DNA stretching and multivalent-cation-induced condensation

Yevgeni Sh. Mamasakhlisov,[1] Brian A. Todd,[2] Artem. V. Badasyan,[1] Anna. V. Mkrtchyan,[1] Vladimir. F. Morozov,[1] and V. Adrian Parsegian[3]
[1]*Department of Molecular Physics, Yerevan State University, 1 Al. Manougian Str., Yerevan 375025, Armenia*
[2]*Department of Physics, Purdue University, West Lafayette, Indiana 47901, USA*
[3]*National Institutes of Health, Bethesda, Maryland 20892, USA*

(Received 9 February 2009; revised manuscript received 28 April 2009; published 24 September 2009)

Motivated by measurements on stretched double-stranded DNA in the presence of multivalent cations, we develop a statistical mechanical model for the compaction of an insoluble semiflexible polymer under tension. Using a mean-field approach, we determine the order of the extended-to-compact transition and provide an interpretation for the magnitude and interval of tensions over which compaction takes place. In the simplest thermodynamic limit of an infinitely long homogeneous polymer, compaction is a first-order transition that occurs at a single value of tension. For finite length chains or for heterogeneous polymers, the transition progresses over an interval of tension. Our theory provides an interpretation for the result of single-molecule experiments in terms of microscopic parameters such as persistence length and free energy of condensation.

DOI: 10.1103/PhysRevE.80.031915 PACS number(s): 87.10.Ca, 87.15.hp, 87.14.gk, 82.37.Rs

I. INTRODUCTION

Despite the biological importance of counterion-induced DNA condensation in the packaging of genomes, the physical mechanisms of this process remain controversial; proposed mechanisms range from complementary ordering of counterions along the helical backbone of DNA [1–3], to correlations between strongly coupled counterions [4–6], to counterion-induced cohesive hydration forces [7–9]. Within this context, single-molecule stretching experiments [10,11] have emerged as a powerful tool to probe the magnitude of condensation energies and to discriminate between alternate theories [12–17].

In these measurements, a well-defined tension is applied to the ends of a single double-helical DNA molecule bathed in a solution containing multivalent counterions (Fig. 1). Under large tensions (>10 pN), the DNA remains stretched and shows force-extension characteristics consistent with a wormlike chain, albeit at persistence lengths smaller than are typically observed in the presence of monovalent ions. At some tension, typically between 0.1–5 pN, the DNA undergoes an abrupt and progressive collapse to “condense” into a compact structure with greatly reduced end-to-end length. The molecule remains collapsed under lower applied forces.

In this paper, we develop a statistical mechanical framework for interpreting the magnitude and interval of the extended-to-condensed transition of single stretched DNA molecules condensed by polyvalent cations. We achieve a clear interpretation of the experimental results in terms of the underlying molecular parameters such as condensation free energy, persistence length, and total DNA length.

In the simplest thermodynamic limit of an infinitely long chain and homogeneous condensation free energies, compaction is a first-order transition that occurs at a single force that is nearly equal to the free-energy per unit length of condensation. This is equivalent to the interpretation given previously for the results of single-molecule stretching experiments [18,19]. Recent experiments with DNA condensation (see, e.g., [1]) and measurements of interactions between short DNA segments [20,21] suggest that interaction free energies may be weakly sequence-specific. For this reason, we also consider chains with heterogeneous interaction energies. This effect and the finite length of polymer chain result in a condensation transition over a finite range of forces. We compare our theoretical results with single-molecule measurements.

II. FREE ENERGY FOR POLYMER MODEL

We assign each segment $\tilde{l}$ of the polymer (i.e., each persistence length) either to a globular (condensed) or to a coiled (extended) region; interactions between regions are neglected. The state of this system may then be described by a set of $\{g_i, c_i\}$ variables, where g_i is the length (in $\tilde{l}$ units) of the ith globular region, and c_i is the length (in the same units) of the ith coil region (Fig. 2). Both coiled and globular regions are considered as a wormlike chain (WLC).

The free energy for each globular region can be calculated in a manner identical to that in Ref. [22],

$$F_{g_i} \simeq G g_i + G_s \frac{R_i r_i}{d^2} + k_B T \frac{\tilde{l}^2 g_i}{R_i^2}, \tag{1}$$

where G is the free energy of intersegment interaction per persistence length. Because of the multivalent ion-mediated attraction in condensed DNA, $G<0$. In the absence of stretching forces, this self-attraction causes DNA to condense into an insoluble phase. The second term takes into account the surface tension or energy of exposing the surface of the toroidal globule to the bath; R_i is the outer and r_i the inner radius of torus; G_s, the surface tension. The third term is the free energy of elastic bending of a stiff chain, where L is the contour length of macromolecule and d is the chain diameter. Sizes of each globular region are considered to satisfy the equation

©2009 The American Physical Society

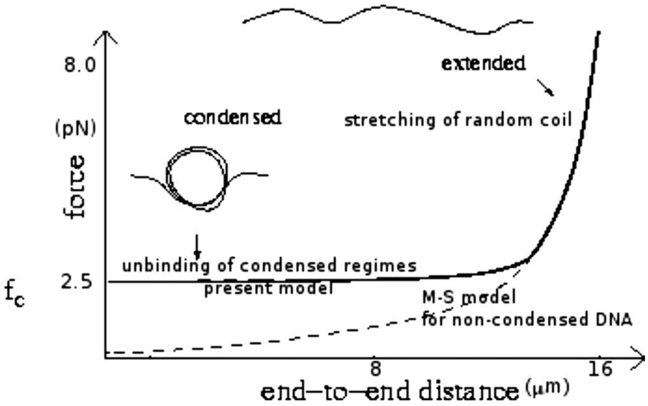

FIG. 1. Double-stranded DNA stretched in the presence of multivalent cations: at large stretching forces the DNA remains in an extended configuration that follows the wormlike chain model (dash line). At decreased force, the DNA condenses and the end-to-end distances rapidly decreases, departing from wormlike chain prediction. The critical force at which the extended-to-condensed transition takes place f_c can be measured from experiments.

$$\tilde{l} g_i d^2 \simeq R_i r_i^2. \tag{2}$$

This simply means that the volume $\tilde{l} g_i d^2$ of condensed polymer in each globule is equal to the volume $R_i r_i^2$ of the globule treated as a torus.

Now we derive an expression for the free energy of the coil part of the chain. We need to estimate the free energy of the WLC at a given value of extending external force f. As f is the parameter under control, the chain end-to-end distance x becomes physically observable quantity, conjugated to the force f,

$$x = \frac{\partial \ln Z_L}{\partial(\beta f)}, \tag{3}$$

where Z_L is the partition function of WLC of length L. $\beta = 1/(k_BT)$ is the inverse temperature. To estimate the free energy of a WLC at a given f, we define the Legendre transform as

$$\Gamma\left(\frac{x}{L}\right) = -\frac{\beta}{L}\mathcal{F}_{coil}[L, f_L(x)] - \beta f \frac{x}{L}, \tag{4}$$

where $\mathcal{F}_{coil}[L, f_L(x)] = -k_BT \ln Z_L[f_L(x)]$. $f_L(x)$ is defined implicitly in Eq. (3) and corresponds to the force-extension expression derived by Marko and Siggia [23] for a WLC. In terms of the Legendre transform, the free energy of the coiled region at the given value of f can be written

$$\mathcal{F}_{coil}(L,f) = \int_0^x dx' f_L(x') - fx = -\frac{TL}{4\tilde{l}}\left(\frac{x}{L}\right)^2\left[\frac{1}{(1-x/L)^2}+2\right], \tag{5}$$

where x is defined by the Marko-Siggia expression $f = f_L(x)$ for given f, $\tilde{l}$ and L. This is identical to the expression used by Blumberg *et al.* [24]. The first term is the potential energy stored in the stretched polymer chain. The second term is the reversible work done against the constant force f. Thus, the free energy of the extended coil of the length L_i containing c_i monomers of length $\tilde{l}$ is equal to

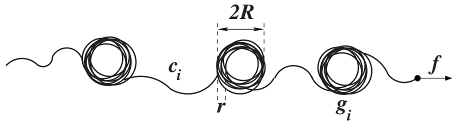

FIG. 2. Schematic representation of the partly ordered dsDNA under stretching: c_i is the number of monomers in the ith coiled subchain and g_i the number of monomers in the ith condensed subchain. R is the external and r the internal radius of the globule.

$$F_{c_i} = \mathcal{F}_{coil}(L_i, f) = -c_i \tilde{f}, \tag{6}$$

where $\tilde{f} = \frac{k_BT}{4}\left(\frac{x}{L}\right)^2\left[\frac{1}{(1-x/L)^2}+2\right]$.

Now that we have estimates for the free energies of coiled and globular regions, we can write the total free energy for the polymer of alternating condensed and extended regions, $\{g_i, c_i\}$, as the sum of corresponding free-energy terms F_{g_i} and F_{c_i} from Eqs. (1) and (6)

$$\frac{F\{g_i,c_i;R_i,r_i\}}{k_BT} = \sum_{i=1}^{n} g_i\left[\frac{G}{k_BT}+\frac{\tilde{l}^2}{R_i^2}\right] + \sum_{i=1}^{n}\left[\frac{G_s}{k_BT}\frac{R_i r_i}{d^2} - c_i\tilde{f}\right]. \tag{7}$$

In the last equation the free energy of the given state $\{g_i, c_i; R_i, r_i\}$ is calculated as additive contributions of regions. We consider condensation under an external stretching force that substantially hinders the long loops formation. Thus, the direct interaction between globular and coiled regions can be neglected. The partition function of the system with free energy given by Eq. (7) can then be written as

$$Z_N = \sum_n \sum_{\{g_i,c_i\}} \prod_{i=1}^{n} \int dR_i dr_i \delta(\tilde{l} g_i d^2 - R_i r_i^2) \times \exp[-\beta F\{g_i, c_i; R_i, r_i\}]. \tag{8}$$

Here the δ function takes into account the normalization condition (2). Thus, we have two groups of degrees of freedom, $\{g_i, c_i\}$ and $\{R_i, r_i\}$ which are realized together at the thermodynamic equilibrium. In the mean-field approximation all the regions, enumerated by $i = 1, \ldots, n$ are equivalent to each other. In the case of long chains, we can assume uniform-sized globules, $R_i = R$ and $r_i = r$. This apparently restricts the number monomers in each globule, but on the mean-field level, where the $\langle g_i \rangle = \nu_g$ $(i = 1, \ldots, n)$ this approximation is adequate.

III. THEORY

Thus, the partition function of the system with free energy given by Eq. (7) can be estimated in the mean-field approximation as

$$Z_N = \sum_n \sum_{\{g_i,c_i\}} \exp[-\beta F_{R,r}\{g_i, c_i\}], \tag{9}$$

where $F_{R,r}\{g_i, c_i\} = F\{g_i, c_i; R_i = R, r_i = r\}$.

Equation (7) parameterizes the free energy of a stretched, self-attracting WLC in terms of the population of globular and coiled regions $\{g_i, c_i\}$.

Suppose that any microstate of the system under consideration can be represented as a sequence of ordered (in our case, globular) (g) and coiled (c) phases. We can reformulate our problem on the language of the Zimm-Bragg parameters s and σ, widely applied to describe temperature driven helix-coil transitions [25,26]. s governs the ordered phase growth, while σ reflects the cooperativity of transition. The Zimm and Bragg parameters in our case will depend on the mean dimensions or toroid R and r and, subsequently, on the mean length of the globular domain. The partition function (9) of the system with free energy given by Eq. (7) becomes

$$Z_N = \sum_n \sum_{\{g_i, c_i\}} \tilde{s}^{N_g} \tilde{\sigma}^n, \tag{10}$$

where

$$\tilde{s} = \exp\left[-\beta G - \tilde{f} - \frac{\tilde{l}^2}{R^2} \right],$$

$$\tilde{\sigma} = \exp\left[-\beta G_s \frac{Rr}{d^2} \right];$$

$$N_g = \sum\nolimits_{i=1}^{n} g_i. \tag{11}$$

The main measurable quantity, in analogy with the traditional helix-coil description "degree of helicity," is the "degree of condensation,"

$$\theta = \frac{\langle N_g \rangle}{N}. \tag{12}$$

Another quantity is the mean length of ordered (condensed in our case) region ν_g, estimated in the Zimm-Bragg theory for homogeneous chain as [26]

$$\nu_g = \langle g_i \rangle \simeq \frac{1}{\sqrt{\tilde{\sigma}}}. \tag{13}$$

Thus, in the mean-field approximation we need to minimize the free energy of the system by parameters R and r using the "normalization" condition (2), where g_i can be substituted by its mean value $\langle g_i \rangle = \nu_g$.

The Zimm-Bragg language has been successfully applied to many types of conformational transitions. For instance, it was recently used for the description of stretch-induced DNA melting [27] as well as for the torque-induced B-P transition in dsDNA [28]. Now, having established relation between DNA condensation under tension and the Zimm-Bragg parametrization, we can use known methods to investigate the state of the system under various stretching forces.

A. Transition point and interval: homogeneous infinite chain

We first consider the simplest thermodynamic limit of an infinitely long homogeneous chain. Applying the generating function analysis in thermodynamic limit, a discontinuity in degree of compaction is found (for details see Appendix A), indicating that it is a first-order phase transition; expressions for the transition point and interval are as follows:

$$\tilde{f}(f_c)|_{N\to\infty} = -\beta G, \tag{14}$$

$$\Delta \tilde{f}(f)|_{N\to\infty} = 0. \tag{15}$$

We conclude that in the $N \to \infty$ limit the transition point is uniquely defined by the effective energy G of attraction between DNA segments and by the persistence length $\tilde{l}$. In this limit, the system can form a globule of infinite radius, while maintaining an infinite volume-to-surface ratio. Consequently, neither the bending nor surface terms play any role. Therefore, in experiments with long DNA the measured critical force f_c reflects only the strength of DNA-DNA attraction and the potential energy of the stretched WLC. When the force drops below the condensation free-energy per unit length minus the potential-energy per unit length of the WLC, the chain collapses completely.

First-order phase transitions are not expected for systems with one spatial dimension [29]. While our system is parameterized along a single axis (distance along the chain) the interface between two phases c and g is not pointlike, but is at the surface of globula. The statistical weight of this interphase junction, defined by Eq. (11) tends to the limit $\tilde{\sigma} \to 0$ in the thermodynamic limit at the critical force f_c. Hence, while the system is parameterized along a single dimension, a phase transition of the first order is still possible. First order is still possible with helix-helix attraction.

The phase behavior of the model with free energy Eq. (7) can be qualitatively compared with those for the widely investigated Poland-Scheraga model [30–33]. The statistical weight of the one elementary coil-helix pair is equal to $w(g,c) = s^g \exp(c\tilde{f}) \exp(-|\ln \sigma| \sqrt{g})$ in Eq. (A3), while the analogous statistical weight for the Poland-Scheraga model is $w(g,c) = s^g \sigma \exp(-\gamma \ln c)$. The order of the phase transition is governed mainly by the nonadditive term assigned to the $(c-g)$ phase boundary. Recall that, for $\gamma > 2$ the Poland-Scheraga model exhibits the phase transition of first order, while for $1 < \gamma < 2$ the transition becomes second order. The power-law dependence of free-energy $\sim \sqrt{g}$ is sharper than logarithmic, thus ensuring the condensation transition to be of first order.

B. Transition point and interval: homogeneous finite chain

To study finite chain length effects, we apply the mean-field approximation, in a way, similar to what had been done for the free energy of the Zimm-Bragg model [34]

$$-\beta \mathcal{F}_N \simeq \max_{n, N_g, R, r} \left[N_g \ln \tilde{s} + n \ln \tilde{\sigma} + \ln \Omega_N(n, N_g) \right], \tag{16}$$

where $\Omega_N(n, N_g)$ is the number of ways to arrange coiled and globular regions, $\Sigma_i g_i = N_g$ and $\Sigma_i c_i = N - N_g$. We maximize Eq. (16) using undefined Lagrange multiplier method with the normalization condition (2) (for details see Appendix B). Different from the usual cooperativity parameter of helix-coil transition, the parameter $\tilde{\sigma}$ is dependent on the mean dimensions of torus (R and r) and changing with the growth of the globular phase. However the interval of transition is defined mainly by the cooperativity parameter at the midpoint of transition and, therefore, can be approximately considered as constant. Compared with the infinite chain case [Eqs. (14) and (15)], the transition shifts to a lower force,

$$\tilde{f}_c \simeq -\beta G - \alpha^{4/5} N^{-2/5}, \tag{17}$$

and acquires a finite interval [26]

$$\Delta\tilde{f}(f) \simeq \sqrt{\tilde{\sigma}} \sim \frac{1}{N}, \tag{18}$$

where $\alpha = \beta G_s \frac{\tilde{l}}{d}$. The lowered force reflects a decrease in the free energy of the condensed state due to bending and surface tension of the finite sized condensate. The finite interval for the transition means that the transition for the finite length homogeneous chain is not strictly first order. However, Eqs. (13) and (18) indicate that the mean length of globular regions is comparable with the chain length and that there should only be a small number of globular regions in the stretched homogeneous chain. This will tend to minimize the contribution of bending and surface tension.

C. Transition point and interval: heterogeneous chain

We consider condensation under an external stretching force that substantially hinders the formation of long loops. That is why the impact of interactions between different globular regions to the free energy is negligible. Thus, condensation of the given ith globule involved only segments neighboring along the chain. According to the proposed model, condensation of homogeneous DNA exhibits an abrupt, single-step condensation curve, θ vs f. However, recent experiments have demonstrated heterogeneity in the interactions between different DNA sequences. To consider the potential impact of this on single-molecule DNA condensation experiments we also consider a polymer chain with heterogeneous interaction strength. We assume that the free energy of domain condensation behaves as a local sequence property in order to consider it in a framework similar to that of the Zimm-Bragg theory.

In the Zimm-Bragg formalism, heterogeneity is incorporated via parameters $\tilde{s}$ [see Eq. (11)] that are different for different types of monomers along the chain (for details see Appendix C). We assume that the free energy of the extended coil is sequence independent. Taking into account Eq. (11), the midpoint of transition is found as

$$\tilde{f}(f_c) = -\beta\langle G\rangle_{seq} - \frac{\tilde{l}^2}{R^2}, \tag{19}$$

where $\langle G\rangle_{seq}$ is the average free energy of intersegment attraction per persistence length and R is the mean value of the large radius of toroid. This indicates that the DNA stretching experiments measure a sequence-averaged free energy of condensation $\langle G\rangle_{seq}$. Following the approach described in [26,34] the interval of transition for heterogeneous chain can be estimated as (see Appendix C)

$$\Delta\tilde{f}(f) \simeq \Omega^2(\beta\delta G)^2 \frac{1}{|\ln\tilde{\sigma}|}, \tag{20}$$

where Ω^2 is variance of homological sequences distribution in the chain and δG is variation of the intersegment interaction free energies. Taking into account Eq. (18), cooperativity parameter scales with chain length as $\tilde{\sigma} \sim N^{-2}$, and the interval of transition is [see Eq. (20)]

$$\Delta\tilde{f}(f) \sim \frac{1}{\ln N}. \tag{21}$$

Compared with Eq. (18) for the homogeneous chain, the interval of the transition for the heterogeneous chain converges much more slowly to the thermodynamic limit than for homogeneous chains ($1/\ln N$ vs $1/N$). Heterogeneous chains also have a shorter average length per globule (or, equivalently, more globules per chain) and the mean length of globular region at the midpoint of transition is [34]

$$\nu_g \simeq \frac{|\ln\tilde{\sigma}|^2}{\Omega^2(\beta\delta G)^2} \sim (\ln N)^2. \tag{22}$$

A similar logarithmic dependence for the mean length of bubble has been obtained recently for twist-induced DNA denaturation in random sequences and for the binding transition [35,36]. The mean value of the large radius of toroidal domain R scales with length of domain as $R \sim \nu_g^{1/5}$ [see Eq. (B8)]. Taking into account Eqs. (19) and (22), the midpoint of transition is finally found as

$$\tilde{f}(f_c) = -\beta\langle G\rangle_{seq} - \alpha^{4/5}\nu_g^{-2/5} \simeq -\beta\langle G\rangle_{seq} - \alpha^{4/5}(\ln N)^{-4/5}. \tag{23}$$

The main difference between condensation under the stretching force and WLC condensation in 3D is the quasi-one-dimensional nature of the transition. The applied stretching force prevents long loop formation and restricted condensation by the segments neighboring along the chain [see Eq. (7) and explanation below]. That is why the growth of the ith globular domain takes place only at the cost of neighboring i and $(i+1)$th coiled domains. Thus, when the relative globule boundary free energy decreases with the globule size, it compensates by the free increase at the cost of neighboring coiled domains with low propensity to condensation. It is important to mention that the boundary free energy is proportional to the surface of the globule while the free energy increase at the cost of condensation of neighbor coiled domains is proportional to the volume of globule. That is why, contrary to the homogeneous case, the heterogeneous globule has no trend to a maximum possible size globule formation.

IV. DISCUSSION

We have managed to describe the extended-to-condensed transition of an insoluble semiflexible polymer as a function of tension, using the language and methods, previously used with the Zimm and Bragg model of helix-to-coil transition. The model we offer allows us to present molecular parameters, such as finite chain length and sequence heterogeneity that may be measured in experiments involving stretched DNA condensation by multivalent cations. The model does not, of course, prove any particular form of DNA/DNA attraction nor is the Bragg-Zimm model to be viewed as more than an enabling procedure with consequent limited meaning and applicability.

In the thermodynamic limit of an infinite homogeneous chain, the condensation transition occurs as a first-order phase transition at a single force value, i.e., critical force. If the condensation free-energy per unit length is much larger than the work per unit length for extending the wormlike chain, $k_BT/\tilde{l}$, then the critical force is nearly equal to the condensation free-energy per unit length. For DNA, this would involve forces much larger than ~ 0.1 pN. For smaller forces, the configurational free energy of the stretched chain becomes significant and the transition occurs at a reduced force given by Eq. (6). This interpretation is equivalent to those used previously to interpret DNA stretching experiments [5,11].

For finite length DNA chains, the condensate must have a finite size, leading to energetic contributions from bending and surface tension. These effects produce a condensation transition that occurs over a finite interval at a reduced force. The precise width that our model predicts for this interval is difficult to estimate because some molecular parameters, particularly the surface tension associated with the condensed state, are not precisely known. However, we can make rough estimates for the critical force and transition interval using the surface tension parameter $\alpha \sim 3$ given in Ref. [37], the contour length of lambda-phage DNA $L=16.7$ μm from the single-molecule stretching experiments of Ref. [11] and the persistence length $\tilde{l}=27$ nm given by least-square fitting the same experimental data to the Marko-Siggia formula. We predict via Eq. (18) (and expression $N=\frac{L}{\tilde{l}}\approx 600$) that the width of transition interval due to finite length effects should be $\Delta f \sim 0.001$ pN. This is sufficiently small that, even given the uncertainty in the microscopic parameters, we expect that finite length effects are not significant in single-molecule stretching experiments of micron-sized DNA. This predicts a much larger effect on the force-distance dependence of stretched DNA stems from heterogeneity in the interaction free energies. According to our model, the interval of transition Δf can be estimated using Eqs. (19) and (20). This predicts that the $\Delta f=0.1-0.2$ pN transition intervals would be observed in experiments with DNA chains with heterogeneous free energies of interaction.

V. CONCLUSIONS

We have modeled the collapse of dsDNA in the presence of extending force, accounting for the finite size of the system and the sequence heterogeneity. As shown, it is possible to formulate the problem in terms of parameters of Zimm and Bragg model, thus opening room for the application of long-known methods and approaches used for the description of helix-to-coil transition. We assume that the conformation of the whole chain can be represented as a sequence of the globular and coiled regions and, that the direct interactions between these ordered and disordered regions can be neglected, due to the presence of extending force. The approach provides a theoretical language to describe experimental force-distance dependence and helps to interpret the measured data. If we assume an infinite chain, collapse occurs as a first-order phase transition at a force nearly equal to the free energy of condensation per unit length. This interpretation is equivalent to those appearing previously [5,11]. For a finite homogeneous chain the interval of transition is small, but finite and scales with chain length as $\sim O(\frac{1}{N})$. For molecules with $N>100$, such as those used previously in experiments, this interval is probably too small to measure. To describe heterogeneous interactions, similar to those previously seen in experiments [1,20], we considered a heterogeneous polymer, modeled as a random sequence of monomers with different values for free energy between globular and coiled phases. In this case, the interval of transition between the extended and collapsed phases estimates as $\sim 1/\ln N$ and for our values of system parameters equals $\Delta f \sim 0.1$ pN. This value is similar to the transition intervals seen in experiments that most closely approach equilibrium conditions [11,14].

ACKNOWLEDGMENTS

V.F.M., Y.Sh.M., and A.V.B. thankfully acknowledge the LPSB/NICHD/NIH for hospitality during short-term visits when part of this paper was written. A.V.B. acknowledges the support from a PRIN-COFIN 2007 grant. This work was supported by the Intramural Research Program of the NIH, Eunice Kennedy Shriver National Institute of Child Health and Human Development.

APPENDIX A: THERMODYNAMICAL LIMIT AND PHASE TRANSITION

To calculate the partition function of the N particle system, let us use the generating function method (see, e.g., in [26]). In our case the generating function can be written as

$$\Xi(p) = \sum_{N=1}^{\infty} p^N Z_N, \tag{A1}$$

where p is any complex number. Then the partition function of the N particle system is calculated as

$$Z_N = \frac{1}{2\pi\imath}\oint dp\,\Xi(p)p^N, \tag{A2}$$

where the contour of integration contains the origin of coordinates, but not the singularities of the function $\Xi(p)$. Then, in the thermodynamics limit $N\to\infty$, the value of Z_N is defined by the singularity of the function $\Xi(p)$ closest to the origin [26]. By taking into account the normalization condition (2), the statistical weight $\tilde{s}^{N_g}\tilde{\sigma}^n$ in the Eq. (10) can be presented as $\tilde{s}^{N_g}\tilde{\sigma}^n=\prod_{i=1}^{n} w(g_i,c_i)$, where

$$w(g,c) = s^g \exp(c\tilde{f})\sigma^{\sqrt{g}} \tag{A3}$$

and

$$s = \exp[\beta G - (\tilde{l}/R)^2]$$

$$\sigma = \exp[-\beta G_s (R\tilde{l}/d^2)^{1/2}]. \tag{A4}$$

To apply the generating function method, the grand canonical partition function is rewritten as

$$\Xi = \sum_{N=1}^{\infty} p^N \sum_{n=1}^{\infty} \sum_{\{g_i,c_i\}} \prod_{i=1}^{n} w(g_i,c_i)\, \delta\left(\sum_i (g_i + c_i) - N\right), \tag{A5}$$

where Dirac δ on the right hand side ensures that the overall system size is N.

As shown in Ref. [26], $\Xi = \frac{\zeta(p)}{1-\zeta(p)}$, where

$$\zeta(p) = \sum_{g=1}^{\infty} s^g p^g \sigma^{\sqrt{g}} \sum_{c=1}^{\infty} p^c \exp(c\tilde{f}), \tag{A6}$$

and in $N \to \infty$ limit the value of $\mathcal{F}_N$ is defined by the nearest to zero pole p_0 of the generating function (A5) as

$$\frac{\mathcal{F}_N}{N} \simeq T \log p_0. \tag{A7}$$

The pole p_0 is determined from $1 = \zeta(p_0)$ equation which can be written as

$$\exp(-\tilde{f}) - p_0 = p_0 \phi(sp_0, \sigma), \tag{A8}$$

where $\phi(sp_0,\sigma) = \Sigma_{g=1}^{\infty}(sp_0)^g \sigma^{\sqrt{g}}$. The solution of Eq. (A8) is found from the intersection between the $\exp(-\tilde{f}) - p_0$ and $p_0\phi(sp_0,\sigma)$. The series $\phi(z,\sigma)$ converges if $|z| \leq 1$ and the function $p\phi(sp,\sigma)$ has finite value if $p = 1/s$. Thus the system with Hamiltonian Eq. (7) exhibits the phase transition at some critical value of $\tilde{f}$,

$$\exp(-f_c) - 1/s = \frac{1}{s}\phi(1,\sigma). \tag{A9}$$

In the thermodynamic limit, and at forces lower than critical $\tilde{f}(f) \leq \tilde{f}(f_c)$, the free energy is

$$\frac{\mathcal{F}_N}{N} \simeq -T \log s. \tag{A10}$$

The degree of chain condensation to describe the force-driven transition can be introduced as

$$\theta = \frac{\langle N_g \rangle}{N}, \tag{A11}$$

where $N_g = \Sigma_{i=1}^{n} g_i$ is the total number of "monomers" in the globular phase and $\langle \cdots \rangle$ is the thermal average. Using Eq. (10), the degree of condensation can be written as

$$\theta = \frac{\partial}{\partial \log s}\left(-\frac{\mathcal{F}_N}{k_B T N}\right). \tag{A12}$$

From Eqs. (A10) and (A12) it follows, that for force values corresponding to $\tilde{f}(f) \leq \tilde{f}(f_c)$, the condensation degree $\theta = 1$. In other words, at these values of $\tilde{f}$ the whole chain is in the compact phase.

The minimum of free energy is reached at $R \to \infty$. At the force values corresponding to $\tilde{f}(f) \leq \tilde{f}(f_c)$ the whole system is in the globular phase and the radius R should be large and, in the thermodynamic limit, infinite.

Finally, at $\tilde{f}(f) \leq \tilde{f}(f_c)$ the free energy is equal to

$$\frac{\mathcal{F}_N}{N} \simeq -G. \tag{A13}$$

In the vicinity of $z = 1$ the series $\phi(z,\sigma)$ can be asymptotically estimated as

$$\phi(z,\sigma) \simeq \frac{2}{\log(\sigma)^2} - \frac{24 \log(1/z)}{\log(\sigma)^4} + 8 \times 5\,! \,\frac{\log(1/z)^2}{\log(\sigma)^6}. \tag{A14}$$

In the transition point $R \to \infty$ and the critical force f_c is estimated as

$$\tilde{f}(f_c) = \frac{G}{k_B T}. \tag{A15}$$

For the force values corresponding to $\tilde{f}(f) \geq \tilde{f}(f_c)$ close enough to the critical point the free energy can be estimated as

$$-\frac{\mathcal{F}_N}{k_B T N} \simeq \log s + s\,\frac{e^{-\tilde{f}(f)} - e^{-\tilde{f}(f_c)}}{1 + \frac{2}{\log(\sigma)^2} + \frac{24}{\log(\sigma)^4}}. \tag{A16}$$

If recall Eq. (A10), for $\tilde{f}(f) \to \tilde{f}(f_c) - 0$ the first derivative of the free energy is

$$\frac{\partial}{\partial \tilde{f}}\left(-\frac{\mathcal{F}_N}{k_B T N}\right) = 0. \tag{A17}$$

At the same time the first derivative is equal to

$$\frac{\partial}{\partial \tilde{f}}\left(-\frac{\mathcal{F}_N}{k_B T N}\right) = 1 \tag{A18}$$

for $\tilde{f}(f) \to \tilde{f}(f_c) + 0$. This indicates a discontinuity in the first derivative of free energy. Therefore the phase transition at the point $\tilde{f}(f) = \tilde{f}(f_c)$, defined by Eqs. (6) and (A15), is of the first order.

APPENDIX B: FINITE LENGTH CHAIN

From Eq. (10) it follows that partition function depends on large R and small r radii of torus in globular regions through the Zimm-Bragg parameters $\tilde{s}$ and $\tilde{\sigma}$ defined in Eq. (11)

$$Z_N = Z_N(R,r) = \sum_n \sum_{\{g_i,c_i\}} \tilde{s}(R,r)^{N_g} \tilde{\sigma}(R,r)^n. \tag{B1}$$

The last expression can be easily transformed to

$$\begin{aligned} Z_N &= \sum_n \sum_{N_g} \Omega(n,N_g)\tilde{s}^{N_g}\tilde{\sigma}^n \\ &= \sum_n \sum_{N_g} \exp[N_g \ln \tilde{s} + n \ln \tilde{\sigma} + \ln \Omega(n,N_g)], \end{aligned} \tag{B2}$$

where $\Omega(n,N_g)$ is the number of ways to arrange the numbers $\{g_i,c_i\}$ such as $\Sigma_i g_i=N_g$ and $\Sigma_i c_i=N-N_g$ In the mean-field approximation

$$Z_N \approx \exp \max_{n,N_g}[N_g \ln \tilde{s} + n \ln \tilde{\sigma} + \ln \Omega(n,N_g)], \tag{B3}$$

and the free energy of Zimm-Bragg model can be presented (see, e.g., in [34]) as

$$\beta\mathcal{F}_N(N_g,n,R,r) = N_g \Delta E + nE_s - S(N,N_g,n), \tag{B4}$$

where $S(N,N_g,n)=\ln \Omega(n,N_g)$ is mixing entropy of the coiled and globular regions,

$$E = -\ln \tilde{s} = \beta G + \tilde{f} + \frac{\tilde{l}^2}{R^2} \tag{B5}$$

and

$$E_s = -\ln \tilde{\sigma} = \alpha \frac{Rr}{\tilde{l}d}, \tag{B6}$$

where $p=\frac{\tilde{l}}{d}$ is the parameter that accounts for the rigidity of dsDNA. The normalization condition (2) can be transformed as

$$\langle g_i \rangle = \frac{N_g}{n} \simeq \frac{Rr^2}{\tilde{l}d^2}, \tag{B7}$$

where n is the average number of globular regions and N_g is the average number of monomers belonging to the globular phase. The free energy Eq. (B4) is minimized with Eq. (B7) condition using the undefined Lagrange multiplier μ. Minimization of the free energy Eq. (B4) over the parameters N_g, n, R, and r straightforwardly results in

$$\begin{aligned} R &\simeq \tilde{l}\alpha^{-2/5}\left(\frac{N_g}{n}\right)^{1/5} \\ r &\simeq d\alpha^{1/5}\left(\frac{N_g}{n}\right)^{2/5}, \end{aligned} \tag{B8}$$

where N_g and n are the equilibrium values, minimizing Eq. (B4). Using Eq. (B8), Zimm-Bragg parameter $\tilde{s}$ in Eq. (11) can be written as $\tilde{s}=\exp(-\Delta\tilde{E})$, where

$$\Delta\tilde{E} \simeq \beta G + \tilde{f} + \alpha^{4/5}\nu_g^{-2/5}, \tag{B9}$$

and $\nu_g=\frac{N_g}{n}$ is the mean length of the globular domain. The transition point of cooperative transitions (e.g., helix coil) is usually defined as the temperature, where the order parameter is equal to 1/2, in other words, as the temperature, at which the half of the system is in ordered state. In the language of Zimm-Bragg's theory, it corresponds to the condition $\tilde{s}=1$. Thus,

$$\tilde{f}(f_c) \simeq -\beta G - \alpha^{4/5}\nu_g^{-2/5}. \tag{B10}$$

The $\tilde{\sigma}$ parameter is responsible for the interval of transition between completely ordered (globular) and completely disordered (extended coil) phases.

Different from the usual cooperativity parameter of helix-coil transition, the parameter $\tilde{\sigma}$ is dependent on the mean dimensions of tore (R and r) and changing with the growth of the globular phase. However the interval of transition is defined mainly by the cooperativity parameter at the midpoint of transition and, therefore can be *approximately considered as constant.*

As reported in [25], near the transition point $\tilde{s}=1\pm\sqrt{\tilde{\sigma}}$. Thus, using Eqs. (B9) and (B10), the force (f) interval of transition can be estimated by equation

$$\Delta\tilde{f}(f) \simeq \sqrt{\tilde{\sigma}}. \tag{B11}$$

To express the interval of transition $\Delta\tilde{f}$ through model parameters, recall that in $N\to\infty$ limit (see Appendix A), the system exhibits first-order phase transition. The $\mathcal{F}$ versus θ curve exhibits two minima and there is a phase coexistence in the vicinity of the phase transition point $\tilde{f}_c$. Near the phase transition the probabilities to find the system in one of two phases are approximately equal, so

$$\exp(-\beta\mathcal{F}_{glob}) \sim \exp(-\beta\mathcal{F}_{coil}). \tag{B12}$$

Here $\mathcal{F}_{glob}$ and $\mathcal{F}_{coil}$ are free energies of completely globular and completely extended dsDNA chain, correspondingly. Thus, within the interval of transition,

$$|\mathcal{F}_{glob} - \mathcal{F}_{coil}| \leq \frac{1}{\beta}. \tag{B13}$$

Due to the estimate

$$\begin{aligned} \mathcal{F}_{coil} &\simeq -\frac{N\tilde{f}(f)}{\beta} \\ \mathcal{F}_{glob} &\simeq NG + \frac{\alpha^{4/5}N^{3/5}}{\beta}, \end{aligned} \tag{B14}$$

and since $\mathcal{F}_{glob}$ is independent from the external force value, with the help of Eq. (B14) the force interval of transition is found to scale as

$$\Delta\tilde{f}(f) \sim \frac{1}{N}. \tag{B15}$$

With taking into account Eq. (B11) the last equation gives

$$\tilde{\sigma} \simeq \frac{1}{N^2}. \tag{B16}$$

According to Zimm-Bragg theory [25] in the vicinity of transition point the average length of ordered region is estimated as $\nu_g \simeq \frac{1}{\sqrt{\tilde{\sigma}}}$. With taking into account Eq. (B16) obtaining

$$\nu_g \simeq N. \tag{B17}$$

Using estimates from Eqs. (B10) and (B17),

$$\tilde{f}(f_c) \simeq -\beta G - \alpha^{4/5} N^{-2/5}. \tag{B18}$$

APPENDIX C: FINITE LENGTH HETEROGENEOUS CHAIN

Derivations presented below are based mainly on approximate approach presented in [26]. For the sake of simplicity let us suppose that the chain consists of the two type of subsequences A and B and the A subsequences concentration is X. Then the free-energy difference between the coiled and globular phases for the subsequence α is $\Delta G_\alpha = G_\alpha + k_B T (\tilde{f} + \frac{\tilde{l}^2}{R^2})$, where $\alpha = A, B$ and R is the mean radius of globular domain [see Eq. (B8)]. G_α is the free energy of condensation of the subsequence α. In the condensed phase DNA stems are packed into the hexagonal lattice and each stem has six nearest neighbors. Thus, the free energy of condensation can be estimated as

$$G_A = -6(|F_2| + X\Delta F),$$

$$G_B = -6(|F_2| + (1-X)\Delta F), \tag{C1}$$

where $\Delta F = |F_1| - |F_2|$, F_1 is the free energy of attraction between homologous subsequences and F_2 is the free energy of attraction between heterologous subsequences. It is supposed $F_1 < F_2 < 0$. Thus, if $X \neq 1/2$, $G_A \neq G_B$.

Following the approach developed in [26] let us define function

$$G(t) = \sum_{\tau=1}^{t} \{\Delta G_A X(\tau) + \Delta G_B [1 - X(\tau)]\}, \tag{C2}$$

where $X(\tau)$ is the local concentration of the A subsequences in the point τ. $G(t)$ is the free-energy difference between the coiled and globular states in the long domain of the length $t \gg 1$. It can be transformed as

$$G(t) = t[\langle \Delta G_X \rangle_{seq} + \xi(t)\,\delta G], \tag{C3}$$

where $\langle \Delta G_X \rangle_{seq} = X G_A + (1-X) G_A + k_B T(\tilde{f} + \frac{\tilde{l}^2}{R^2})$ is the mean value of the free-energy difference between coiled and globular states, $\xi(t) = \frac{1}{t}\Sigma_{\tau=1}^{t} \eta(t)$ and $\eta(t)$ is the deviation of the local concentration $X(\tau)$ from X. For the long enough domain $(t \gg 1)$ $G(t) \approx t\langle \Delta G_X \rangle_{seq}$. The critical force value f_c is defined by condition

$$0 = \langle \Delta G_X \rangle_{seq}\big|_{f=f_c} \tag{C4}$$

Thus, the critical force value f_c is satisfied by equation

$$\tilde{f}(f_c) = -\beta \langle G \rangle_{seq} - \frac{\tilde{l}^2}{R^2}, \tag{C5}$$

where $\langle G \rangle_{seq} = X G_A + (1-X) G_A$. Taking into account Eq. (C4) the mean value of the free energy difference between coiled and globular states is found as

$$\langle \Delta G_X \rangle_{seq} = k_B T(\tilde{f} - \tilde{f}_X) = k_B T \delta\tilde{f}_X, \tag{C6}$$

where $\tilde{f}_X = X\tilde{f}^A + (1-X)\tilde{f}^B$ and $\tilde{f}^\alpha = -\beta G_\alpha - \frac{\tilde{l}^2}{R^2}$. The free energy of globularization of domain of the length l is

$$\Delta G(l) = G(t+l) - G(t+l) + F_s, \tag{C7}$$

where $F_s = -\frac{1}{\beta}\ln \tilde{\sigma}$ is the boundary free energy of a torous globule. According to Eq. (B16) it can be considered as constant, $F_s \sim \ln N$. The fluctuation of concentration necessary for globularization of domain of the length l is defined by condition $\Delta G(l) = 0$ and

$$\xi(l) = \frac{k_B T \delta\tilde{f}_X + F_s l^{-1}}{k_B T \delta\tilde{f}}, \tag{C8}$$

where $\delta\tilde{f} = \tilde{f}^A - \tilde{f}^B$ and $\tilde{f}^\alpha = -\beta G_\alpha - \frac{\tilde{l}^2}{R^2}$. The probability distribution for the random variable $\xi(l)$ can be estimated as GAUSSIAN,

$$P_l(\xi) = \left(\frac{2\pi\Omega^2}{l}\right)^{-1/2} \exp\left(-\frac{l\xi^2}{2\Omega^2}\right), \tag{C9}$$

where $\Omega^2 = X(1-X)$. Thus, the probability of formation of globular domain of the length l is found,

$$P_l \sim \frac{1}{\sqrt{l}} \exp\left(-\frac{[k_B T \delta\tilde{f}_X \sqrt{l} + F_s l^{-1/2}]^2}{2\Omega^2 (k_B T)^2 (\delta\tilde{f})^2}\right). \tag{C10}$$

The probability distribution Eq. (C10) has a sharp maximum at the $l_{\max} = \frac{F_s}{k_B T \delta\tilde{f}_X}$ and can be approximated as

$$P_{l_{\max}} \sim \exp\left(-\frac{\tilde{f} - \tilde{f}_X}{2\Delta\tilde{f}}\right), \tag{C11}$$

where $\Delta\tilde{f}$ is the width of transition,

$$\Delta\tilde{f} = \frac{\Omega^2 k_B T (\delta f)^2}{F_s} = \Omega^2 (\beta\delta G)^2 \frac{1}{|\ln \tilde{\sigma}|} \tag{C12}$$

and $\delta G = G_A - G_B$. The described above approach is approximate and the mean length of ordered domain ν_g can be estimated more precisely than $l_{\max}$ as

$$\nu_g \simeq \frac{|\ln \tilde{\sigma}|^2}{\Omega^2 (\beta\delta G)^2} \tag{C13}$$

using the method developed in [34].

[1] A. A. Kornyshev, D. J. Lee, S. Leikin, and A. Wynveen, Rev. Mod. Phys. **79**, 943 (2007).

[2] A. A. Kornyshev and S. Leikin, J. Chem. Phys. **107**, 3656 (1997).

[3] A. A. Kornyshev and S. Leikin, Phys. Rev. Lett. **82**, 4138 (1999).

[4] I. Rouzina and V. A. Bloomfield, J. Phys. Chem. **100**, 9977 (1996).

[5] R. Zhang and B. I. Shklovski, Physica A **349**, 563 (2005).

[6] B. I. Shklovskii, Phys. Rev. Lett. **82**, 3268 (1999).

[7] V. A. Parsegian, R. P. Rand, and D. C. Rau, Chem. Scr. **25**, 28 (1985).

[8] D. C. Rau and V. A. Parsegian, Biophys. J. **61**, 246 (1992).

[9] S. Leikin, D. C. Rau, and V. A. Parsegian, Proc. Natl. Acad. Sci. U.S.A. **91**, 276 (1994).

[10] J. Yang and D. C. Rau, Biophys. J. **89**, 1932 (2005).

[11] B. A. Todd and D. C. Rau, Nucleic Acids Res. **36**, 501 (2008).

[12] C. G. Baumann, V. A. Bloomfield, S. B. Smith, C. Bustamante, M. D. Wang, and S. M. Block, Biophys. J. **78**, 1965 (2000).

[13] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **94**, 6185 (1997).

[14] B. A. Todd, V. A. Parsegian, A. Shirahata, T. J. Thomas, and D. C. Rau, Biophys. J. **94**, 4775 (2008).

[15] Y. Murayama, Y. Sakamaki, and M. Sano, Phys. Rev. Lett. **90**, 018102 (2003).

[16] K. Besteman, K. Van Eijk, and S. Lemay, Nat. Phys. **3**, 641 (2007).

[17] N. V. Hud and I. D. Vilfan, Annu. Rev. Biophys. Biomol. Struct. **34**, 295 (2005).

[18] K. Yoshikawa, M. Takahashi, V. V. Vasilevskaya, and A. R. Khokhlov, Phys. Rev. Lett. **76**, 3029 (1996).

[19] Y. Murayama and M. Sano, J. Phys. Soc. Jpn. **70**, 345 (2001).

[20] G. S. Baldwin, N. J. Brooks, R. E. Robson, A. Wynveen, A. Goldar, S. Leikin, J. M. Seddon, and A. A. Kornyshev, J. Phys. Chem. B **112**, 1060 (2008).

[21] A. G. Cherstvy, A. A. Kornyshev, and S. Leikin, J. Phys. Chem. B **108**, 6508 (2004).

[22] A. Yu. Grosberg and A. V. Zhestkov, J. Biomol. Struct. Dyn. **3**, 859 (1986).

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

[24] S. Blumberg *et al.*, Biophys. J. **88**, 1692 (2005).

[25] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic Press, New York, 1970).

[26] A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules*, AIP Series in Polymers and Complex Materials (AIP, New York, 1994).

[27] I. Rouzina and V. A. Bloomfield, Biophys. J. **80**, 882 (2001).

[28] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, Nature (London) **424**, 338 (2003).

[29] L. D. Landau and E. M. Lifshitz, *Statistical Physics, Course of Theoretical Physics* (Butterworth-Heinemann, Oxford, 1984), Vol. 5.

[30] D. Poland and H. A. Sheraga, J. Chem. Phys. **45**, 1456 (1966).

[31] D. Poland and H. A. Sheraga, J. Chem. Phys. **45**, 1464 (1966).

[32] M. E. Fisher, J. Chem. Phys. **45**, 1469 (1966).

[33] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).

[34] A. A. Vedenov, A. M. Dykhne, and M. D. Frank-Kamenetskii, Sov. Phys. Usp. **14**, 715 (1972); A. A. Vedenov and A. M. Dykhne, Sov. Phys. JETP **28**, 187 (1969).

[35] L.-H. Tang and H. Chate, Phys. Rev. Lett. **86**, 830 (2001).

[36] T. Hwa, E. Marinari, K. Snippen, and L.-H. Tang, Proc. Natl. Acad. Sci. U.S.A. **100**, 4411 (2003).

[37] A. G. Cherstvy, J. Phys.: Condens. Matter **17**, 1363 (2005).