



# HHS Public Access

Author manuscript

*J Phys Chem Lett.* Author manuscript; available in PMC 2017 March 17.

Published in final edited form as:

*J Phys Chem Lett.* 2016 March 17; 7(6): 1042–1046. doi:10.1021/acs.jpcllett.6b00246.

## Sequence Affects the Cyclization of DNA Minicircles

Qian Wang and B. Montgomery Pettitt\*

Department of Biochemistry and Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, Texas 77555-0304, United States

### Abstract

Understanding how the sequence of a DNA molecule affects its dynamic properties is a central problem affecting biochemistry and biotechnology. The process of cyclizing short DNA, as a critical step in molecular cloning, lacks a comprehensive picture of the kinetic process containing sequence information. We have elucidated this process by using coarse-grained simulations, enhanced sampling methods, and recent theoretical advances. We are able to identify the types and positions of structural defects during the looping process at a base-pair level. Correlations along a DNA molecule dictate critical sequence positions that can affect the looping rate. Structural defects change the bending elasticity of the DNA molecule from a harmonic to subharmonic potential with respect to bending angles. We explore the subelastic chain as a possible model in loop formation kinetics. A sequence-dependent model is developed to qualitatively predict the relative loop formation time as a function of DNA sequence.

### Graphical abstract

---

\*Corresponding Author: mpettitt@utmb.edu.

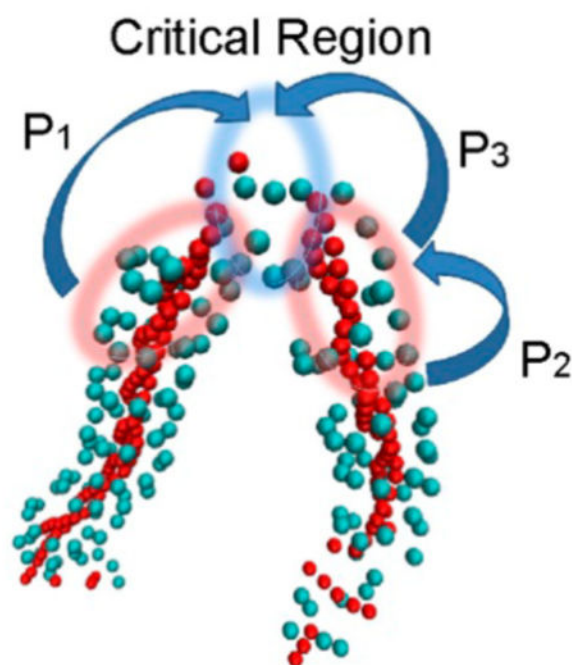
#### Notes

The authors declare no competing financial interest.

#### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcllett.6b00246.

DNA sequences studied in the work; protocol for the free-energy calculations, cyclization simulation protocol and angle distribution measurement; illustration of the kinks and bubbles existing in the cyclization trajectory; distribution of the time of appearance of structural defects  $T_{\text{defects}}$  divided by the time of looping  $T_{\text{loop}}$  for model sequence TA; and probabilities of angles between DNA segments with different numbers of mismatched base pairs. (PDF)



The formation of circular DNA is critical to many molecular cloning technologies, in particular, for certain gene therapies.<sup>1</sup> A central question is how the sequence of a DNA molecule affects its various dynamic and structural properties when there are geometric restrictions.<sup>2</sup> The circularization or looping of short DNA segments<sup>3–5</sup> has drawn considerable experimental attention.<sup>6–9</sup> DNA looping is also a critical step in cellular processes such as gene transcription and expression. Under many circumstances, transcription is regulated by binding a protein to a DNA loop. Interestingly, the DNA loop can be very short (<150bp) and form quickly. A typical example is the Lac Repressor–DNA loop. Thus, studying the looping of short DNA segments<sup>3–5</sup> is central to a variety of processes. Within the 150 base pair (bp) persistence length,<sup>10</sup> unconstrained double-stranded DNA (dsDNA) has been often described by the wormlike chain (WLC) model; however, the cyclization rate of DNA fragments <100 bps is orders magnitude larger than the WLC predictions in experimental studies.<sup>11–14</sup> There are several hypotheses to explain this phenomenon that involve some form of nonlinear elasticity, including forming transient kinks<sup>15,16</sup> (KWLC model), bubbles,<sup>17</sup> and the linear subelastic chain model<sup>12</sup> (LSEC).

Many previous theoretical works have treated DNA as a sequence nonspecific polymer. Thus, a central question, how the sequence of a DNA molecule affects its various dynamic properties, has not been well addressed. Understanding the consequences of coupling mechanical defects such as kinks and bubbles to the kinetics could allow one to manipulate the looping rate by sequence variations. We show that sequence determines possible mechanical geometries that dictate the kinetics of cyclization. We find that experimentally measured DNA looping rates are consistent with locally sublinear elasticity created by specific transient sequence dependent defects. Our theory qualitatively predicts the relative loop formation time as a function of DNA sequence.

Here we first investigate the cyclization process using molecular dynamics simulations with a coarse-grained model. Simulations provide a route to study the dynamic consequences from structural information at the base pair level<sup>18–20</sup> and include the sequence information. Through simulations, we identify the critical sequential positions in the cyclization process. Next, we will use this information to develop a Markov model to predict the looping time for various sequences.

The issue for simulating the cyclization process is the long time scale (up to hours). We consider a sequence-dependent a coarse-grained (CG) model.<sup>21–24</sup> We note that results from this CG model are not strictly representative of real time, but the time elapsed can be considered to scale roughly with the number of integration time steps. In addition, we sample with a biasing potential to further accelerate the simulations. The biased potential was devised to be weak enough to reduce but not eliminate the free-energy barrier to cyclization.

DNA sequence named “TA” (see supplement or ref 14) was chosen to compare with experiments,<sup>14</sup> which provide the looping time of “TA” and four other DNA molecules with the same length of “TA” but with different sequences. This makes it possible for us to compare our analytical predictions directly to the experimental observations. The total length of “TA” is 73bp (63bp +10-nt overhang), considerably shorter than the average persistence length. During cyclization, the two complementary overhangs can form contacts if they are spatially close enough, leading to a stable loop (Figure 1A).

The free-energy profile for DNA cyclization is shown in Figure 1B (black curve). It is clear that there is a modest free-energy barrier during the cyclization process. Because of increasing curvature, the free energy gradually increases when the distance between residues 72 and 82,  $R_{72-82}$ , reduces from  $27\sigma$  to  $6\sigma$ . When  $R_{72-82} < 6\sigma$ , the free energy decreases sharply because new complementary contacts begin to form and stabilize the system. This can be inferred from the fact that the probability of base-pair opening for bp72 (Figure 1C) quickly decreases from one to close to zero at this region; however, this contact still has ~5% probability to reopen. This is not surprising considering that the nick between bp72 and bp0 should increase the instability (Figure 1A). In contrast, for bp11 that is far away from the nick, the probability of bp opening remains essentially at zero. Interestingly, bp25, which is also far away from the nick, has ~5% probability to open. Those results indicate that the probability of base-pair opening depends on sequence. After the pulling force was assigned, the free-energy barrier still exists in the range from  $R_{72-82} = 6-10\sigma$  (Figure 1B, red curve). Thus, although our cyclization simulation is biased, it provides useful information on how a linear-shaped DNA molecule searches dynamically to form a loop.

Next, cyclization simulations were performed (see Supporting Information). The percentage of unlooped structures as a function of time is shown in Figure 2. The looping rate can be obtained by fitting the curve to a single exponential function ( $R > 0.99$ ), which matches with experiments. Structural defects were found to be inevitable in the cyclization process. In all 200 looping trajectories, we found no single successful cyclization trajectory with an absence of kinks or bubbles (Supplemental Figure 1).

The probability of each base pair of sequence TA to open as a function of time is shown in Figure 3A. We found that the two dominant regions having bp openings are 24–26 and 35–37, while other base pairs are quite stable. This result remains the same regardless of the strength of the external biased potential  $k = 0.003$  or  $0.0015 \epsilon/\sigma^2$ . AT tracts are known to be the thermodynamic “weak points” along a dsDNA based on multiple previous studies;<sup>25,26</sup> kinks have been observed to concentrate at TpA base stacks in both simulations<sup>27,28</sup> and experiments.<sup>29–31</sup> Thus, one might assume that bp24–26 and bp35–37 would be weak points; however, bp24–26 is “CCA” and bp35–37 is “GCA”, both thermodynamically more stable than bp15–20 (ATTTT).

A probable explanation for this phenomenon is that for a highly bent structure, the location of the structural defects is controlled by a positive correlation along a DNA minicircle for bp  $i$  and bp  $i+N/2$ , in conjunction with its local thermodynamic stability. This positive correlation was also observed in a previous experimental study.<sup>32</sup> For the sequence TA, bp0 and bp63 are unstable because there are nicks next to them (Figure 1A). Because of the positive correlation, base pairs near bp36 ( $= 0 + 73/2$ ) and bp26 ( $= 63 - 73/2$ ) would be predicted to be unstable as well.

The relation between the relative time of structural defects appearing,  $T_{\text{defects}}$ , and looping  $T_{\text{loop}}$  is shown in Figure 3B. When  $T_{\text{defects}}/T_{\text{loop}} = 1$ , loop formation happens as structural defects occur, suggesting that those defects are critical in the cyclization process. For a random sequence location bp16–18, the distribution of  $T_{\text{defects}}/T_{\text{loop}}$  is near that expected for a random process (Supplemental Figure 2A). In contrast, for bp24–26 and bp35–37, the probability is dramatically increased,  $T_{\text{defects}}/T_{\text{loop}} \approx 1$ , especially for bp24–26 (Figure 3B). Our results demonstrate for our test sequence that bp24–26 and bp35–37 are two critical positions for the cyclization process. Note that when the length of overhang equals zero, the two critical positions merge into one position locating the center of the sequence.

Because bp24–26 and bp35–37 have the highest frequency of bp openings and they are strongly correlated with the final cyclization, we hypothesize that mutation at those places will greatly affect the loop closure rate. Three sequences were designed: TA\_mut, mutation from C/G to A/T at bp24–26 and bp35–37; TA\_mis1, mutation from CC to AA in one strand of TA at bp24 (so there are 2 bp mismatched); and TA\_mis2, similar to TA\_mis1 but the mutation is at bp11. For TA\_mut, the looping rate is increased 1.5 times compared with TA (Figure 2, red curve). This is because the mutation from stable base pairs C/G to less stable base pairs A/T at bp24–26 and bp35–37 increases the probability to form defects at those positions. In addition, creating bp mismatches at those two places would further accelerate the looping process because a bubble is highly probable. Indeed, TA\_mis1 loops 1 order of magnitude faster than TA (Figure 2). To exclude other possible influences brought by bp mismatches, we created similar mismatches at bp11, and its looping rate was found to be about the same as the original TA model sequence. Our results clearly show that in cyclization, position bp( $x+N/2$ ) controls the cyclization rate, where  $x$  is an unstable position (such as having nicks nearby) and  $N$  is the total bp number. This is a geometric result that allows a sharp bend to bring the ends together kinetically in the most thermodynamically economical way.

To investigate how structural defects affect the elastic properties of a DNA molecule, we ran simulations for an 18bp dsDNA with different bp mismatches (see Supporting Information). The angular distribution of DNA segments with 7bp length is shown in Figure 4. For NM (number of mismatched bps) = 0, it is widely accepted that free DNA behaves like an elastic rod so the bending energy is approximately  $E = k\theta^2$  with probability  $P \approx \sin \theta e^{-\beta k\theta^2}$ . Thus, its potential of mean force  $-\ln(P)$  should be well fit by  $y = A - \ln(\sin \theta) + B\theta^2$ . Indeed, the fitting quality is very good ( $R = 0.994$ ), fully supporting the idea that the bending energy varies essentially quadratically with the bending angle. When NM increases, the fitting quality decreases, indicating that the bending energy gradually deviated from the harmonic assumption. When NM = 5, a linear elasticity model  $y = A - \ln(\sin \theta) + B\theta$  provides a higher quality fit. We fitted the resulting angular distribution using the form  $y = A - \ln(\sin \theta) + B\theta^a$  (Supplemental Figure 3B). When NM = 0,  $a = 2.1$ , close to the harmonic model. When NM = 5,  $a$  decreases to 1.2, close to the linear model. Our results demonstrate that structural defects can change the elastic properties of a DNA molecule from harmonic to subharmonic (linear) behavior, which greatly favors the cyclization.

A so-called subelastic chain model<sup>12</sup> (LSEC) was proposed to explain the high bendability of certain DNA molecules. LSEC is based on the observation that if a long DNA molecule is divided into small segments  $\sim 7$ bp, the bending energy increases linearly, instead of quadratically, with the bending angle; however, LSEC is an empirical energy function with an unclear mechanistic origin. Our CG simulations show that, a “bare” or unconstrained DNA is well described by the WLC model. Only when the total number of mismatches, NM, is large enough can the elasticity of the DNA be approximately described by the LSEC model. This also probably explains why the WLC model is good enough for studies without considering the magnesium;<sup>33,34</sup> the probability of transient structural defects decreases in the absence of magnesium. LSEC model is powerful when the probability of defects is large enough (large NM); however, moderate NM can cause the elasticity of a DNA molecule to display a behavior between harmonic and linear form (Supplemental Figure 3B). Thus, many conditions such as the concentration of magnesium or the length of an AT track in the circle can affect the validity of the LSEC model.

On the basis of the above discussions, a sequence-dependent model (Figure 5A) is proposed to predict the looping time for a given DNA sequence. The sequence positions were divided to two regions: (1) Region C, the critical position to affect the looping time. This region contains the bp index  $|i - N/2| < \delta_1$  and  $|i - (N/2 - s)| < \delta_1$ .  $N$  is the total bp number,  $s$  is the length of overhangs (= 10bp in our simulation), and  $\delta_1$  describes the region width. (2) Region NC, this region is adjacent to the region C. If any base pair has defects in the region NC, the local stress will be released by local melting or kinking, and the probability of forming defects in region C actually decreases. In other words, region NC is negatively correlated with region C. This negative correlation has been studied in our previous work.<sup>23</sup> This region includes bp  $i$  satisfying  $\delta_1 < |i - N/2| < \delta_2$ , and  $\delta_1 < |i - (N/2 - s)| < \delta_2$ . We take  $\delta_1 = 1$  and  $\delta_2 = 5$  in this work, but it will not affect the results qualitatively. We neglect all other regions except region N and NC because they have a minor influence on the cyclization time (Figure 2, black and blue curves).

We modeled the looping as a Markov process searching positions that have structural defects. Any base pair  $i$  can have a defect with a probability  $p(i)$ . If  $i$  belongs to region C, then the structure will loop immediately. If  $i$  belongs to region NC; then, the structure cannot loop and has to wait for the next search until a defect happens in the region C. Normalized  $p(i)$  satisfies

$$p(i) = e^{-\Delta G(i-1, i, i+1)/k_B T} / \sum_{i \in C, i \in NC} e^{-\Delta G(i-1, i, i+1)/k_B T}$$

$G(i-1, i, i+1)$  is the free energy of having defects at base pair  $i$ , when its neighbor is  $i-1$  and  $i+1$ . Note that the negative correlation is implied by this normalization process because  $p(i \in NC)$  increases will lead to  $p(i \in C)$  decreases. As we have noted before,  $G$  can be decomposed as Kirkwood superposition relationship

$$\Delta G(i-1, i, i+1) = \Delta G(i-1, i) + \Delta G(i, i+1) - \Delta G^o(i)$$

The values of  $G(i-1, i)$ ,  $G(i, i+1)$ , and  $G^o(i)$  can be found in our previous work.<sup>23</sup> To reduce the error of estimation in region C, we assume

$$p(i) = \max(p(k)), i \in C, k \in C$$

The looping time  $T$  is approximated as  $T = X * T_0$ .  $X$  is how many positions have ever been chosen during the Markov searching process before loop finishes and  $T_0$  is a time unit chosen to scale the CG time to real time. We cannot determine the exact value of  $T_0$ , but it should not qualitatively affect the result. Predictions on six different DNA sequences are shown in Figure 5B. We predicted that the looping time order should be TA < E8A10 ~ E8A17 < R73 < E8A26 ~ E8A38, matching with the experimental result<sup>14</sup> very well.

In conclusion, both kinks and bubbles were observed in the dynamic simulations, which provides the mechanical origin of the subelastic model. As long as there are these types of structural defects, the subelastic model is valid to describe the looping process of small DNA circles or so-called mini circles. Several conditions are correlated to induce such structural defects; they depend on the size the loop, external torsional stress, magnesium concentrations, and many other factors. Here we point out that the defects related to the looping process happen in specific, critical sequential positions. The probability to have structural defects can be affected by the sequence in the critical regions and the sequence in the adjacent regions through negative correlations. This appears to be a general rule and could be used in studies of the relative looping time of various sequences with similar lengths.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

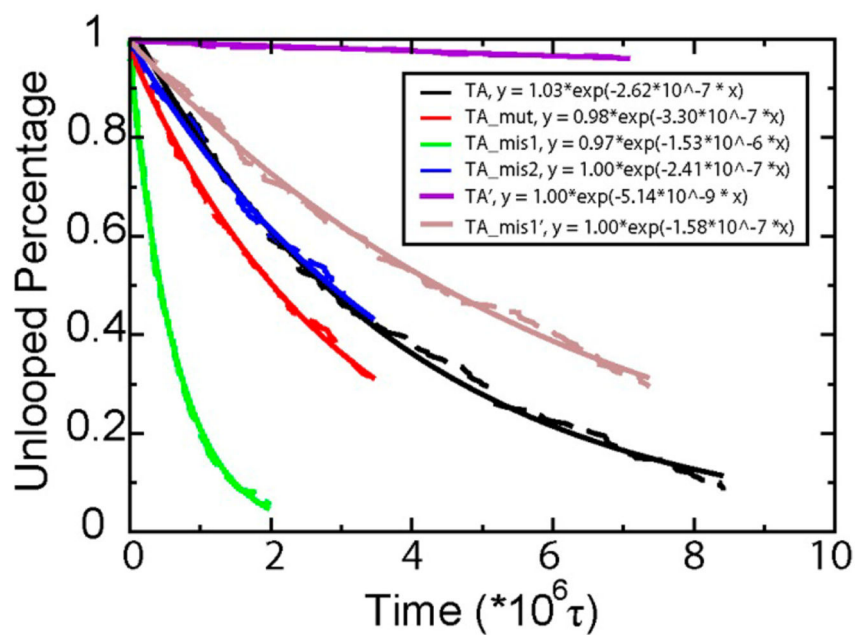
B.M.P. acknowledges many interesting conversations with Prof. Lynn Zecheidrich about DNA minicircles. We gratefully acknowledge the financial support of the National Institutes of Health (GM 066813) and the Robert A. Welch Foundation (H-0037). A portion of the research was performed via the NSF Xsede and in conjunction with the Texas Advanced Computing Center (TACC) at The University of Texas at Austin. The Sealy Center for Structural Biology scientific computing staff is acknowledged for computational support.

## References

1. Zhao N, Fogg JM, Zechiedrich L, Zu Y. Transfection of shRNA-encoding Minivector DNA of a few hundred base pairs to regulate gene expression in lymphoma cells. *Gene Ther.* 2011; 18(3):220–224. [PubMed: 20962872]
2. Irobalieva RN, Fogg JM, Catanese DJ, Sutthibutpong T, Chen M, Barker AK, Ludtke SJ, Harris SA, Schmid MF, Chiu W, Zechiedrich L. Structural diversity of supercoiled DNA. *Nat Commun.* 2015; 6:8440. [PubMed: 26455586]
3. Crothers DM, Drak J, Kahn JD, Levene SD. DNA Bending, Flexibility, and Helical Repeat by Cyclization Kinetics. *Methods Enzymol.* 1992; 212:3–29. [PubMed: 1518450]
4. Levene SD, Giovan SM, Hanke A, Shoura MJ. The thermodynamics of DNA loop formation, from J to Z. *Biochem Soc Trans.* 2013; 41:513–518. [PubMed: 23514145]
5. Manzo, C., Finzi, L. *Methods in Enzymology, Vol 475: Single Molecule Tools, Pt B: Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods.* Vol. 475. Elsevier Academic Press, Inc; San Diego: 2010. Quantitative Analysis of DNA-Looping Kinetics from Tethered Particle Motion Experiments; p. 199-220.
6. Dickerson RE, Chiu TK. Helix bending as a factor in protein/DNA recognition. *Biopolymers.* 1997; 44(4):361–403. [PubMed: 9782776]
7. Vilar JMG, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev.* 2005; 15(2):136–144. [PubMed: 15797196]
8. Wong OK, Guthold M, Erie DA, Gelles J. Interconvertible lac repressor-DNA loops revealed by single-molecule experiments. *PLoS Biol.* 2008; 6(9):e232. [PubMed: 18828671]
9. Purohit PK, Kondev J, Phillips R. Mechanics of DNA packaging in viruses. *Proc Natl Acad Sci U S A.* 2003; 100(6):3173–3178. [PubMed: 12629206]
10. Baumann CG, Smith SB, Bloomfield VA, Bustamante C. Ionic effects on the elasticity of single DNA molecules. *Proc Natl Acad Sci U S A.* 1997; 94(12):6185–6190. [PubMed: 9177192]
11. Yuan CL, Chen HM, Lou XW, Archer LA. DNA bending stiffness on small length scales. *Phys Rev Lett.* 2008; 100(1):4.
12. Wiggins PA, Van der Heijden T, Moreno-Herrero F, Spakowitz A, Phillips R, Widom J, Dekker C, Nelson PC. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat Nanotechnol.* 2006; 1(2):137–141. [PubMed: 18654166]
13. Cloutier TE, Widom J. Spontaneous sharp bending of double-stranded DNA. *Mol Cell.* 2004; 14(3):355–362. [PubMed: 15125838]
14. Vafabakhsh R, Ha T. Extreme Bendability of DNA Less than 100 Base Pairs Long Revealed by Single-Molecule Cyclization. *Science.* 2012; 337(6098):1097–1101. [PubMed: 22936778]
15. Vologodskii A, Frank-Kamenetskii MD. Strong bending of the DNA double helix. *Nucleic Acids Res.* 2013; 41(14):6785–6792. [PubMed: 23677618]
16. Wiggins PA, Phillips R, Nelson PC. Exact theory of kinkable elastic polymers. *Phys Rev E.* 2005; 71(2):19.
17. Yan J, Marko JF. Localized single-stranded bubble mechanism for cyclization of short double helix DNA. *Phys Rev Lett.* 2004; 93(10):4.
18. Mitchell JS, Laughton CA, Harris SA. Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA. *Nucleic Acids Res.* 2011; 39(9):3928–3938. [PubMed: 21247872]
19. Lankas F, Lavery R, Maddocks JH. Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure.* 2006; 14(10):1527–1534. [PubMed: 17027501]

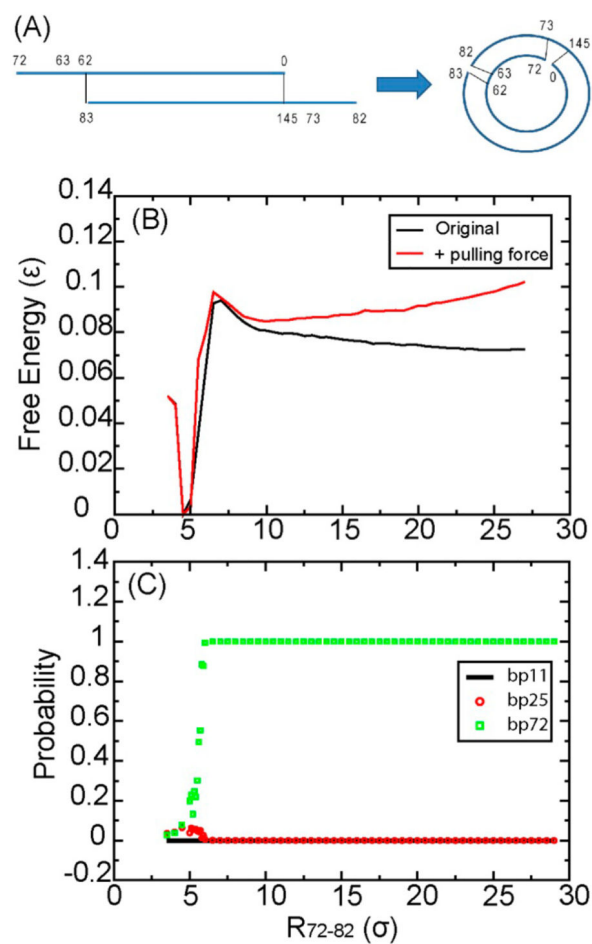
20. Merlitz H, Rippe K, Klenin KV, Langowski J. Looping dynamics of linear DNA molecules and the effect of DNA curvature: A study by Brownian dynamics simulation. *Biophys J*. 1998; 74(2):773–779. [PubMed: 9533690]
21. Ouldridge TE, Louis AA, Doye JPK. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J Chem Phys*. 2011; 134(8):085101. [PubMed: 21361556]
22. Sulc P, Romano F, Ouldridge TE, Rovigatti L, Doye JPK, Louis AA. Sequence-dependent thermodynamics of a coarse-grained DNA model. *J Chem Phys*. 2012; 137(13):135101. [PubMed: 23039613]
23. Wang Q, Pettitt BM. Modeling DNA Thermodynamics under Torsional Stress. *Biophys J*. 2014; 106(5):1182–1193. [PubMed: 24606942]
24. Wang Q, Myers CG, Pettitt BM. Twist-Induced Defects of the P-SSP7 Genome Revealed by Modeling the Cryo-EM Density. *J Phys Chem B*. 2015; 119(15):4937–4943. [PubMed: 25793549]
25. SantaLucia J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A*. 1998; 95(4):1460–1465. [PubMed: 9465037]
26. SantaLucia J, Hicks D. The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct*. 2004; 33:415–440. [PubMed: 15139820]
27. Randall GL, Zechiedrich L, Pettitt BM. In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucleic Acids Res*. 2009; 37(16):5568–5577. [PubMed: 19586933]
28. Kannan S, Kohlhoff K, Zacharias M. B-DNA under stress: over-and untwisting of DNA during molecular dynamics simulations. *Biophys J*. 2006; 91(8):2956–2965. [PubMed: 16861282]
29. Drew HR, Weeks JR, Travers AA. Negative Supercoiling Induces Spontaneous Unwinding of a Bacterial Promoter. *EMBO J*. 1985; 4(4):1025–1032. [PubMed: 2990904]
30. Travers A. DNA dynamics: Bubble ‘n’ flip for DNA cyclisation? *Curr Biol*. 2005; 15(10):R377–R379. [PubMed: 15916938]
31. Protozanova E, Yakovchuk P, Frank-Kamenetskii MD. Stacked-unstacked equilibrium at the nick site of DNA. *J Mol Biol*. 2004; 342(3):775–785. [PubMed: 15342236]
32. Lionberger TA, Demurtas D, Witz G, Dorier J, Lillian T, Meyhofer E, Stasiak A. Cooperative kinking at distant sites in mechanically stressed DNA. *Nucleic Acids Res*. 2011; 39(22):9820–9832. [PubMed: 21917856]
33. Schopflin R, Brutzer H, Muller O, Seidel R, Wedemann G. Probing the Elasticity of DNA on Short Length Scales by Modeling Supercoiling under Tension. *Biophys J*. 2012; 103(2):323–330. [PubMed: 22853910]
34. Shi XS, Herschlag D, Harbury PAB. Structural ensemble and microscopic elasticity of freely diffusing DNA by direct measurement of fluctuations. *Proc Natl Acad Sci U S A*. 2013; 110(16):E1444–E1451. [PubMed: 23576725]



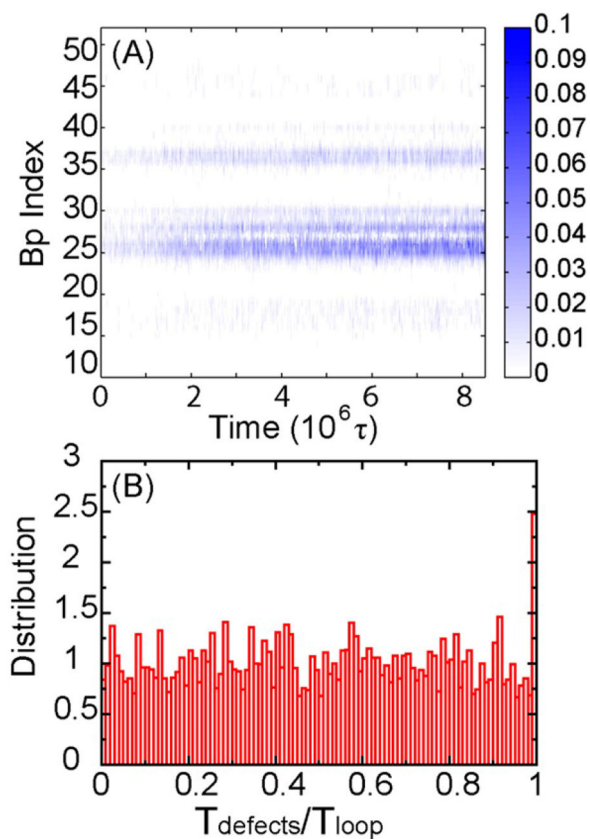


**Figure 1.**

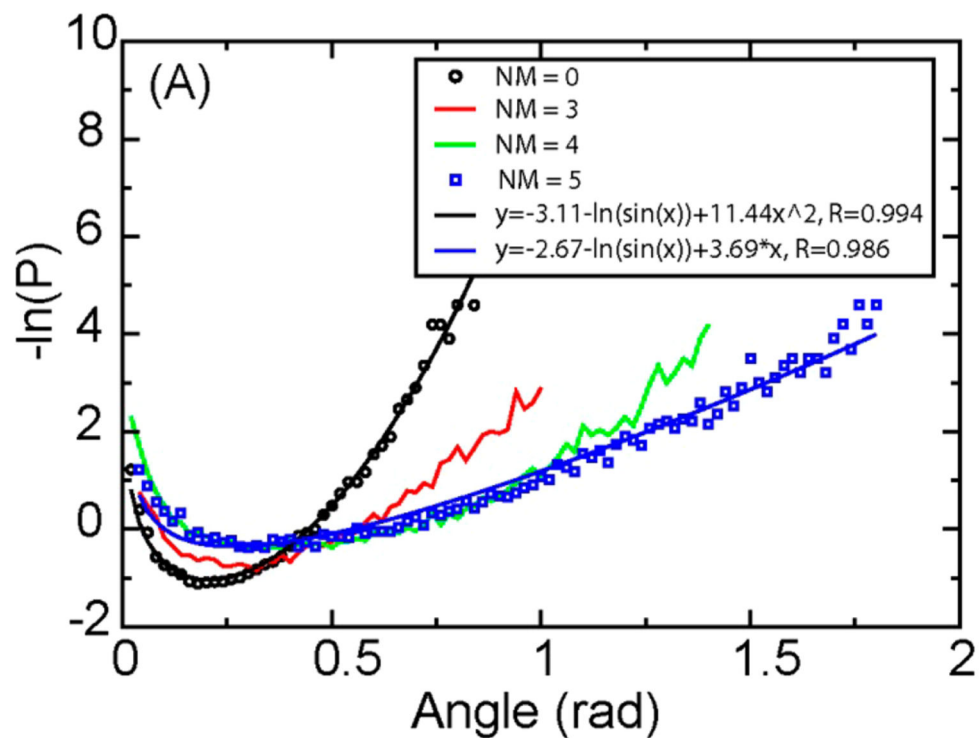
(A) Illustration of the cyclization process of a dsDNA with total bp = 73. The numbers in the Figure show the bp index. Two overhangs are at nucleotides 63–72 in one strand and nt73–82 in its complementary strand. (B) Free-energy profile of DNA TA as a function of the distance between bp72 and bp82,  $R_{72-82}$ . (C) Base pair opening probability of bp11, 25, and 72.  $e = 25$  kJ/mol.  $\sigma = 8.52$  Å.



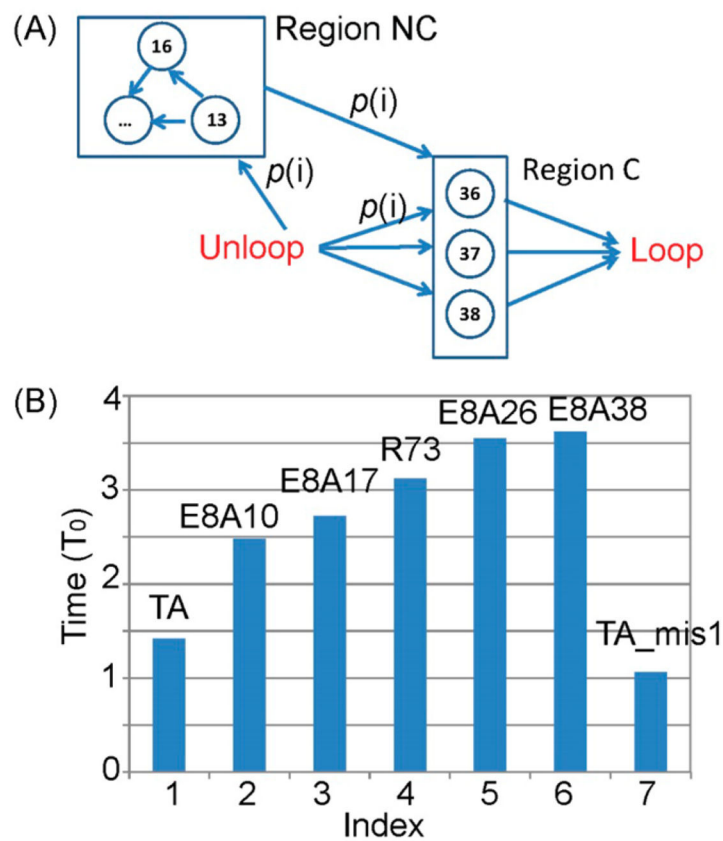
**Figure 2.** Percentage of unlooped structures as a function of time for four DNA sequences. Raw simulation data are shown in dashed lines, and solid lines fit results with a single exponential function. TA, TA\_mut, TA\_mis1, and TA\_mis2 are results with a biased potential  $0.003 \text{ e}/\sigma^2 = 0.001 \text{ kJ/mol}/\text{\AA}^2$ . TA' and TA\_mis1' are results with a biased potential  $0.0015 \text{ e}/\sigma^2$ .  $\tau = 200$  integration steps.



**Figure 3.** (A) Probability of each base pair of DNA TA to open as a function of time. (B) Distribution of the time for the appearance of structural defects  $T_{\text{defects}}$  (bp 24–26) divided by the time of looping  $T_{\text{loop}}$  for model sequence TA.



**Figure 4.** Probabilities of angles between DNA segments, with different numbers of mismatched base pairs (NM). The length of the DNA segment is 7bp. Black line: results using  $y = A - \ln(\sin \theta) + B\theta^2$ ; blue line: results using  $y = A - \ln(\sin \theta) + B\theta$ .



**Figure 5.** (A) Markov model to represent the looping process. (B) Prediction on the looping time for different DNA sequences.  $T_0$  is a time unit to match the CG time to real time. The error is  $<0.01 T_0$ .