdot \hat{d}_3(k_2)\rangle$, for all values of k_1, k_2, and k_3. Hence, using the translational invariance of the elastic energy, we obtain

$$\langle\hat{d}_3(1) \cdot \hat{d}_3(N+1)\rangle = [\langle\hat{d}_3(1) \cdot \hat{d}_3(2)\rangle]^N = \exp\left(-\frac{L}{l_p^n}\right),$$
(5)

where $L = Nl_0$ and

$$l_p^n = -l_0[\ln(\langle\hat{d}_3(1) \cdot \hat{d}_3(2)\rangle)]^{-1}.$$
(6)

l_p^n is a constant independent of DNA length, as suggested by Fig. 1(b). Since l_p^n is by definition independent of the translational degrees of freedom, and there is no coupling between rotational and translational degrees of freedom in Eq. (1), it can be seen that in the continuous limit l_p^n is equal to the WLC persistence length, l_p^{WLC}, which is given by $l_p^{WLC} = 2(A_1^{-1} + A_2^{-1})^{-1}$ [4].

By a similar argument, we can calculate the persistence length for the second definition, that is, Eq. (3). As a first step, we write the vectors $\hat{u}(1)$ and $\hat{u}(N+1)$ in terms of the basis $\{\hat{d}_i(2)\}$ and $\{\hat{d}_i(N)\}$ respectively. The symmetry properties of the vector $\hat{u}(k)$ under the reflection transformation $\mathcal{R}$ imply that $\langle\hat{u}(k) \cdot \hat{d}_1(k)\rangle = \langle\hat{u}(k) \cdot \hat{d}_2(k)\rangle = 0$ and also $\langle\hat{u}(k) \cdot \hat{d}_1(k+1)\rangle = \langle\hat{u}(k) \cdot \hat{d}_2(k+1)\rangle = 0$. Using this symmetry-related result (see Appendix A) as well as the locality property of the elastic energy, and replacing $\langle\hat{d}_3(2) \cdot \hat{d}_3(N+1)\rangle$ with $\exp(-\frac{L-l_0}{l_p^n})$ from Eq. (5), we obtain

$$\langle\hat{u}(1) \cdot \hat{u}(N+1)\rangle = \alpha \exp\left(-\frac{L}{l_p^n}\right),$$
(7)

where

$$\alpha = \exp\left(\frac{l_0}{l_p^n}\right)\langle\hat{u}(1) \cdot \hat{d}_3(1)\rangle\langle\hat{u}(1) \cdot \hat{d}_3(2)\rangle.$$
(8)

In writing Eq. (8), we also used the translational invariance of the elastic energy to replace $\langle\hat{u}(N+1) \cdot \hat{d}_3(N+1)\rangle$ with $\langle\hat{u}(1) \cdot \hat{d}_3(1)\rangle$. From Eqs. (3) and (7) it can be seen that the persistence length in the second definition, that is, l_p^c, depends on DNA length via the equation $\frac{1}{l_p^c} = \frac{1}{l_p^n} - \frac{\ln\alpha}{L}$.

Equations (5) and (7), with l_p^n being independent of DNA length, agree qualitatively with Fig. 1(b). To prove the quantitative agreement, we obtained an approximate analytical expression for α. Since large deformations are unlikely for a relaxed DNA, with the elastic energy given in Eq. (1), we expect $\bar{\bar{\Theta}} \equiv \bar{\Theta} - \langle\bar{\Theta}\rangle \ll 1$, and $\bar{\bar{\rho}} \equiv \frac{\bar{\rho} - \langle\bar{\rho}\rangle}{l_0} \ll 1$. Writing α in Eq. (8) in terms of the base pair degrees of freedom, and expanding the resultant expression in a Taylor series to second order in $\bar{\bar{\Theta}}$ and $\bar{\bar{\rho}}$, we obtain (see Appendix C)

$$\alpha = 1 - \frac{2}{l_0 K_{sh}} + \frac{l_0}{2 l_p^n},$$
(9)

where K_{sh} is the effective shear constant, defined as

$$\frac{1}{K_{sh}} = \frac{1}{2 l_0}(\langle\rho_1\rangle^2 + \langle\rho_2\rangle^2).$$
(10)

For a free DNA in 3D whose elastic energy is given by Eq. (1), we have $K_{sh} = 2(K_1^{-1} + K_2^{-1})^{-1}$.

Solid curves in Fig. 1(b) and its inset show the theoretical results for a 3D DNA [i.e., Eqs. (2), (3), (5), and (7)], where we have assumed $l_p^n = l_p^{WLC}$ and we calculated α from Eq. (9). It can be seen that there is a perfect agreement between theory and simulation.

Assuming $\widetilde{u}^{(m)}(k)$ is the unit vector along the line which connects kth base pair to the base pair $k + m$, $m \geqslant 1$, the definition of l_p^c can be generalized by using the vector $\widetilde{u}^{(m)}$ instead of $\hat{u}$ in Eq. (3). To calculate l_p^c in this generalized definition, we perform coarse graining of DNA in a higher level, so that every m base pairs are replaced with one *coarse-grained base pair*. The degrees of freedom of the coarse-grained base pairs are denoted by $\widetilde{\Theta}_i^{(l)}$ and $\widetilde{\rho}_i^{(l)}$, with $i = 1, 2, 3$, where the tilde ($\widetilde{}$) sign is used to distinguish the coarse-grained degrees of freedom from the ordinary ones, and $l = ml_0$ is the *coarse-graining length*. Then $\widetilde{u}^{(m)}$ can be written in terms of the coarse-grained translational degrees of freedom and the axes of the coarse-grained midframe coordinate system, similar to Eq. (4). Under the reflection transformation $\mathcal{R}$ at the base pair level, the coarse-grained degrees of freedom transform similar to ordinary ones (see Appendix B for more details). As a result, Eq. (7) can be written in a generalized form as

$$\langle\widetilde{u}^{(m)}(1) \cdot \widetilde{u}^{(m)}(N+1)\rangle = \alpha(l)\exp\left(-\frac{L}{l_p^n}\right) \equiv \exp\left(-\frac{L}{l_p^c(l)}\right),$$
(11)

where

$$\alpha(l) = \exp\left(\frac{l}{l_p^n}\right) \langle \widetilde{\hat{u}}^{(m)}(1) \cdot \hat{d}_3(1) \rangle \, \langle \widetilde{\hat{u}}^{(m)}(1) \cdot \hat{d}_3(m+1) \rangle .$$

(12)

Approximate analytical expressions, to the first order of $\bar{\bar{\Theta}}$ and $\bar{\rho}$, can be easily obtained for the coarse-grained degrees of freedom. Using these expressions one can calculate $\alpha(l)$ to the second order of $\bar{\bar{\Theta}}$ and $\bar{\rho}$ to obtain (Appendix C)

$$\alpha(l) = 1 - \frac{2}{l\,\widetilde{K}_{sh}(l)} + \frac{l}{2\,l_p^n},$$

(13)

where $\widetilde{K}_{sh}(l)$ is the coarse-grained effective shear constant given by

$$\frac{1}{\widetilde{K}_{sh}(l)} = \frac{1}{2l}(\langle \widetilde{\rho}_1^{(l)} \rangle^2 + \langle \widetilde{\rho}_2^{(l)} \rangle^2) = \frac{1}{K_{sh}} + \frac{1}{12}\frac{1}{l_p^n}\left(l^2 - l_0^2\right).$$

(14)

On the right-hand side of Eq. (14), the first term comes from pure shearing at the base pair level and the second term reflects the contribution of the bending of the base pairs to the coarse-grained shear.

The argument presented above is also valid in 2D, where DNA is confined to a plane. In this case, the centers of the base pairs are forced to remain on a plane. To simplify the theoretical approach, we also assume that the normal vectors to the base pairs are confined to the plane. Since the confinement of DNA in 2D does not alter the locality and homogeneity of the elastic energy, or its symmetry properties, Eq. (3) is valid in 2D for all coarse-graining lengths. Thus $\alpha(l)$ is still given by Eq. (13), where l_p^n, K_{sh}, and $\widetilde{K}_{sh}(l)$ are defined by Eqs. (6), (10), and (14) respectively. The confinement in 2D only affects the values l_p^n and K_{sh}. If it is assumed that bending and shearing in DNA are isotropic (i.e. $A_1 = A_2$ and $K_1 = K_2$), we have $l_p^{n\,(2D)} = 2\,l_p^{n\,(3D)}$ and $K_{sh}^{(2D)} = 2\,K_{sh}^{(3D)}$, and consequently according to Eqs. (13) and (14), $\widetilde{K}_{sh}^{(2D)}(l) = 2\,K_{sh}^{(3D)}(l)$ and $\alpha^{2D}(l) - 1 = \frac{1}{2}(\alpha^{3D} - 1)$. For $(\alpha - 1) \ll 1$ it then follows that $l_p^{c\,(2D)}(l) = 2\,l_p^{c\,(3D)}(l)$. Dashed curves in Fig. 1(b) show the theoretical results for a DNA in 2D, assuming $m = 1$. It can be seen that the theory agrees with the simulation, and the isotropic approximation is valid.

Figure 3 shows α as a function of the coarse-graining length l, as calculated from 3D Monte Carlo simulation (squares, red [gray]) and 2D molecular dynamics simulations (circles, blue [gray]). The theoretical results presented above are also shown (solid red [gray] curve and dashed blue [gray] curve respectively), which nicely fit the simulation data. It can be seen in the figure that there is a *critical* value for the coarse-graining length, l_c, for which $\alpha = 1$ and thus $l_p^c = l_p^n$. The critical coarse-graining length can be calculated from Eq. (13), and for the elastic parameters used here, it turns out to be about $5\,bp$. One has $l_p^c < l_p^n$ for $l < l_c$, while $l_p^c > l_p^n$ for $l > l_c$.

Figure 4 shows l_p^c as a function of DNA length for four different values of coarse-graining length. Empty and filled symbols correspond to the 2D and 3D simulation data respectively, and curves show the theoretical results. As can be seen in the figure, although all curves converge to l_p^n for
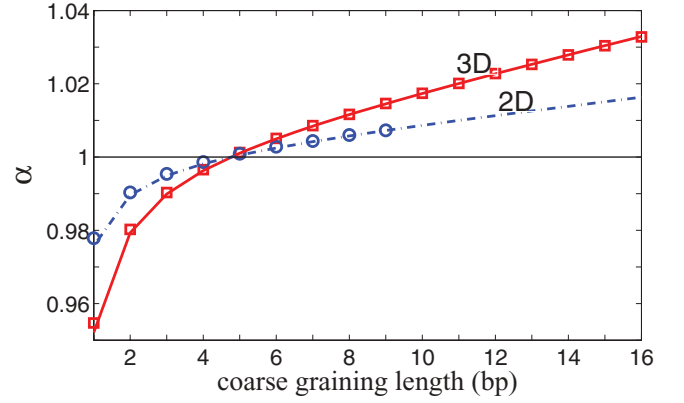

FIG. 3. (Color online) Plot of α^{3D} (squares, red [gray]) and α^{2D} (circles, blue [gray]) for different values of coarse-graining length. Solid curve (red [gray]) and dashed curve (blue [gray]) correspond to 3D and 2D theoretical results respectively.

long DNA molecules, the values of l_p^c can be significantly different from l_p^n at small length scales, depending on the coarse-graining length.

The generalized definition of the persistence length l_p^c is closely related to the method used in AFM experiments to measure the persistence length [6], where the coarse-graining length l is related to the AFM resolution. In this sense, our results suggest that the method used in the AFM experiments can dramatically affect the measured values of the persistence length and the bending flexibility of DNA at small length scales, and thus the correction coefficient α should be taken into account in the analysis to get a correct value for the persistence length. Data shown in Fig. 4 suggest that it is possible to determine α in a high-resolution AFM experiment, by measuring l_p^c as a function of DNA length at small length scales. It should be noted that α carries, by construction, information about average shear modulus of DNA. Having α in any resolution length l, one can find the effective shear modulus, $\widetilde{k}_{sh}(l)$ via Eq. (12). The contribution of the average shear modulus of DNA, K_{sh}, dominates the contribution of
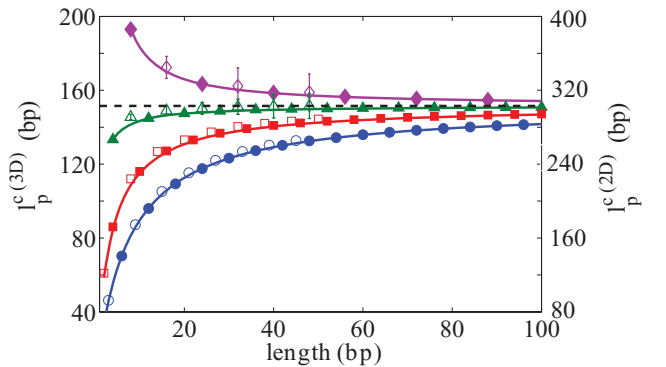
FIG. 4. (Color online) Plots of l_p^c vs DNA length for different values of the coarse-graining length: $l = 1bp$ (circles, blue [dark gray]), $l = 2bp$ (squares, red [light gray]), $l = 4bp$ (triangles, green [drak gray]), and $l = 8bp$ (diamonds, purple [light gray]). Empty and filled symbols correspond to 2D and 3D simulations respectively, and solid curves show the theoretical results. The dashed line indicates the value of l_p^n.

bending flexibility in Eq. (14) for $l < (l_0^2 + 12\, l_p^n\, K_{\text{sh}}^{-1})^{\frac{1}{2}}$, which is about 2.5 nm for the elastic parameters used here. This is a feasible method to measure DNA shear modulus by high-resolution AFM experiments.

IV. CONCLUSION

In summary, we have discussed differences between two possible definitions of DNA persistence length in a coarse-grained model. One definition is related to the orientation of the base pairs, which is only observable on NMR or crystallographic pictures, and the other definition is based on connecting the centers of successive base pairs, which is applicable to AFM experiments. It has been demonstrated that although the two definitions converge to a common value for sufficiently long DNA molecules, they are significantly different at short length scales.

APPENDIX A: SYMMETRY PROPERTIES OF THE UNIT VECTORS UNDER THE REFLECTION TRANSFORMATION

There are different ways to define the relative degrees of freedom of the base pairs [9,14,15]. We chose the so-called Cambridge University Engineering Helix Calculation Scheme (CEHS) representation [9] since it is easier to deal with mathematically, but it should be noted that the basic results presented in the paper will not change if one chooses another valid definition for the degrees of freedom. In the CEHS representation the orientation of the $(k + 1)$-th base pair with respect to the kth base pair is given by the relative rotation matrix $\mathfrak{r}(k)$ as $\hat{d}_i(k + 1) = \sum_{j=1}^{3} \mathfrak{r}_{ji}(k)\,\hat{d}_j(k)$, and the relative rotation matrix is written as an Euler rotation in the form

$$\mathfrak{r}(k) = \mathfrak{R}_3\left(\frac{\Omega(k)}{2} - \phi(k)\right)\mathfrak{R}_2[\Gamma(k)]\,\mathfrak{R}_3\left(\frac{\Omega(k)}{2} + \phi(k)\right),$$
(A1)

where $\mathfrak{R}_i(\phi)$ denotes the rotation about the axis $\hat{x}_i$ by angle ϕ [9]. The three rotational degrees of freedom for the $(k + 1)$-th base pair are then given by $\Theta_1(k) = \Gamma(k)\cos[\phi(k)]$, $\Theta_2(k) = \Gamma(k)\sin[\phi(k)]$, and $\Theta_3(k) = \Omega(k)$. The midframe coordinate system for the kth base pair step is defined in CEHS representation via relation $\hat{d}_i^M(k) = \sum_{j=1}^{3} \mathfrak{r}_{ji}^M(k)\,\hat{d}_j(k)$, where

$$\mathfrak{r}^M(k) = \mathfrak{R}_3\left(\frac{\Omega(k)}{2} - \phi(k)\right)\mathfrak{R}_2\left(\frac{\Gamma(k)}{2}\right)\mathfrak{R}_3[\phi(k)]. \quad (A2)$$

The translational degrees of freedom for the $(k + 1)$-th base pair are then given by $\rho_i(k) = \vec{r}(k) \cdot \hat{d}_i^M(k)$, $i = 1, 2, 3$, where $\vec{r}(k)$ is the relative position vector for the $(k + 1)$-th base pair with respect to the kth base pair.

The reflection transformation $\mathcal{R}$, as defined in the text, transforms the angle $\phi(k)$ to $\phi(k) + \pi$ for every k, while keeping $\Gamma(k)$ and $\Omega(k)$ unchanged. Therefore, the rotation matrices

$\mathfrak{r}(k)$ and $\mathfrak{r}^M(k)$ transform under the reflection transformation as

$$\mathfrak{r}(k) \longrightarrow \mathfrak{R}_3(\pi)\,\mathfrak{r}(k)\,\mathfrak{R}_3(\pi) \tag{A3}$$

and

$$\mathfrak{r}^M(k) \longrightarrow \mathfrak{R}_3(\pi)\mathfrak{r}^M(k)\,\mathfrak{R}_3(\pi)\,. \tag{A4}$$

Then, it follows that the unit vectors $\hat{d}_i(k)$ and $\hat{d}_i^M(k)$ transform under the reflection transformation as

$$\hat{d}_i(k+1) \cdot \hat{d}_j(k) \longrightarrow \begin{cases} -d_i(k+1) \cdot \hat{d}_j(k) & \text{if } i \text{ or } j \neq 3 \\ d_i(k+1) \cdot \hat{d}_j(k) & \text{otherwise} \end{cases}$$
(A5)

and

$$\hat{d}_i^M(k) \cdot \hat{d}_j(k) \longrightarrow \begin{cases} -d_i^M(k) \cdot \hat{d}_j(k) & \text{if } i \text{ or } j \neq 3 \\ d_i^M(k) \cdot \hat{d}_j(k) & \text{otherwise.} \end{cases}$$
(A6)

Since the reflection transformation changes the signs of $\rho_1(k)$ and $\rho_2(k)$ while keeping $\rho_3(k)$ unchanged, it follows from Eq. (A6) that the unit vectors $\hat{u}(k)$ [defined in Eq. (4) of the main text] transforms like

$$\hat{u}(k) \cdot \hat{d}_j(k) \longrightarrow \begin{cases} -\hat{u}(k) \cdot \hat{d}_j(k) & \text{if } j \neq 3, \\ \hat{u}(k) \cdot \hat{d}_j(k) & \text{if } j = 3. \end{cases} \tag{A7}$$

It is easy to show that $\hat{u}(k) \cdot \hat{d}_j(k + 1)$ transforms in a similar way.

APPENDIX B: COARSE-GRAINED DEGREES OF FREEDOM AND THEIR SYMMETRY PROPERTIES

For coarse graining DNA to a higher level, we replace every m base pairs with one coarse-grained base pair. The corresponding coarse-grained degrees of freedom are denoted by $\widetilde{\Theta}_i^{(l)}$ and $\widetilde{\rho}_i^{(l)}$, with $l = m\, l_0$ being the coarse-graining length. Defining the coarse-grained degrees of freedom in terms of the base pair degrees of freedom is straightforward. Using the formalism presented in the previous section we have $\hat{d}_i(k + m) = \sum_{j=1}^{3} \widetilde{\mathfrak{r}}_{ji}^{(l)}(k)\,\hat{d}_j(k)$, where

$$\widetilde{\mathfrak{r}}^{(l)}(k) = \prod_{p=k}^{k+m-1} \mathfrak{r}(p)\,. \tag{B1}$$

Similar to Eq. (A3), the coarse-grained relative rotation matrix $\widetilde{\mathfrak{r}}^{(l)}(k)$ can be written as an Euler rotation in the form

$$\widetilde{\mathfrak{r}}^{(l)}(k) = \mathfrak{R}_3\left(\frac{\widetilde{\Omega}^{(l)}(k)}{2} - \widetilde{\phi}^{(l)}(k)\right)\mathfrak{R}_2[\widetilde{\Gamma}^{(l)}(k)]\,\mathfrak{R}_3$$
$$\times \left(\frac{\widetilde{\Omega}^{(l)}(k)}{2} + \widetilde{\phi}^{(l)}(k)\right). \tag{B2}$$

The coarse-grained rotational degrees of freedom can then be defined as $\widetilde{\Theta}_1^{(l)}(k) = \widetilde{\Gamma}^{(l)}(k)\cos[\widetilde{\phi}^{(l)}(k)]$, $\widetilde{\Theta}_2^{(l)}(k) = \widetilde{\Gamma}^{(l)}(k)\sin[\widetilde{\phi}^{(l)}(k)]$, and $\widetilde{\Theta}_3^{(l)}(k) = \widetilde{\Omega}^{(l)}(k)$. Also, the coarse-grained midframe coordinate system, whose axes are denoted by $\widetilde{\hat{d}}_i^{M\,(l)}$, $i = 1, 2, 3$, can be defined via equation

$$\tilde{\bar{d}}_i^{M\,(l)}(k) = \sum_{j=1}^3 \tilde{\mathfrak{r}}_{ji}^{M\,(l)}(k)\,\hat{d}_j(k), \text{ where}$$

$$\tilde{\mathfrak{r}}^{M\,(l)}(k) = \mathfrak{R}_3\!\left(\frac{\widetilde{\Omega}^{(l)}(k)}{2} - \widetilde{\phi}^{(l)}(k)\right)\mathfrak{R}_2\!\left(\frac{\widetilde{\Gamma}^{(l)}(k)}{2}\right)\mathfrak{R}_3[\widetilde{\phi}^{(l)}(k)]. \tag{B3}$$

The coarse-grained translational degrees of freedom can then be defined by the projections of the vector which connects the base pair k to $k+m$ on the axes of the coarse-grained midframe coordinate system and are given by

$$\widetilde{\rho}_i^{(l)}(k) = \sum_{j=1}^3 \sum_{p=k}^{k+m-1} \rho_j(p)\,\hat{d}_j^M(p)\cdot\tilde{\bar{d}}_i^{M\,(l)}(k), \quad i = 1,2,3. \tag{B4}$$

From the transformation properties of $\mathfrak{r}(k)$ and Eq. (B1), it can be seen that under the reflection transformation at the base pair level, $\tilde{\mathfrak{r}}^{(l)}(k)$ transforms similar to $\mathfrak{r}(k)$, that is, Eq. (A3). It is then follows from Eq. (B2) that the reflection transforms $\widetilde{\phi}^{(l)}(k)$ to $\widetilde{\phi}^{(l)}(k)+\pi$ for every k, while keeping $\widetilde{\Gamma}^{(l)}(k)$ and $\widetilde{\Omega}^{(l)}(k)$ unchanged. As a result, the coarse-grained rotational degrees of freedom transform similar to the ordinary rotational degrees of freedom, that is,

$$\begin{aligned}
\widetilde{\Theta}_1^{(l)}(k) &\longrightarrow -\widetilde{\Theta}_1^{(l)}(k), \\
\widetilde{\Theta}_2^{(l)}(k) &\longrightarrow -\widetilde{\Theta}_2^{(l)}(k), \\
\widetilde{\Theta}_3^{(l)}(k) &\longrightarrow \widetilde{\Theta}_3^{(l)}(k).
\end{aligned} \tag{B5}$$

Also, it can be seen from Eq. (B3) that $\tilde{\mathfrak{r}}^{M\,(l)}(k)$ transforms similar to $\mathfrak{r}^M(k)$, that is, Eq. (A4). Hence it is easy to show that we have

$$\tilde{\bar{d}}_i^{M\,(l)}(k)\cdot\hat{d}_j^M(p) \longrightarrow \begin{cases} -\tilde{\bar{d}}_i^{M\,(l)}(k)\cdot\hat{d}_j^M(p) & \text{if } i \text{ or } j \neq 3 \\ \tilde{\bar{d}}_i^{M\,(l)}(k)\cdot\hat{d}_j^M(p) & \text{otherwise,} \end{cases} \tag{B6}$$

for every p and k. Therefore, it follows from Eqs. (B4) and (B6) that the coarse-grained translational degrees of freedom transform similar to the ordinary translational degrees of freedom, that is,

$$\begin{aligned}
\widetilde{\rho}_1^{(l)}(k) &\longrightarrow -\widetilde{\rho}_1^{(l)}(k), \\
\widetilde{\rho}_2^{(l)}(k) &\longrightarrow -\widetilde{\rho}_2^{(l)}(k), \\
\widetilde{\rho}_3^{(l)}(k) &\longrightarrow \widetilde{\rho}_3^{(l)}(k).
\end{aligned} \tag{B7}$$

APPENDIX C: CORRECTION COEFFICIENT IN THE FIRST APPROXIMATION

The correction coefficient $\alpha(l)$, as given in Eq. (12) of the main text, can be written in terms of the coarse-grained degrees of freedom. Considering the definition of $\widetilde{\rho}_i^{(l)}(k)$ in Eq. (B4), the unit vector $\widetilde{u}^{(l)}(k)$ can be written as

$$\widetilde{u}^{(l)}(k) = \frac{1}{|\widetilde{\bar{\rho}}^{(l)}(k)|}\sum_{i=1}^3 \widetilde{\rho}_i^{(l)}(k)\,\tilde{\bar{d}}_i^{M\,(l)}(k), \tag{C1}$$

where $|\widetilde{\bar{\rho}}^{(l)}| \equiv \sqrt{(\widetilde{\rho}_1^{(l)\,2} + \widetilde{\rho}_2^{(l)\,2} + \widetilde{\rho}_3^{(l)\,2})}$. From Eqs. (B2) and (B3) we obtain

$$\tilde{\bar{d}}_i^{M\,(l)}(k)\cdot\hat{d}_j(k) = \left\{\mathfrak{R}_3\!\left(\frac{\widetilde{\Omega}^{(l)}(k)}{2} - \widetilde{\phi}^{(l)}(k)\right)\mathfrak{R}_2\!\left(\frac{\widetilde{\Gamma}^{(l)}(k)}{2}\right)\mathfrak{R}_3[\widetilde{\phi}^{(l)}(k)]\right\}_{ji} \tag{C2}$$

and

$$\tilde{\bar{d}}_i^{M\,(l)}(k)\cdot\hat{d}_j(k+m) = \left\{\mathfrak{R}_3\!\left(-\frac{\widetilde{\Omega}^{(l)}(k)}{2} - \widetilde{\phi}^{(l)}(k)\right)\mathfrak{R}_2\!\left(-\frac{\widetilde{\Gamma}^{(l)}(k)}{2}\right)\mathfrak{R}_3[\widetilde{\phi}^{(l)}(k)]\right\}_{ji}. \tag{C3}$$

Also, from Eq. (B2) together with the Eq. (5) of the main text, it follows that

$$\langle\cos\widetilde{\Gamma}^{(l)}(k)\rangle = \exp\!\left(-\frac{l}{l_p^n}\right). \tag{C4}$$

Using Eqs. (B4) and (C1)–(C4), together with Eq. (12) of the main text, we can write

$$\alpha(l) = \langle\cos\widetilde{\Gamma}^{(l)}\rangle^{-1}\left\langle\frac{\widetilde{\rho}_3^{(l)}}{|\widetilde{\bar{\rho}}^{(l)}|}\cos\frac{\widetilde{\Gamma}^{(l)}}{2} - \frac{\widetilde{\rho}_1^{(l)}\cos\widetilde{\phi}^{(l)} - \widetilde{\rho}_2^{(l)}\sin\widetilde{\phi}^{(l)}}{|\widetilde{\bar{\rho}}^{(l)}|}\sin\frac{\widetilde{\Gamma}^{(l)}}{2}\right\rangle$$

$$\times\left\langle\frac{\widetilde{\rho}_3^{(l)}}{|\widetilde{\bar{\rho}}^{(l)}|}\cos\frac{\widetilde{\Gamma}^{(l)}}{2} + \frac{\widetilde{\rho}_1^{(l)}\cos\widetilde{\phi}^{(l)} - \widetilde{\rho}_2^{(l)}\sin\widetilde{\phi}^{(l)}}{|\widetilde{\bar{\rho}}^{(l)}|}\sin\frac{\widetilde{\Gamma}^{(l)}}{2}\right\rangle. \tag{C5}$$

In Eq. (C5) we have omitted the indexes that indicate the number of the base pair steps, since the elastic energy is translationally invariant and every base pair step is equivalent to any other. To calculate the statistical averages in Eq. (C5), we expand the

right-hand side of the equation in terms of $\bar{\bar{\Theta}}$ and $\bar{\rho}$. Equations (B1)–(B4) relate the coarse-grained degrees of freedom to the ordinary ones. Hence, it is straightforward to calculate $\widetilde{\Theta}^{(l)}$ and $\widetilde{\bar{\rho}}^{(l)}$ to the first order of $\bar{\bar{\Theta}}$ and $\bar{\rho}$. We obtain

$$\widetilde{\Theta}_1^{(l)}(k) = \sum_{p=1}^{m} \left[\cos\left(\psi_p^m\right) \bar{\Theta}_1(k+p-1) + \sin\left(\psi_p^m\right) \bar{\Theta}_2(k+p-1) \right] + O(\bar{\Theta}^2, \bar{\rho}^2), \tag{C6}$$

$$\widetilde{\Theta}_2^{(l)}(k) = \sum_{p=1}^{m} \left[-\sin\left(\psi_p^m\right) \bar{\Theta}_1(k+p-1) + \cos\left(\psi_p^m\right) \bar{\Theta}_2(k+p-1) \right] + O(\bar{\Theta}^2, \bar{\rho}^2), \tag{C7}$$

$$\widetilde{\Theta}_3^{(l)}(k) = l\omega_0 + \sum_{p=1}^{m} \bar{\Theta}_3(k+p-1) + O(\bar{\Theta}^2, \bar{\rho}^2), \tag{C8}$$

$$\widetilde{\rho}_1^{(l)}(k) = l_0 \sum_{p=1}^{m} \left[\cos\left(\psi_p^m\right) \bar{\rho}_1(k+p-1) + \sin\left(\psi_p^m\right) \bar{\rho}_2(k+p-1) \right] + l_0 \sum_{p=1}^{m} \left[-\left(\frac{m}{2} + \frac{1}{2} - p\right) \sin\left(\psi_p^m\right) \bar{\Theta}_1(k+p-1) \right.$$
$$\left. + \left(\frac{m}{2} + \frac{1}{2} - p\right) \cos\left(\psi_p^m\right) \bar{\Theta}_2(k+p-1) \right] + O(\bar{\Theta}^2, \bar{\rho}^2), \tag{C9}$$

$$\widetilde{\rho}_2^{(l)}(k) = l_0 \sum_{p=1}^{m} \left[-\sin\left(\psi_p^m\right) \bar{\rho}_1(k+p-1) + \cos\left(\psi_p^m\right) \bar{\rho}_2(k+p-1) \right] - l_0 \sum_{p=1}^{m} \left[\left(\frac{m}{2} + \frac{1}{2} - p\right) \cos\left(\psi_p^m\right) \bar{\Theta}_1(k+p-1) \right.$$
$$\left. + \left(\frac{m}{2} + \frac{1}{2} - p\right) \sin\left(\psi_p^m\right) \bar{\Theta}_2(k+p-1) \right] + O(\bar{\Theta}^2, \bar{\rho}^2), \tag{C10}$$

and

$$\widetilde{\rho}_3^{(l)}(k) = l + l_0 \sum_{p=1}^{m} \bar{\rho}_3(k+p-1) + O(\bar{\Theta}^2, \bar{\rho}^2), \tag{C11}$$

where $\psi_p^m = \left(\frac{m+1}{2} - p\right)\omega_0 l_0$. By expanding $\alpha(l)$ in Eq. (C5) in powers of $\widetilde{\Theta}^{(l)}$ and $\widetilde{\bar{\rho}}^{(l)}$ to the second order, and using Eqs. (C6)–(C11), one can easily calculate the averages in Eq. (C5) to obtain Eqs. (13) and (14) of the main text. These two equations are valid to the second order of $\bar{\bar{\Theta}}$ and $\bar{\bar{\rho}}$, and in the special case of $m = 1$ are reduced to Eqs. (9) and (10) of the main text respectively.

[1] R. Vafabakhsh and T. Ha, Science **337**, 1097 (2012).

[2] P. C. Nelson, Science **337**, 1045 (2012).

[3] J. F. Marko and E. D. Siggia, Macromolecules **27**(4), 981 (1994).

[4] P. Nelson, *Biological Physics: Energy, Information, Life* (W.H. Freeman, New York, 2008).

[5] J. F. Marko and E. D. Siggia, Macromolecules **28**(26), 8759 (1995).

[6] P. A. Wiggins, T. V. der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, Nat. Nanotechnol. **1**, 137 (2006).

[7] A. Podesta, L. Imperadori, W. Colnaghi, L. Finzi, P. Milani, and D. Dunlap, Biophys. J. **89**, 2558 (2005).

[8] A. Podesta, M. Indrieri, D. Brogioli, G. S. Manning, P. Milani, R. Guerra, L. Finzi, and D. Dunlap, J. Microsc. **215**, 236 (2004).

[9] M. A. El-Hassan and C. R. Calladine, J. Mol. Biol. **251**, 648 (1995).

[10] N. Becker, L. Wolff, and R. Everaers, Nucleic Acid Res. **34**, 5638 (2006).

[11] K. B. Towels, J. F. Beausang, H. G. Garcia, R. Philips, and P. C. Nelson, Phys. Biol. **6**, 025001 (2009).

[12] B. Mergell, M. R. Ejtehadi, and R. Everaers, Phys. Rev. E **68**, 021911 (2003).

[13] H. Kamberaj, R. J. Low, and M. P. Neal, J. Chem. Phys. **122**, 1906216 (2005).

[14] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska, Nucleic Acids Res. **37**, 5917 (2009).

[15] F. Lankaš, O. Gonzalez, L. M. Heffler, G. Stoll, M. Moakherd, and J. H. Maddocks, Phys. Chem. Chem. Phys. **11**, 10565 (2009).