Sequence Affects the Cyclization of DNA Minicircles: Supporting Information

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, TX 77555-0304

1. DNA Sequences

Four DNA sequences (TA, TA-mut, TA-mis1 and TA-mis2) simulated are shown below.

Sequence TA:

TTAGGATAACCGCCAATTTTGCGCCATCTGTCGCGCATGCACGCAAATTCGCCACGA
TCTCGATGCCTAAGAC

Sequence TA_mut (mutations are shown in red):

TTAGGATAACCGCCAATTTTGC**TATA**TCTGTCG**TATA**TGCACGCAAATTCGCCACGA
TCTCGATGCCTAAGAC

Sequence TA_mis1 (mutations are shown in red, notice that for this sequence the mutation is only on one strand):

TTAGGATAACCGCCAATTTTGCG**AA**ATCTGTCGCGCATGCACGCAAATTCGCCACGA
TCTCGATGCCTAAGAC

Sequence TA_mis2 (mutations are shown in red, notice that for this sequence the mutation is only on one strand):

TTAGGATAAC**AA**CCAATTTTGCGCCATCTGTCGCGCATGCACGCAAATTCGCCACGA
TCTCGATGCCTAAGAC

Six DNA sequences tested by the analytical model are shown below (from 3' to 5')

Sequence TA

TTAGGATAACCGCCAATTTTGCGCCATCTGTCGCGCATGCACGCAAATTCGCCACGA
TCTCGATGCCTAAGAC

Sequence E8A10

TGATTGAGCTCTATTGTGCCGGTTGTCTTTTTTTTTTTGTCGTGGCACCTCCGCTATTTA
TTTTTGCCTAAGAC

Sequence E8A17

TGATTGAGCTCTATTGTGCCGGTTTTTTTTTTTTTTTTTCGTGGCACCTCCGCTATTTA
TTTTTGCCTAAGAC

Sequence R73

ACTTTGCCACCTCCACTCCTATCAGTACCAGTTAATCTTCCCTCAGATATAACTCCAT
GATCGTGCCTAAGAC

Sequence E8A38

TGATTGAGCTCTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTATTTATT
TTTGCCTAAGAC

Sequence E8A26

TGATTGAGCTCTATTGTGCCTTTTTTTTTTTTTTTTTTTTTTTTTTTCCTCCGCTATTTATT
TTTGCCTAAGAC

2. Coarse-grained (CG) Models

The coarse-grained model used to represent DNA molecules is a modification of the Ox1 model and can be found in previous work [1-3]. The temperature was set to 310K and ionic strength was set to 0.5M. Newtonian dynamics was applied and an Andersen-like thermostat method [4] was used to maintain the temperature. We note that lacking Langevin-like forces from solvent the dynamics of this model is not strictly representative of real time but the time elapsed can be considered to scale with the number of integration steps.

3. Free Energy Measurement

The free energy of cyclization was calculated by umbrella sampling method. A harmonic potential was added between bp72 and bp82: $E = k(R_{72-82} - R_0)^2$ while $k = 1.0\varepsilon/\sigma^2$ (ε = 25 kJ/mol and σ = 8.52Å). $R_{72\text{-}82}$ is the distance between bp72 and 82. R_0 ranges from 3.5 to 30.0σ. There are a total of 62 windows. The reason to choose $R_{72\text{-}82}$ as the reaction coordinate is to mimic an experiment [5] which puts a fluorescence tag on bp72 and 82. Simulation at each reaction window was repeated 5 times from different random seeds for velocities. For each simulation 10^8 time steps were performed. The integration step = 0.005τ. $\tau \sim$ 3ps but since there is no direct mapping from a CG time scale to real time, we will just use τ as time-like unit in the text. The free energy profile was calculated using a WHAM procedure [6].

4. Cyclization Simulation

Cyclization simulations were initiated from fully extended structures. For each sequence, 200 trajectories were performed from different, random structures. Simulations were performed until at least 90% trajectories formed loops. The integration step was set to 0.005τ and data was recorded every 10^5 steps. The criteria of forming a loop is $R_{72\text{-}82} < 4\sigma$ and remaining in place for at least 5000τ. In our simulations, once the loop is formed, we found it will not fall apart again due to the strong interactions between two strands. Because the actual cyclization can take up to hours to finish, which is beyond the current computation power, two weak pulling potentials were added to help the cyclization process:

$$E_1 = k(R_{0-72} - R_1)^2$$

$$E_2 = k(R_{82-83} - R_2)^2$$

$R_{0\text{-}72}$ is the distance between bp0 and 72 and $R_{82\text{-}83}$ is the distance between bp82 and 83. k = 0.003 ε/σ^2. $R_1 = R_2 = 0.59\sigma \sim 5$Å, which approximately equals to the distance between two bps in a closed loop form. The pulling potentials reduced but did not completely eliminate the free energy barrier (see results below). In order to further prove that the pulling potential will not qualitatively change the main conclusion of this work, we also reduced the potential to half (k = 0.0015 ε/σ^2) and repeated the cyclization simulations for sequences TA and TA_mis1.

5. Angle Distribution for DNA molecules with and without Defects

In order to study the bending elasticity of a normal DNA and DNA with defects, we first designed a normal B-form DNA with bp=18 (CGCGCGCGCGCGCGCGCG). Next, in order to create defects on this DNA molecule, we made substitutions or mutations (C->A and G->A) on one strand and kept the original C/G bp in the other strand, leading to bp mismatches. During the simulation, those mismatched bps formed defects such as kinks or bubbles. The probability of defect formations was controlled by the total number of mismatched bps (NM). We tested NM = 0, 3, 4 and 5. For each sequence, the simulation lasted $5*10^5\tau$ and was repeated 16 times from different random seeds. The total number of the recorded configurations for each sequence was 160,000. For each configuration, the angle θ between two lines was calculated: one line points from the bp9 to bp2 and the other points from bp16 to bp9. This amounts to the angle between two continuous DNA segments of 7 bp length. This analysis was to match with that from a previous experimental study [7] dividing DNA into small segments of 2.5nm ~ 7bp.

Fig. 1. Illustration of the kinks and bubbles existing in the cyclization trajectory. Three kinds of structures were observed: (A) type I kink – no base pair opening but there is a large roll angle (> 45 degrees) for a base stacking; (B) type II kink – one base pair opening and (C) bubble – at least 3 continuous base pair openings. The percentage of the structures containing type I kink defects in a successful trajectory was found to be 13.8%, while that of the type II kink was 7.5% and that of bubble was 0.4%.

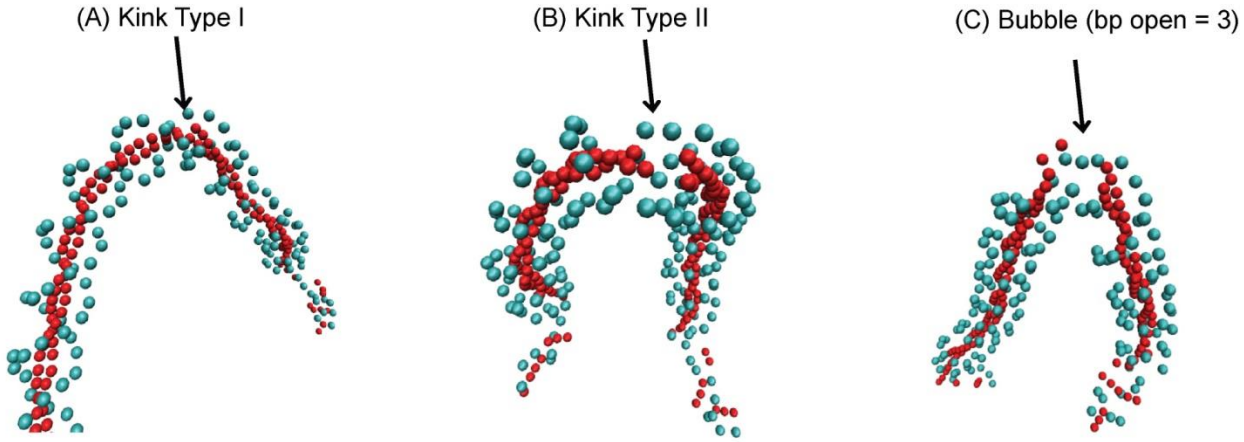


Fig. 2. Distribution of the time of appearing structural defects $T_{defects}$ divided by the time of looping T_{loop} for model sequence TA: (A) bp16-18; (B) bp35-37 and (C) bp24-26.

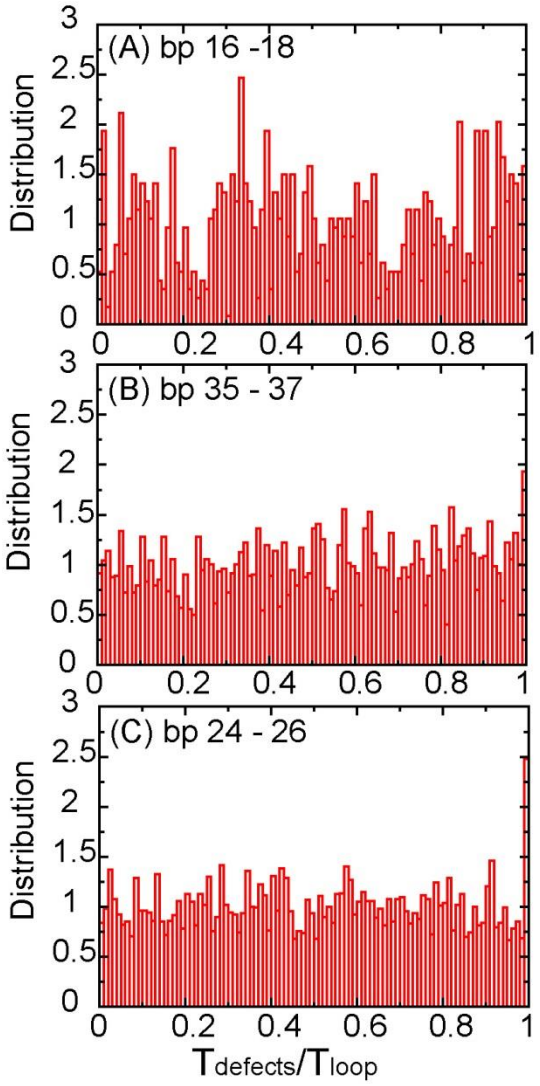
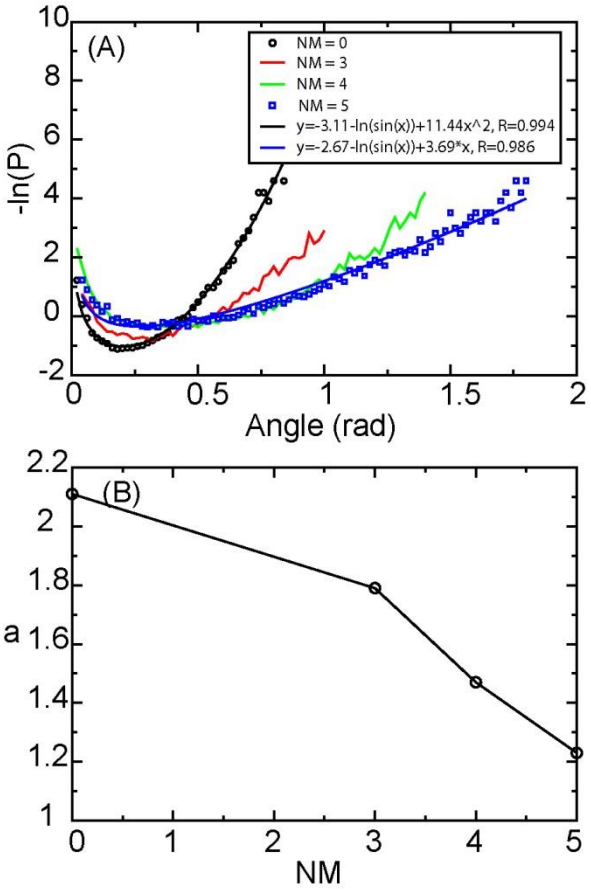


Fig. 3. (A) Probabilities of angles between DNA segments, with different numbers of mismatched base pairs (NM). The length of DNA segment is 7bp. Black line: fitting results using the function $y = A - \ln(\sin\theta) + B\theta^2$ *and blue line: fitting results using the function* $y = A - \ln(\sin\theta) + B\theta$ *. The unit of angle is rad. R is the correlation coefficient. (B) Fitting the raw data in (A) using* $y = A - \ln(\sin\theta) + B\theta^a$



[1] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, Journal of Chemical Physics **134**, 22 (2011).
[2] P. Sulc *et al.*, Journal of Chemical Physics **137**, 14 (2012).
[3] Q. Wang, and B. M. Pettitt, Biophys. J. **106**, 1182 (2014).
[4] H. C. Andersen, Journal of Chemical Physics **72**, 2384 (1980).
[5] R. Vafabakhsh, and T. Ha, Science **337**, 1097 (2012).
[6] A. Grossfield, 2013).
[7] P. A. Wiggins *et al.*, Nat. Nanotechnol. **1**, 137 (2006).