Range of Interaction between DNA-Bending Proteins is Controlled by the Second-Longest Correlation Length for Bending Fluctuations

Houyin Zhang and John F. Marko

The range of interaction between DNA-bending proteins is controlled by the second-longest correlation length for bending fluctuations

Houyin Zhang[1, *] and John F. Marko[2, †]

[1]*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA*
[2]*Department of Physics and Astronomy and Department of Molecular Biosciences,
Northwestern University, Evanston, Illinois 60208, USA*

When a DNA molecule is stretched, the zero-force correlation length for its bending fluctuations – the persistence length A – bifurcates into two different correlation lengths – the shorter "longitudinal" correlation length $\xi_\parallel(f)$ and the longer "transverse" correlation length $\xi_\perp(f)$. In the high-force limit, $\xi_\parallel(f) = \xi_\perp(f)/2 = \sqrt{k_B T A/f}/2$. When DNA-bending proteins bind to the DNA molecule, there is an effective interaction between the protein-generated bends mediated by DNA elasticity and bending fluctuations. Surprisingly, the range of this interaction is not the longest correlation length associated with transverse fluctuations of the tangent vector along the polymer, but instead is the second longest longitudinal correlation length $\xi_\parallel(f, \mu)$. The effect arises from the protein-bend contribution to the Hamiltonian having an axial rotational symmetry which eliminates its coupling to the transverse fluctuations.

In living cells, binding of proteins to DNA controls gene expression and packaging of the genome [1–3], requiring DNA-binding proteins to communicate and cooperate over a wide range of length scales. Many DNA-binding proteins distort the double helix, which has led to development of micromechanical methods for analyis of protein-DNA interactions [4–7]. Notably, tension along DNA molecules can induce cooperative interactions between distant DNA-bending proteins [8–10]. Numerical analysis of a semiflexible "worm-like chain" (WLC) showed the force-generated interaction range between two quenched bends (corresponding to bound proteins) to be $\xi(f) \approx \sqrt{A/\beta f}/2$ [10], in the high-force range $f > k_B T/A \approx 0.1$ pN. The dominant (longest) correlation length for the WLC is that for bending fluctuations, $\xi_\perp(f) \approx \sqrt{A/\beta f}$ [11]. Surprisingly, the protein-induced bends have an interaction range exactly half of $\xi_\perp(f)$. Here we show that this arises from the rotational symmetry of the part of the energy describing the protein-DNA interaction. We will illustrate this for a specific model, but we will also argue that our results are general for a wide class of similar models of proteins interacting with DNA.

To investigate the force-dependent interactions between DNA-binding proteins we use a discretized WLC model including protein-binding effects [10] (essentially the model of Refs. [12] and [13]). The DNA is treated as a chain of N links each of length b, with each link able to point in any direction with a Boltzmann weight that depends on the angle between adjacent links. This form of the model allows it to be treated by "transfer matrix" [14] methods, whereby summation over individual link degrees of freedom can be reduced to matrix multiplications. This method has a long history of use in biopolymer physics, from early studies of the helix-coil transition [15], to recent studies of semiflexible polymer elasticity [16, 17].

In our model, proteins are taken to bind "nonspecifically" (with no sequence-dependent affinity) to the nodes between adjacent links. The DNA-protein complex energy is a sum over the links,

$$\beta E = \sum_{i=1}^{N} \left\{ \frac{a}{2}|\hat{t}_{i+1} - \hat{t}_i|^2(1 - n_i) + \left[\frac{a'}{2}(\hat{t}_{i+1} \cdot \hat{t}_i - \gamma)^2 - \mu \right] n_i \right.$$
$$\left. -\beta b f \hat{t}_i \cdot \hat{z} - \eta n_i n_{i+1} \right\}, \qquad (1)$$

where $\hat{t}_i$ is an unit vector which describes the direction of each link, the dimensionless parameter a ($= A/b$) is the naked DNA bending rigidity, a' is the rigidity of a protein-occupied node, and n_i is the protein occupation degree of freedom for the i-th node ($n_i = 0$ for an unoccupied node, or 1 for a protein-bound node). Protein binding is controlled by the chemical potential μ which is a function of bulk protein concentration c ($\mu = \ln[c/c_0]$ where c_0 is a protein-dependent constant determining the binding affinity), and the bends induced by protein binding are described by $\gamma = \cos\theta$, where θ is the preferred angle between adjacent links resulting from protein binding. The DNA-protein complex is stretched by the applied force in the $\hat{z}$ direction. Finally, η is the intrinsic binding cooperativity, describing the strength of direct interaction between adjacent bound proteins.

This model describes a variety of DNA-binding proteins [10, 12, 18], depending on the various parameter values chosen. For most situations of interest, the total DNA contour length L ($= Nb$) can be taken to be much bigger than the link length b, and a periodic boundary condition can be used ($\hat{t}_N = \hat{t}_0$). We compute the equilibrium partition function $Z = \text{Tr}(\mathbf{T}^N)$ following the method of Ref. [10], where $\mathbf{T}$ is the transfer matrix for a

link

$$\mathbf{T} = \begin{pmatrix} (\langle \hat{t}_i, 1|\mathbf{T}|\hat{t}_{i+1}, 1\rangle) & (\langle \hat{t}_i, 1|\mathbf{T}|\hat{t}_{i+1}, 0\rangle) \\ (\langle \hat{t}_i, 0|\mathbf{T}|\hat{t}_{i+1}, 1\rangle) & (\langle \hat{t}_i, 0|\mathbf{T}|\hat{t}_{i+1}, 0\rangle) \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{A}e^{\mu+\eta} & \mathbf{A}e^{\mu} \\ \mathbf{B} & \mathbf{B} \end{pmatrix} . \qquad (2)$$

Here $\mathbf{A}$ and $\mathbf{B}$ are matrices with elements $\langle \hat{t}|\mathbf{A}|\hat{t}'\rangle = e^{-\frac{a'}{2}(\hat{t}\cdot\hat{t}'-\gamma)^2 + \beta bf\hat{t}\cdot\hat{z}}$ and $\langle \hat{t}|\mathbf{B}|\hat{t}'\rangle = e^{-\frac{a}{2}|\hat{t}'-\hat{t}|^2 + \beta bf\hat{t}\cdot\hat{z}}$.

The matrices may be expressed in terms of spherical harmonics, $i.e.$,

$$\langle lm|\,\mathbf{A}\,|l'm'\rangle \equiv \int d\hat{t} \int d\hat{t}' Y_{lm}^*(\hat{t})\, \langle \hat{t}|\,\mathbf{A}\,|\hat{t}'\rangle\, Y_{l'm'}(\hat{t}') , \quad (3)$$

where the integrals are on the unit sphere. Due to axial symmetry around the stretching (z) direction, the matrix elements are diagonal in m, $i.e.$, Eq. 3 is proportional to $\delta_{mm'}$. Closed-form expressions for the matrix elements of $\mathbf{A}$ and $\mathbf{B}$ can be found in Sec. S1 of Supplemental Material [19].

Protein occupation at any node along the chain can be computed in this transfer matrix framework using the "operators" $\langle \hat{t}, n|\hat{\mathbf{n}}|\hat{t}', n'\rangle = \delta_{\hat{t}\hat{t}'}\delta_{nn'}n$, or

$$\hat{\mathbf{n}} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} , \qquad (4)$$

where $\mathbf{1}$ and $\mathbf{0}$ stand for the unit matrix and zero matrix in the $\{|l,m\rangle\}$ space, respectively. The average and two-point correlations for the occupation variables are

$$\langle n_i \rangle = \mathrm{Tr}(\hat{\mathbf{n}}\mathbf{T}^N)/\mathrm{Tr}(\mathbf{T}^N) ,$$
$$\langle n_i n_{i+j} \rangle = \mathrm{Tr}(\hat{\mathbf{n}}\mathbf{T}^j\hat{\mathbf{n}}\mathbf{T}^{N-j})/\mathrm{Tr}(\mathbf{T}^N) ,$$
$$G^{\hat{\mathbf{n}}}(r, f) \equiv \langle n_i n_{i+j} \rangle - \langle n_i \rangle\langle n_{i+j} \rangle , \qquad (5)$$

where $G^{\hat{\mathbf{n}}}(r, f)$ is the protein occupation correlation function for two DNA-binding proteins separated by contour length $r = jb$.

The three components of the tangent vector can also be expressed as "operators": $\langle \hat{t}, n|\hat{\mathbf{t}}^i|\hat{t}', n'\rangle = \delta_{\hat{t}\hat{t}'}\delta_{nn'}t^i$, $i = x, y,$ or z. It will be useful to express $\hat{\mathbf{t}}^x$ and $\hat{\mathbf{t}}^y$ in terms of raising and lowering operators $\hat{\mathbf{t}}^{\pm} = \hat{\mathbf{t}}^x \pm i\hat{\mathbf{t}}^y$. Matrix elements of these operators in the $\{|l,m\rangle\}$ basis are straightforward to compute (Sec. S2 of Supplemental Material [19]).

These matrices have some important symmetries; $\hat{\mathbf{t}}^{\pm}$ and $\hat{\mathbf{t}}^z$ are diagonal in n (for the moment we suppress the simple n-dependence of these operators). Second, $\hat{\mathbf{t}}^z$ is diagonal in m due to rotational symmetry around the z direction; in the $\{|l,m\rangle\}$ space with an appropriate order of basis vectors, $\hat{\mathbf{t}}^z$ is block-diagonal (nonzero entries shown as purple or dark gray entries of Fig. 1 for $l, l' \leq 3$). On the other hand, in this basis the raising (lowering) operator $\hat{\mathbf{t}}^+$ ($\hat{\mathbf{t}}^-$) is a block lower (upper) next-to-diagonal matrix [nonzero entries indicated as the orange or light gray lower (upper) next-to-diagonal elements in Fig. 1].
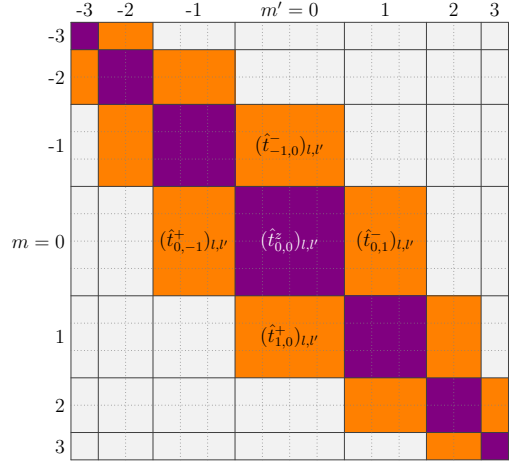

FIG. 1. (color online) Structure of the three tangent component operators $\hat{\mathbf{t}}^z$, $\hat{\mathbf{t}}^+$, and $\hat{\mathbf{t}}^-$ in $\{|l,m\rangle\}$ space for the case $l, l' \leq 3$ (ignoring the n-dependence). For matrix $\hat{\mathbf{t}}^z$, only matrix elements along the main diagonal blocks are possibly non-zero (purple or dark gray); all other matrix elements are zero. For matrix $\hat{\mathbf{t}}^+$ ($\hat{\mathbf{t}}^-$), only matrix elements in the lower (upper) next-to-diagonal blocks are possibly non-zero (orange or light gray); all other matrix elements are zero.

The raising and lowering operators also have some key symmetries in the $\{|l,m\rangle\}$ basis:

$$\begin{matrix} \hat{\mathbf{t}}^+_{l,m,n;l',m-1,n} & = & -\hat{\mathbf{t}}^+_{l',1-m,n;l,-m,n} \\ \| & & \| \\ \hat{\mathbf{t}}^-_{l',m-1,n;l,m,n} & = & -\hat{\mathbf{t}}^-_{l,-m,n;l',1-m,n} \end{matrix} . \qquad (6)$$

The average z-component of the tangent vector is

$$\langle t_i^z \rangle = \mathrm{Tr}(\hat{\mathbf{t}}^z\mathbf{T}^N)/\mathrm{Tr}(\mathbf{T}^N) , \qquad (7)$$

which also equals the molecule end-to-end extension as a fraction of total contour length. The transverse components of the tangent vector average to zero by symmetry: $\langle t_i^x \rangle = \langle t_i^y \rangle = 0$.

The longitudinal correlations are

$$G^{\hat{\mathbf{t}}^z}(r, f) \equiv \langle t_i^z t_{i+j}^z \rangle - \langle t_i^z \rangle\langle t_{i+j}^z \rangle , \qquad (8)$$
$$\langle t_i^z t_{i+j}^z \rangle = \mathrm{Tr}(\hat{\mathbf{t}}^z\mathbf{T}^j\hat{\mathbf{t}}^z\mathbf{T}^{N-j})/\mathrm{Tr}(\mathbf{T}^N) ,$$

and the transverse correlations are

$$G^{\hat{\mathbf{t}}^x}(r, f) \equiv \langle t_i^x t_{i+j}^x \rangle = \langle t_i^+ t_{i+j}^- \rangle/2 = G^{\hat{\mathbf{t}}^y}(r, f) , \quad (9)$$
$$\langle t_i^+ t_{i+j}^- \rangle = \mathrm{Tr}(\hat{\mathbf{t}}^+\mathbf{T}^j\hat{\mathbf{t}}^-\mathbf{T}^{N-j})/\mathrm{Tr}(\mathbf{T}^N) .$$

Before studying force-generated protein-protein interactions, we examine the bending correlation lengths for naked DNA. In the absence of protein ($\mu = -\infty$, driving $n_i = 0$), Eq. 1 becomes a 1-d Heisenberg model. We use a segment size $b = 5$ nm (approximately 15 bp) and DNA bending stiffness $a = 10$ corresponding to persistence length $A = ab = 50$ nm. We consider the long-DNA limit $L \gg A$ for all computations, and a cutoff on

l is chosen to obtain ten-digit numerical precision (typically $l_{\max} = 14$). The averages of the z-component of the tangent vector $\langle \hat{\mathbf{t}}^z \rangle$ and the tangent vector correlation functions $G^{\hat{\mathbf{t}}^z}(r, f)$ and $G^{\hat{\mathbf{t}}^x}(r, f)$ are computed from Eqs. 7, 8 and 9 (Sec. S3 of Supplemental Material [19]). For large j, we extract correlation lengths from the expected asymptotic form of the correlation functions:

$$G^{\hat{\mathbf{t}}^z}(r, f) = C_\parallel(f) e^{-r/\xi_\parallel(f)} , \qquad (10)$$

where $\xi_\parallel(f)$ and $C_\parallel(f)$ are the longitudinal correlation length and amplitude. Similarly, we extracted the transverse correlation length $\xi_\perp(f)$ and amplitude $C_\perp(f)$.

Fig. 2(a) shows $\xi_\parallel(f)$ (bottom pink crosses) and $\xi_\perp(f)$ (top orange pluses) as a function of force. For small forces, the two correlation lengths approach the same limit $\xi_\parallel(f) \approx \xi_\perp(f) \approx 50$ nm, the persistence length for a dsDNA. However, as force is increased, the two correlation lengths vary differently. As one enters the high-force range ($f > k_B T/A \approx 0.1$ pN) the correlation lengths bifurcate into two curves, both $\sim f^{-1/2}$ but with different coefficients: $\xi_\perp = 2\xi_\parallel = \sqrt{k_B TA/f}$. $\xi_\parallel(f)$ approaches (numerically) exactly half of $\xi_\perp(f)$ for large forces.

Fig. 2(b) shows the longitudinal correlation function amplitude $C_\parallel(f)$ (bottom pink crosses), the zero-distance correlation function $G^{\hat{\mathbf{t}}^z}(0, f)$ (bottom green squares), the transverse correlation function amplitude $C_\perp(f)$ (top orange pluses), and the zero-distance correlation function $G^{\hat{\mathbf{t}}^x}(0, f)$ (top aqua circles). The longitudinal amplitude is (numerically) the square of the transverse correlation function amplitude for large forces.

To analytically understand the relationship between the transverse and longitudinal correlation functions for naked DNA, we consider the high-force limit ($f \gg k_B T/A$) of the continuous WLC model where a Gaussian approximation for tangent vector fluctuations becomes appropriate [11]. In this limit, since $\hat{t}$ fluctuates only slightly around $\hat{z}$, then t_x and t_y are small quantities, and $t_z = 1 - (t_x^2 + t_y^2)/2 + \mathcal{O}(t_x^4 + t_y^4)$. Following Ref. [11] (Sec. S4 of Supplemental Material [19]),

$$G^{\hat{\mathbf{t}}^x}(r, f) = \sqrt{k_B T/4Af}\, e^{-r/\sqrt{k_B TA/f}} + \cdots ,$$
$$G^{\hat{\mathbf{t}}^z}(r, f) = (k_B T/4Af)\, e^{-r/\sqrt{k_B TA/4f}} + \cdots . \quad (11)$$

Therefore, the longitudinal correlation function is essentially the square of the transverse correlation function in the large force limit, indicating that $\xi_\parallel(f) = \sqrt{k_B TA/4f} = \xi_\perp(f)/2$ and $C_\parallel(f) = k_B TA/4Af = C_\perp^2(f)$. Fig. 2 includes analytic results for $\xi_\parallel(f)$ (Fig. 2(a), bottom red line), $\xi_\perp(f)$ (Fig. 2(a), top blue line), $C_\parallel(f)$ (Fig. 2(b), bottom red line) and $C_\perp(f)$ (Fig. 2(b), top blue line). In the high-force range, the analytic results match our numerical calculation.

The asymptotic behaviors of the transverse and longitudinal correlation functions can also be obtained from the discrete WLC transfer matrices, using $G^{\hat{\mathbf{t}}^z}(r, f)$ and $G^{\hat{\mathbf{t}}^x}(r, f)$ defined in Eq. 8 and Eq. 9, respectively. In the large-N limit (Sec. S5 of Supplemental Material [19]) ,

$$G^{\hat{\mathbf{t}}^z}(r, f) = \sum_{k=1} (\mathbf{t}_0^z)_{0,k} (\mathbf{t}_0^z)_{k,0} \left(\frac{\lambda_k^0}{\lambda_0^0} \right)^j ,$$
$$G^{\hat{\mathbf{t}}^x}(r, f) = \frac{1}{2} \sum_{k=1} \left(\mathbf{t}_{0,-1}^+\right)_{0,k} \left(\mathbf{t}_{-1,0}^-\right)_{k,0} \left(\frac{\lambda_k^1}{\lambda_0^0} \right)^j , \quad (12)$$

where the λ_k^m ($m = 0, \pm 1, \pm 2, \cdots; k = |m|, |m|+1, |m|+2, \cdots$) are the eigenvalues of the m-th diagonal block of the "naked DNA" matrix $\mathbf{B}$ and $\lambda_{|m|}^m > \lambda_{|m|+1}^m > \lambda_{|m|+2}^m > \cdots$. Sec. S6 of the Supplemental Material [19] shows how applied force f affects the eigenvalues $\{\lambda_k^m\}$, i.e., the transfer matrix spectrum of naked DNA.

$(\mathbf{t}_m^z)_{k_1, k_2}$ is the matrix element of the m-th diagonal block of the matrix $\hat{\mathbf{t}}^z$ (ignoring n-dependence) expressed by taking the eigenvectors of the m-th diagonal block of the naked matrix $\mathbf{B}$ ($\{|\lambda_k^m\rangle\}$) as the basis, $(\mathbf{t}_m^z)_{k_1, k_2} = \langle \lambda_{k_1}^m | \hat{\mathbf{t}}_m^z | \lambda_{k_2}^m \rangle$. Similarly, $\left(\mathbf{t}_{m,m-1}^+\right)_{k_1, k_2} = \langle \lambda_{k_1}^m | \hat{\mathbf{t}}_{m,m-1}^+ | \lambda_{k_2}^{m-1} \rangle$ and $\left(\mathbf{t}_{m,m-1}^-\right)_{k_1, k_2} = \langle \lambda_{k_1}^{m-1} | \hat{\mathbf{t}}_{m-1,m}^- | \lambda_{k_2}^m \rangle$.

For large distances (large j) the correlation functions (12) is dominated by the $k = 1$ contributions, giving

$$\xi_\parallel(f) = \frac{b}{\ln(\lambda_0^0/\lambda_1^0)} , \quad \xi_\perp(f) = \frac{b}{\ln(\lambda_0^0/\lambda_1^1)} . \quad (13)$$

We have verified that the calculated results of $\xi_\parallel(f)$ and $\xi_\perp(f)$ by Eq. 13 computed by matrix diagonalization [bottom green squares and top aqua circles in Fig. 2(a)] match the result computed directly from the multiplication of matrices [compare to bottom pink crosses and top orange pluses in Fig. 2(a)].

We now consider the interaction between DNA-bending proteins. We computed the protein occupation correlation function $G^{\hat{\mathbf{n}}}(r, f)$ along a DNA in dilute protein solution, the decay of which provides the correlation length for the inter-protein interaction, for the case $\gamma = 0$ and $a' = 50$, corresponding to a rather stiff protein-DNA complex with a preferred bend angle of $90°$. We extract the protein occupancy correlation length, just as in Eq. 10; the result for $\mu = -5$ and $\eta = 0$ is shown in Fig. 3 (bottom grey filled circles; also see Sec. S7 of Supplemental Material [19]). For comparison, in Fig. 3 we also show the tangent vector correlation lengths of protein-bound DNA (bottom pink crosses and top orange pluses) and those of naked DNA (bottom green squares and top aqua circles).

Fig. 3 reveals that the protein occupancy correlation length is equal to the longitudinal tangent vector correlation length $\xi_\parallel(f, \mu)$. This can be understood from the diagonal-in-m form (Fig. 1) of the protein occupancy matrix $\hat{\mathbf{n}}$, which is shared by $\mathbf{A}$, $\hat{t}^z$, and $\mathbf{B}$ reflecting their axial rotational symmetry. For the protein occupancy correlation function, just as for $G^{\hat{\mathbf{t}}^z}(r, f)$, blocks of different m are decoupled, leading to the correlation function
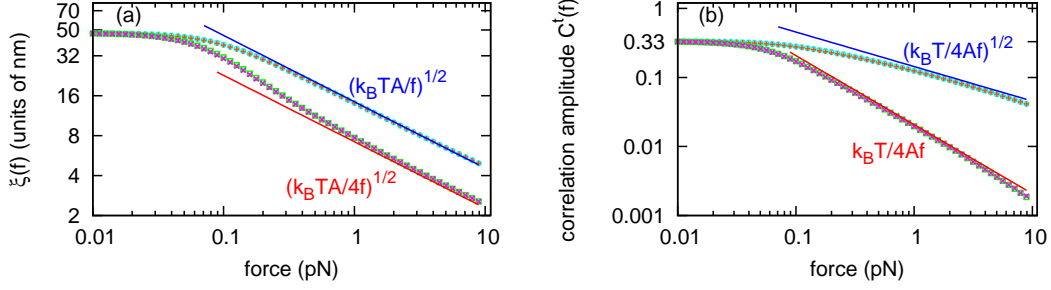
FIG. 2. (color online) Tangent vector correlation lengths and amplitudes for naked DNA. (a) The longitudinal correlation length $\xi_\parallel(f)$ calculated by three different methods: by Eq. 10 (bottom pink crosses), by Eq. 11 (bottom red line), and by Eq. 13 (bottom green squares, overlapping with bottom pink crosses). The transverse correlation length $\xi_\perp(f)$ calculated by three different methods: by Eq. 10 (top orange pluses), by Eq. 11 (top blue line), and by Eq. 13 (top aqua circles, overlapping with top orange pluses). In high-force range, $\xi_\perp(f) = 2\xi_\parallel(f) = \sqrt{k_B T A / f}$. (b) The longitudinal correlation function amplitude $C_\parallel(f)$ calculated by Eq. 10 (bottom pink crosses), agreeing with $G^{\hat{t}^z}(0, f)$ (bottom green squares) and the high-force limit from Eq. 11 (bottom red line). The transverse correlation function amplitude $C_\perp(f)$ calculated by Eq. 10 (top orange pluses), agreeing with $G^{\hat{t}^x}(0, f)$ (top aqua circles) and the high-force limit from Eq. 11 (top blue line).
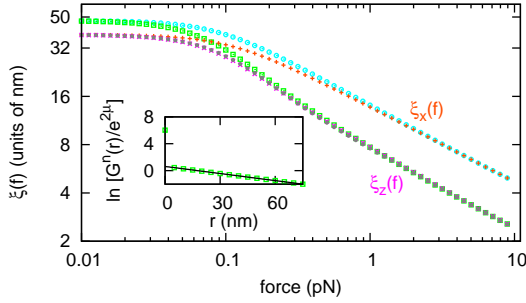

FIG. 3. (color online) Protein occupation correlation length along DNA (bottom grey filled circles) in dilute protein solution ($\mu = -5$, $\eta = 0$, $\gamma = 0$, $a' = 50$). For reference, the tangent vector correlation lengths along DNA in the same solution (bottom pink crosses and top orange pluses) and those along a naked DNA (bottom green squares and top aqua circles) are shown. Inset shows $\ln[G^{\hat{n}}(r)/e^{2\mu}]$ vs. distance at $f = 0.1$ pN (green squares) and a linear fit of it (black line).

decay being determined by the largest two eigenvalues of the ($m = 0$)th diagonal blocks of transfer matrix **T**, which has the largest eigenvalue of **T**. Thus, the correlation function of the protein occupation degrees of freedom have a decay length of $\xi_\parallel(f, \mu)$. For large forces, this is exactly half of the longest correlation length in the problem, that of $\xi_\perp(f, \mu)$; this longer transverse correlation length arises from the off-block-diagonal structure of $\hat{t}^\pm$ which allows selection of the two largest eigenvalues from any *two* adjacent diagonal blocks of **T**. The protein occupancy operator $\hat{n}$ has higher symmetry (Eq. 4), which makes $\langle n_i^2 \rangle = \langle n_i \rangle$ and explains the jump of protein occupation correlation function at $r = 0$.

Our main result, namely $\xi_\parallel$ being the range of protein-protein interactions mediated by bending fluctuations along a stretched DNA, follows from the rotational symmetry of the protein-DNA interaction operator **A**. This

result is general for proteins which distort DNA structure as long as the distortions are isotropically distributed around the double helix axis (the same result follows for other values of γ, a' and μ, Sec. S8 of Supplemental Material [19]). For non-sequence-specific binding this will be the case since two proteins may bind in essentially any relative axial orientation by shifting their binding positions over a 10.5 base-pair range (the $h \approx 3.6$ nm helix repeat for dsDNA), a length scale small relative to the inter-protein separations we are considering. Twist fluctuations of inter-protein DNA will further randomize the relative axial orientation of pairs of DNA-bound proteins. Of course, for very large forces where $\xi_\parallel$ becomes comparable to h, a more detailed model taking into account the helix structure would need to be employed, but in that regime, with $\xi_\parallel$ comparable in size to a typical DNA-binding protein, one would expect direct interactions between adjacent proteins to dominate (Sec. S7 of Supplemental Material [19]).

One might ask whether the longer $\xi_\perp$ can ever be the protein-protein interaction range. This would require the relative angular orientations of the two proteins around the axial (stretching) direction to be fixed: one might imagine using a sequence-specific DNA-binding protein and binding sites "phased" to be at the same orientation around the double helix (spaced multiples of 10.5 bp) in an attempt to do this. However, since angular (twist) fluctuations between the two sites will increase with interprotein distance, one would expect to see an interaction decaying with the longer $\xi_\perp$ only at short distances, with the shorter $\xi_\parallel$ dominating at longer distances.

Finally, we note that while the individual protein-generated distortions considered here are isotropically distributed around the double helix axis, the *relative* orientation of two nearby bends can be expected to be angularly correlated. This arises from the fact that two

oppositely-directed bends of the double helix compensate one another's distortion [8–10]. Therefore one can expect force on a DNA to direct the self-organization of DNA-bending-proteins into clusters (via the interaction we have analyzed) with correlated bend directions. If the bends are chiral (the general case since DNA and proteins are both chiral), nearby proteins will be chirally organized, *i.e.* forming locally helical DNA-protein complexes. An important case is that of nucleosomes, which have well-defined entry/exit angles for the DNA bound to them, with broken chiral symmetry; moderate applied tension can be expected to drive the packing and helical organization of adjacent nucleosomes in chromatin.

---

* Email address: houyinzhang2011@u.northwestern.edu
† Email address: john-marko@northwestern.edu

[1] J. Stavans and A. Oppenheim, Phys. Biol. **3**, R1 (2006).
[2] E.P.C. Rocha, Annu. Rev. Genet. **42**, 211-33 (2008).
[3] *Genome Organization and Function in the Cell Nucleus*, ed. K. Rippe, Wiley-VCH (New York, 2011).
[4] B.M. Ali, R. Amit, I. Braslavsky, A. B. Oppenheim, O. Gileadi, and J. Stavans, Proc. Natl. Acad. Sci. U.S.A. **98**, 10658 (2001).
[5] J. van Noort, S. Verbrugge, N. Goosen, C. Dekker, R.T. Dame, Proc. Nat. Acad. Sci. USA **101**, 6969 (2004).
[6] A. Meglio, E. Praly, F. Ding, J.F. Allemand, D. Bensimon, V. Croquette, Curr. Opin. Struct. Biol. **19**, 615 (2009).
[7] K.R. Chaurasiya, T. Paramanathan, M.J. McCauley, M.C. Williams, Phys. Life. Rev. 7, 299 (2010).
[8] J. Rudnick and R. Bruinsma, Biophys. J. **76**, 1725 (1999).
[9] E.F. Koslover and A.J. Spakowitz, Phys. Rev. Lett. **102**, 178102 (2009).
[10] H. Zhang and J.F. Marko, Phys. Rev. E **82**, 051906 (2010).
[11] J.F. Marko and E.D. Siggia, Macromolecules **28**, 8759 (1995).
[12] J. Yan and J.F. Marko, Phys. Rev. E **68**, 011905 (2003).
[13] J. Yan, R. Kawamura, and J.F. Marko, Phys. Rev. E **71**, 061905 (2005).
[14] H.A. Kramers, G.H. Wannier, Phys. Rev. **60**, 252 (1941).
[15] B.H. Zimm, J.K. Bragg, J. Chem. Phys. **31** 526 (1959).
[16] L. Livadaru, R.R. Netz, H.J. Kreuzer, Macromolecules **36**, 3732 (2003).
[17] C. Storm, P.C. Nelson, Phys. Rev. E **67** 051906 (2003).
[18] B. Xiao, H. Zhang, R.C. Johnson, and J.F. Marko, Nucleic Acids Res. **39**, 5568 (2011).
[19] Supplemental Material is available at (URL).