

Localized Single-Stranded Bubble Mechanism for Cyclization of Short Double Helix DNA

Jie Yan and John F. Marko

Department of Physics, University of Illinois at Chicago, 845 West Taylor Street, Chicago Illinois 60607-7059, USA
(Received 24 May 2004; published 3 September 2004)

Recent experiments indicate that double-stranded DNA molecules of approximately 100 base pairs in length have a probability of cyclization which is up to 10^5 times larger than that expected based on the known bending modulus of the double-helix. We argue that for short molecules, the formation of a few base pairs of single-stranded DNA can provide a “flexible hinge” that facilitates loop formation. A detailed calculation shows that this mechanism explains the experimental data.

DOI: 10.1103/PhysRevLett.93.108108

PACS numbers: 87.14.Gg, 82.35.Pq, 87.15.Aa, 87.15.La

Bending of stiff double-stranded DNA (dsDNA) into loops is essential to many processes in living cells. Two important examples include regulation of gene expression via contact of a nearby regulatory sequence to the beginning of a gene [1], and the packaging of DNA into nucleosomes [2], the basic structural unit underlying the chromosome. In these cases, DNA circles shorter than 30 nm (100 base pairs) form. This is remarkable since this is significantly shorter than the DNA persistence length of $A = 50$ nm (150 bp): such small loops are usually imagined to be possible only with the help of DNA-bending proteins.

This view has been challenged by a recent test-tube experiment [3] on 94 and 116 base pair DNAs showing loop formation probabilities $>10^4$ times larger than would be predicted by the persistence-length-based semiflexible polymer model of DNA bending [4] (see Fig. 1). The experiments also showed that there was a very strong sequence dependence of the loop formation probability. Longer DNAs (322 bp) formed loops with the probability expected from models based on semiflexible polymer theory, ruling out the possibility of experimental error.

In this Letter, we propose that these results can be explained by internal strand-separation fluctuations which transiently convert a double helix to much more flexible single-stranded DNA (ssDNA). Although energetically expensive and therefore rare excitations, these internal “bubbles” become *favorable* to smooth bends when forming loops of less than about 150 base pairs. A simple semiquantitative calculation supports this picture; we also present a detailed calculation using a novel transfer-matrix technique for calculation of end-to-end distances along semiflexible polymers. This indicates that a realistic estimate of the energetic cost of bubble formation leads to a large enough enhancement of cyclization probability for short DNAs to be a plausible explanation of the experimental data.

The conventional theory of cyclization of short DNAs uses the semiflexible polymer model [5]. The bending energy of a molecule of length L is taken to be that of thin-beam elastic theory [6]:

$$\frac{E}{k_B T} = \frac{A}{2} \int_0^L ds \left(\frac{\partial \hat{\mathbf{t}}}{\partial s} \right)^2. \quad (1)$$

Here, $\hat{\mathbf{t}}(s)$ is the unit tangent vector at arclength position s . Writing the energy in $k_B T$ units makes the dimension of the elastic constant A a length: it is called the *persistence length*, and is the correlation length for thermal fluctuations of the tangent vector [5]. To see this, consider the energy of a thermally excited bend by one radian in a length L , which is ($E \approx k_B T A/L$). This is comparable to

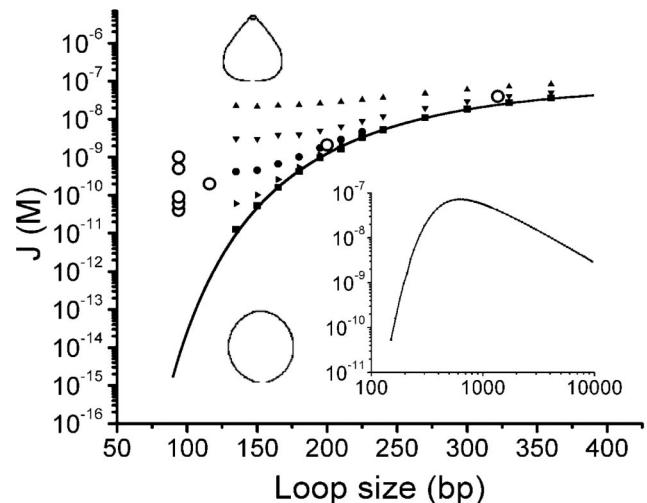


FIG. 1. Cyclization j -value for short DNAs. Experimental data [3] are indicated by open circles (\circ). Theoretical results are shown for the semiflexible polymer model (\blacksquare), and for the model of the text with $\mu/k_B T = 9$ (\blacktriangle), 10 (\blacktriangledown), 11 (\bullet), and 12 (\blacktriangleright). A bubble excitation energy of $11k_B T$ produces a large enhancement of the cyclization probability for chains smaller than 200 bp; note that in this short-chain regime, a small change in μ causes a large change in cyclization probability. Inset sketches indicate the circular and teardrop configurations discussed in the text. The solid line is the empirical formula from [14] for the semiflexible model. Inset shows j -value of the semiflexible polymer ($\mu = \infty$) calculated from 150 bp to 10^4 bp, showing a peak position ~ 500 bp.

$k_B T$ when $L \approx A$; thus thermally excited bends of about one radian will occur over regions of length roughly A .

The energy required to smoothly bend a DNA of length L into a circle to allow its cyclization will therefore be $E/k_B T = 2\pi^2 A/L$. When L is comparable to or smaller than A (50 nm or 150 bp), this energy becomes large compared to $k_B T$ (we note that the cyclization reaction, which is catalyzed by the enzyme T4 DNA ligase, requires the juxtaposed ends to be nearly parallel). For the 94 bp DNAs studied by Cloutier and Widom [3], this is $E = 31k_B T$, large enough to render cyclization of a 94 bp dsDNA unobservable.

The most likely explanation we see for the results of Cloutier and Widom [3] is based on excitation of a small region of strand-separated DNA. This will act as a “flexible hinge” since single-stranded DNAs are very flexible; the persistence length of ssDNA has been observed to be roughly 0.7 nm or one base [7] (the length of ssDNA is 0.7 nm per base, longer than that of the dsDNA which is 0.34 nm per helically folded base pair). Thus, the persistence length of a strand-separated bubble region will be roughly 2 bp, much shorter than the 150 base pair persistence length of dsDNA. Such a flexible hinge can greatly reduce the bending energy cost of cyclizing a short DNA.

However, we must also estimate the free energy cost of bubble formation. There are two contributions to this free energy: the sequence-dependent base-pairing and stacking free energy measured in experiments on melting of short DNAs [8]; for the conditions relevant to the experiments of Cloutier and Widom (25 C, 0.1 NaCl pH 7.5 aqueous solution), this ranges from $1k_B T$ to $4k_B T$ per base pair opened, depending on sequence. These estimates do not include the entropic cost of requiring the ssDNAs to close into a bubble. For a 3 bp bubble, this cost is approximately $3k_B T$ [9]. Therefore, the total free energy of a 3 bp bubble under the chemical conditions mentioned above ranges from $\epsilon = 6k_B T$ to $15k_B T$.

If a 3 bp bubble is excited near the middle of our molecule of length L , it will permit sharp bending at that point. However, we will still have to have some bending of the molecule to allow it to cyclize. The minimum-energy configuration that accomplishes this is the teardrop shape (Fig. 1 inset), which has a bending energy of 0.71 times the circle [10], or in $k_B T$ units, $14A/L$. Thus, the free energy difference between the bubble-teardrop configuration and the smoothly bent circle configurations is

$$E_{\text{teardrop}} - E_{\text{circle}} = k_B T \left[\epsilon - 5.7 \frac{A}{L} \right]. \quad (2)$$

Given $A = 150$ bp and $\epsilon = 10k_B T$, we can estimate the molecule length L^* at which the bubble-teardrop and circle configurations are equal in free energy by solving for when Eq. (2) is zero, giving $L^* = 85$ bp. This very simple calculation indicates that for sufficiently short

molecules, the bubble-hinge state will be lower in free energy than a smoothly bent state.

The above estimate, although suggestive, is not precise as it does not account for conformational fluctuations that play a major role in cyclization of short DNAs [11]. We have therefore carried out a detailed analysis of a model which combines double-helix bending with bubble-hinge excitations. Our model is a discretized version of Eq. (1), based on a series of N segments, each of length b in double-helix form; the total length of our polymer is $L = Nb$. Its Hamiltonian is:

$$\beta E = \sum_{i=1}^{N-1} \left[\frac{(\delta_{n_i,0} a + \delta_{n_i,1} a')}{2} (\hat{\mathbf{t}}_{i+1} - \hat{\mathbf{t}}_i)^2 + \beta \mu \delta_{n_i,1} \right], \quad (3)$$

where the $\hat{\mathbf{t}}_i$ are tangent vectors describing segment orientations. The n_i are two-state variables, indicating whether segment i is either in double-helix form ($n_i = 0$) or contains a ssDNA bubble ($n_i = 1$). In this Letter, the dsDNA bending elastic constant corresponds to the persistence length of (1) through $a = A/b$; the bending persistence length of the ssDNA bubble is ba' . We will use $b = 1$ nm (3 bp), and therefore $a = 50$, and $a' = 1$. The parameter μ is the free energy associated with creation of a bubble on a segment, and will be taken to be approximately $10k_B T$.

We wish to compute the thermal equilibrium probability density for the two end segments of this polymer to be found together and parallel to one another. This amounts to computing the expectation value of $\delta^2(\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_N) \delta^3(b \sum_{j=1}^{N-1} \hat{\mathbf{t}}_j)$ (the choice of $N-1$ as the upper limit for the latter sum corresponds to forcing overlap of the two end segments; an alternative calculation where the sum is taken to N gives the same results since we use a segment length of only 1 nm). Decomposing the three-dimensional delta function into wave number components gives us

$$J = \frac{1}{N_A} \int \frac{d^3 k}{(2\pi)^3} \int d^2 t_1 \cdots d^2 t_N \delta^2(\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_N) e^{i\mathbf{k} \cdot \sum_{j=1}^{N-1} \hat{\mathbf{t}}_j} \sum_{n_1, \dots, n_{N-1}} e^{-\beta E} \times \frac{1}{\int d^2 t_1 \cdots d^2 t_N \sum_{n_1, \dots, n_{N-1}} e^{-\beta E}}. \quad (4)$$

This quantity, expressed in units of mols/litre, or M, is often called “ j -factor” in the biochemical literature [11,12]. Here, N_A is Avogadro’s number.

After carrying out the sums over the n_i variables, (4) may be written in terms of a \mathbf{k} -dependent transfer matrix $T_{\mathbf{k}}(\hat{\mathbf{t}}, \hat{\mathbf{t}}') = e^{i\mathbf{k} \cdot \hat{\mathbf{t}}} [e^{-a(\hat{\mathbf{t}} - \hat{\mathbf{t}}')^2/2} + e^{-a'(\hat{\mathbf{t}} - \hat{\mathbf{t}}')^2/2 - \beta \mu}]$, as

$$J = \frac{1}{N_A} \int \frac{d^3 k}{(2\pi)^3} \frac{\text{Tr}(T_{\mathbf{k}}^{N-1})}{\int d^2 t_1 d^2 t_N T_{\mathbf{0}}^{N-1}}, \quad (5)$$

where the integrals over the tangent vectors $\hat{\mathbf{t}}_i$ in (4) are replaced by the matrix multiplications of (5). The $\delta^2(\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_N)$ in the numerator has become the trace in (5).

The matrix $T_{\mathbf{k}}$ can be computed in the basis of spherical harmonics to be

$$\begin{aligned} \langle lm|T_{\mathbf{k}}|l'm'\rangle &= \int d^2t d^2t' Y_{lm}^*(\hat{\mathbf{t}}) T_{\mathbf{k}}(\hat{\mathbf{t}}, \hat{\mathbf{t}}') Y_{l'm'}(\hat{\mathbf{t}}') \\ &= 4\pi(-1)^m \delta_{mm'} \sum_{l_2} i^{l_2} (2l_2 + 1) \sqrt{(2l + 1)(2l' + 1)} \begin{pmatrix} l_2 & l & l' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_2 & l & l' \\ 0 & m & -m \end{pmatrix} \\ &\quad \times [e^{-a} i_{l'}(a) + e^{-\beta\mu - a'} i_{l'}(a')] j_{l_2}(bk), \end{aligned} \quad (6)$$

using spherical harmonic expansions for the exponential functions in $T_{\mathbf{k}}$ and expressing all integrals of spherical harmonics in terms of $3j$ symbols[13]. Here, j_l and i_l are the spherical Bessel function and the modified spherical Bessel function of the first kind, respectively. The simple form of (6) allows the matrix multiplications of (5), and then the integral over \mathbf{k} to be done numerically with MATHEMATICA using the Gauss Kronrod method. The calculation is made finite by cutting off the l -sums at some maximum angular momentum: depending on the situation, we have used l_{\max} up to 18 to obtain convergence.

Calculation results are shown in Fig. 1. First, for $\mu = \infty$, bubbles never occur, and our model reduces to the conventional semiflexible polymer model (■). Our results are close to those from Monte Carlo[12] and approximate numerical calculations [11], showing a peak in the cyclization probability density near 500 bp, a long-length decay $\propto L^{-3/2}$ (Fig. 1, inset) [10], and most important for this Letter, a severe suppression for less than 300 bp which is in discord with the experimental data [3]. Calculations for <135 bp become inconveniently lengthy on the computers (AMD Opteron PCs) we have used, because the tightly bent configurations require a large $l_{\max} > 20$ for convergence of the calculation. Our computation can be extended to smaller chains by the use of more powerful computers. Our results for the simple semiflexible chain agree with the result of Shimida and Yamakawa (Fig. 1, solid curve) [14].

The filled circle points in Fig. 1 show the cyclization probability density for $\mu = 11k_B T$. For large L , the result is essentially identical to that for $\mu = \infty$: the ssDNA bubbles are too rare to change the large-scale polymer properties. However, below the 500 bp peak, the cyclization probability does not show a rapid decrease, and passes close to the experimental data. Even rare appearances (probability per segment $e^{-11} \approx 2 \times 10^{-5}$) of flexible joints along a 150 bp DNA boost the probability of cyclization by more than 100 times.

An important feature of the experimental data for 94 bp DNAs is a strong sensitivity to sequence composition. The six different 94 bp molecules studied by Cloutier and Widom [3] show cyclization probabilities varying

over a range of nearly 100. Our model, being based on rare fluctuations, depends exponentially on their energy μ . The triangular points of Fig. 1 shows the cyclization probability for $\mu = 9$ (▲), 10 (▼), and 12 (►) $k_B T$. A change in μ by $1k_B T$ causes a roughly tenfold change in cyclization probability for 135 bp. Extension of the calculation to incorporate sequence dependence is straightforward; the main problem is reliable prediction of the sequence dependence of ssDNA bubbles since existing models for sequence dependence of DNA melting [8] are based on measurements of strand separation of whole molecules, and may not be well calibrated for small internal bubbles.

There should be strong sensitivity of cyclization probability for short molecules, to even small changes in the probability of ssDNA bubbles. The strong enhancement of cyclization probability observed experimentally at 25 C should thus be enormously sensitive to temperature, since AT-rich DNA sequences begin to melt by 50 C. At 35 C, the free energy cost of opening AT-rich 3 bp regions will thus be reduced by a few $k_B T$.

An interesting effect reported by Cloutier and Widom is the quite strong sensitivity of cyclization probability to molecular length. A 1 bp change in length was found to generate, in some cases, a greater than two-fold change in j factor. While this dependence on molecular details is beyond the scope of our model, it emphasizes how sensitive the results are to molecular architecture.

It is possible that a portion of the cyclization enhancement observed might be due to permanent bends [15]. However, the experiments of Cloutier and Widom suggest that permanent bends are not the main factor driving the anomalously large cyclization. The molecules used do not contain known sequence motifs causing the sharp permanent bends necessary to compete with our flexible bubble mechanism. Furthermore, a sharp permanent bend should lead to anomalous electrophoretic migration, which was not observed (see D bands of Fig. 5 of Ref. [3]). Thus, we are in accord with the conclusion of Cloutier and Widom that permanent bends are not responsible for the bulk of the effect they observe. However, the role of spontaneous bending could be experimentally determined by carrying out cyclization experiments on 94 bp molecules engi-

neered to carry known spontaneously bent sequences. Our calculations may be generalized to the case where a fixed bend is present in the molecule [16] to predict the effect of the bend in this case.

We have argued that cyclization of short DNAs proceeds via formation of a localized structural defect in the double helix, which turns out to be a transition state lower in free energy than a smooth bend. This type of defect is extremely rare and in most experimental situations can be ignored. However, in this interpretation of the cyclization experiment of Cloutier and Widom [3], tightly bent configurations are selected, and in that subset of conformations, flexible-hinge bubbles dominate. Finally, we note that our transfer-matrix approach makes possible essentially exact numerical calculation of end-to-end distributions for many variants of the semiflexible polymer model relevant to biophysical experiments; for example, models for folding of DNA by proteins that bind along its length [16].

This research was supported by the NSF through Grant No. DMR-0203963. We thank J. Widom and R. Owczarzy for helpful discussions.

-
- [1] S. Oehler, M. Amouyal, P. Kolkhof, B. Wilcken-Bergmann, and B. Muller-Hill, *EMBO J.* **13**, 3348 (1994).

- [2] T. J. Richmond and C. A. Davey, *Nature (London)* **423**, 145 (2003).
- [3] T. E. Cloutier and J. Widom, *Mol. Cell* **14**, 355 (2004).
- [4] P. J. Hagerman, *Annual Review of Biophysics and Biophysical Chemistry* **17**, 265 (1988).
- [5] M. Doi and S. F. Edwards, *Theory of Polymer Dynamics* (Oxford University, New York, 1985).
- [6] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, New York, 1986).
- [7] S. B. Smith, Y. Cui, and C. Bustamante, *Science* **271**, 795 (1996).
- [8] J. Santalucia, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1460 (1998).
- [9] D. H. Matthews, J. Sabina, M. Zuker, and D. H. Turner, *J. Mol. Biol.* **288**, 911 (1999), see Table 17.
- [10] H. Yamakawa and W. H. Stockmayer, *J. Chem. Phys.*, **57**, 2843 (1972).
- [11] Y. Zhang and D. M. Crothers, *Biophys. J.* **84**, 136 (2003).
- [12] A. Podtelezhnikov and A. Vologodskii, *Macromolecules* **30**, 6668 (1997).
- [13] Eric W. Weisstein, Wigner 3j-Symbol, from MathWorld—A Wolfram Web Resource, <http://mathworld.wolfram.com/Wigner3j-Symbol.html>.
- [14] J. Shimada and H. Yamakawa, *Macromolecules* **17**, 689 (1984).
- [15] D. M. Crothers, T. E. Haran, and J. G. Nadeau, *J. Biol. Chem.* **265**, 7093 (1990).
- [16] J. Yan and J. F. Marko, *Phys. Rev. E* **68**, 011905 (2003).