THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

DNA Looping Mediated by Site-Specific SfiI–DNA Interactions

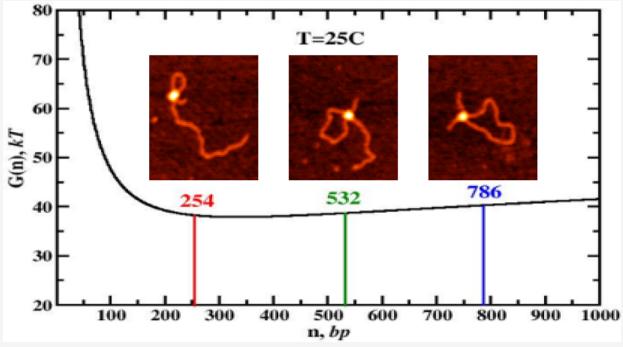
Sridhar Vemulapalli, Mohtadin Hashemi, Anatoly B. Kolomeisky,* and Yuri L. Lyubchenko*

Cite This: J. Phys. Chem. B 2021, 125, 4645–4653

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Interactions between distant DNA segments play important roles in various biological processes, such as DNA recombination. Certain restriction enzymes create DNA loops when two sites are held together and then cleave the DNA. DNA looping is important during DNA synapsis. Here we investigated the mechanisms of DNA looping by restriction enzyme SfiI by measuring the properties of the system at various temperatures. Different sized loop complexes, mediated by SfiI–DNA interactions, were visualized with AFM. The experimental results revealed that small loops are more favorable compared to other loop sizes at all temperatures. Our theoretical model found that entropic cost dominates at all conditions, which explains the preference for short loops. Furthermore, specific loop sizes were predicted as favorable from an energetic point of view. These predictions were tested by experiments with transiently assembled SfiI loops on a substrate with a single SfiI site.



INTRODUCTION

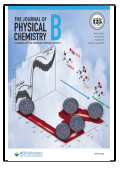
Protein–DNA interactions are critically important for supporting the majority of fundamental cellular processes, such as transcription, DNA repair, and DNA recombination.[1−3] Complex topological structures can occur when two spatially distant segments on DNA are brought together by specialized proteins, leading to a site-specific protein–DNA synaptic complex called a synaptosome.[4−8] The formation of the synaptosome is a general phenomenon during transcriptional regulation[9] (e.g., by Lac repressor), site-specific recombination,[7,10,11] and various gene rearrangement systems.[12−14] Synaptosome formation leads to DNA looping by bringing together spatially distant segments.[15−19] The size of the loops is determined by a variety of structural and chemical interactions between protein and DNA molecules. Several theoretical models have been proposed to describe the role of different factors, such as DNA length, protein–DNA interactions, DNA mechanics, and geometric factors, in the loop formation process.[20−24] However, the molecular mechanisms of underlying loop formation processes still remain not well understood. In addition, there are a limited number of experimental studies that might test these theoretical ideas.

Double-stranded DNA (dsDNA) can be characterized as a semiflexible polymer chain with a persistence length, l_p, of ∼150 bp under typical cellular conditions.[25−27] The persistence length of dsDNA is important for understanding the processes that involve the interaction of multiple sites on the same DNA, e.g., due to different proteins binding to specific sites on the DNA. The parameter l_p reflects the ability of the DNA chain to bend, and it depends on several factors, including ionic strength, DNA sequence, DNA defects, and temperature.[28−31] Local kinks in dsDNA lead to higher intrinsic bendability and affect the stability of loops formed by proteins.[28,32,33] However, the microscopic picture of the protein-mediated loop formation is still not fully understood.

SfiI is a type IIF restriction endonuclease that can associate with two distinct cognate sites. SfiI binds to DNA as a homotetramer and attaches to a recognition sequence of 5′-GGCCNNNN^NGGCC-3′, where N denotes any base and ^ marks the cleavage position before cleavage. To cleave the DNA chain, SfiI requires magnesium ions as cofactors; replacing Mg^{2+} with Ca^{2+} results in the formation of stable SfiI–DNA synaptic complexes.[5,34,35] The interaction between the two cognate sites on the same DNA molecule leads to the formation of looped DNA structures.[35−37] These properties make SfiI an ideal system to quantitatively characterize the loop formation process and to elucidate the microscopic factors that affect this process.

In this study, we employed atomic force microscopy (AFM) to directly visualize the SfiI–DNA synaptic complexes. The DNA used in this work contained three specific SfiI recognition sites, resulting in the formation of DNA loops with 254 bp, 532 bp, and 786 bp lengths. The assembly of the

Received: January 27, 2021
Revised: April 21, 2021
Published: April 29, 2021



https://doi.org/10.1021/acs.jpcb.1c00763
J. Phys. Chem. B 2021, 125, 4645–4653

looped complexes was monitored at 4, 25, and 50 °C. A simple theoretical model based on equilibrium polymer physics arguments was developed and applied to understand the experimental observations. The probability of the formation of DNA loops of different sizes is explained by considering free-energy changes during the formation of different protein−DNA complexes. The comparison with experiments reveals the importance of entropic factors in the protein-mediated DNA looping under all experimental conditions. Theoretical predictions for free-energy profiles of DNA loop formation were further tested by experiments with a single specific protein binding site making possible transient loop formation. Physical−chemical arguments to explain these observations are presented.

EXPERIMENTAL PROCEDURE

Materials. SfiI enzyme: SfiI restriction enzyme with low BSA (20 units/μL) was purchased from the New England Biolabs (Beverly, MA).

DNA Substrate. Two DNA constructs were used in this study and were obtained by the PCR of a pUC19 plasmid from Bio Basic (Markham, ON, CA), which included an 885 bp DNA sequence with three SfiI cognate sequences, GGCC-TCGAG-GGCC. The SfiI recognition sequence was previously proposed as the strongest sequence with high specificity.[34]

Three-SfiI-Site DNA Construct. PCR was performed to obtain a final construct of 1036 bp, with three SfiI sites and two flanks of 113 bp and 98 bp, as shown in Figure 1. As

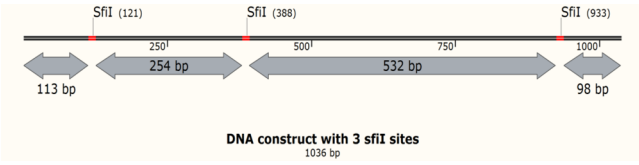


Figure 1. Schematic of the three-SfiI-site DNA construct. The SfiI sites are represented as red bars. The arm lengths and the intersite distance between the SfiI sites are given in base pairs.

shown in Figure 1, the SfiI sites on the DNA were separated by 254 bp between the first and second sites and 532 bp between the second and third sites; thus, there are 786 bp between the first and third sites. The PCR product was run on a 1% agarose gel, the product bands were excised, and DNA was extracted and purified using the Qiagen DNA gel extraction kit (Qiagen Inc., Valencia, CA). The final DNA concentration was determined by absorbance at 260 nm using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE). A restriction digest reaction using SfiI and the purified DNA was run and characterized on a gel to confirm the presence of recognition sites, shown in Figure S1. Sanger sequencing of the DNA was also performed to confirm the presence of three recognition sites, shown in Figure S2.

Single-SfiI-Site DNA Construct. PCR was performed to obtain a construct of 1036 bp with a single SfiI recognition site flanked by 925 bp and 98 bp arms, shown in Figure S3. The PCR product was purified and quantified as described above.

The PCR primers for the three-site construct were the forward primer, 5′-GGGGATGTGCTGCAAGG-3′, and the reverse primer, 5′-TGTGTGGAATTGTGAGCGG-3′. The primers for the single-site construct were the forward primer, 5′-ATTCTGAGAATAGTGTATGCGG-3′, and the reverse primer, 5′-CCAAGCACCAGAAGCC-3′. Primers were designed using SnapGene software (version 5.2, GSL Biotech, Chicago, IL) and ordered from IDT (Coralville, IA).

SfiI−DNA Synaptosome Assembly. The synaptosome assembly reaction mixture consisted of 1 μL of 10× buffer A [10 mM HEPES (pH 7.5), 50 mM NaCl, 2 mM CaCl2, 0.1 mM EDTA], 2 μL of DNA substrate (86 ng/μL), 1 μL of 1 mM DTT, 1 μL of SfiI enzyme, and 5 μL of DI water. The reaction mixture was prepared, incubated for 15 min at the desired temperature, and diluted through serial dilutions to obtain the final concentration used for AFM. The same procedure was followed for SfiI−DNA complex assembly at 4, 25, and 50 °C; reagents, except SfiI protein, were pre-equilibrated at the desired temperature before complex assembly and deposition.

SfiI−DNA Transient Loop Assembly. Transient loop assembly was performed at 25 °C with the reaction mixture consisting of 1 μL of 10× buffer A, 1 μL of 1 mM DTT, 1 μL of SfiI enzyme, and 2 μL of DI water. One μL of the reaction mixture was then mixed with 10 μL of 2 nM DNA and immediately (<5 s) used for AFM sample preparation.

Atomic Force Microscopy. A freshly cleaved mica was functionalized with a 167 μM solution of 1-(3-aminopropyl)-silatrane, as described previously.[38,39] Five μL of SfiI−DNA assembly reaction mixture sample was diluted using 1× buffer A and deposited on the functionalized mica surface, incubated for 2 min, rinsed with DI water, and dried with a gentle stream of argon. In the case of transient loop assembly, 5 μL of the reaction mixture was deposited on the functionalized mica surface, incubated for 2 min, rinsed with DI water, and dried with a gentle stream of argon. A typical image of 1 × 1 μm^2 with 512 pixels/line was obtained under ambient conditions with a MultiMode AFM system (Bruker) using TESPA probes (Bruker Nano, Camarillo, CA, USA).

Data Analysis. The contour lengths of the free DNA and the SfiI−DNA complexes were measured using the FemtoScan software (Advanced Technologies Center, Moscow, Russia). The contour length measurement of free DNA was performed by tracing a line starting from the one end of the DNA, continuing along the backbone to the other end (Figure S4). A total of N = 500 particles were analyzed, a histogram was assembled, and the distribution fit with a single Gaussian function. The peak was at 1037 ± 32 bp (standard deviation) (Figure S5). The conversion factor was obtained by dividing the mean length of DNA, in nanometers, with the total length in bp (1036 bp); the conversion factor was 0.335. A similar procedure was followed to measure the contour length of single-SfiI-site DNA and obtained the conversion factor (0.340) details provided (Figure S6).

To measure the loop sizes of the DNA−SfiI complexes, a line starting from the center of the SfiI tetramer (bright feature on the AFM images; Figure 2), along the length of the DNA loop, back to the center of the tetramer was drawn and analyzed, as shown in Figure S7. A total of N = 1000 SfiI−DNA synaptosome complexes were analyzed at each temperature. Statistical analysis was performed by assembling each data set in histograms (bin size of 20 bp) and fitting with a multiple peak Gaussian function. Data for each peak $\pm 2\sigma$ were tabulated and compared using a non-parametric Kolmogorov−Smirnov method[40] to determine the significance of change for each peak at different temperatures.

In the case of transient loop assembly with a single SfiI site construct, a total of N = 250 SfiI−DNA complexes forming loops (Figure 5A) were analyzed. Statistical analysis was

https://doi.org/10.1021/acs.jpcb.1c00763
J. Phys. Chem. B 2021, 125, 4645−4653

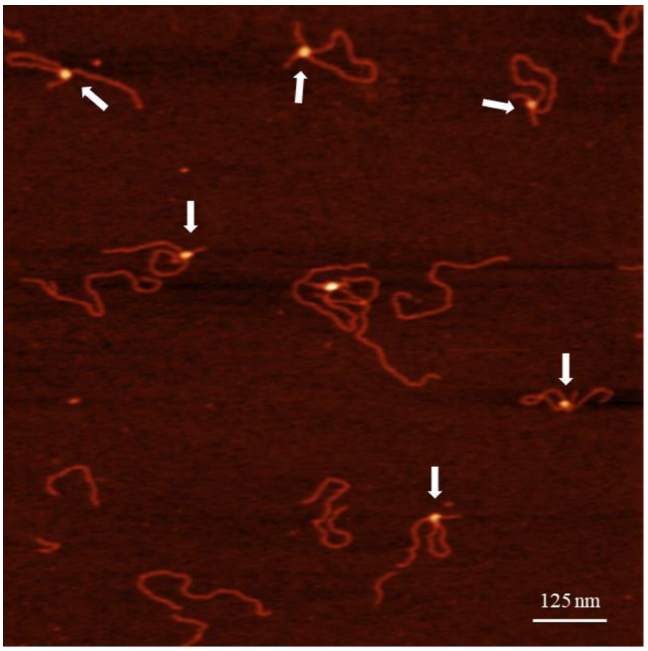


Figure 2. A typical 1 × 1 μm^2 AFM image of SfiI in complex with the three-site DNA construct at 50 °C. Complexes of different loop sizes are indicated with arrows; SfiI tetramers are seen as bright features at the intersection of DNA.

performed, and histograms, with a bin size of 20 bp, indicated the most probable loop size around 300−400 bp, as shown in Figure 5B. A total of N = 130 SfiI−DNA transient loop complexes were analyzed to measure the length of the short flank of the DNA to the center of the tetramer (bright feature in Figure S9) to confirm the position of SfiI; the histogram is shown in Figure 5C. A single-peak Gaussian function estimates the peak at 101 ± 9 bp, indicating the positioning of SfiI on the recognition site. All statistical analysis was performed using OriginPro (Origin Laboratories, Northampton, MA).

RESULTS

DNA with Three and Single SfiI Sites. The final product of PCR was purified and quantified, as described in the Experimental Procedure section. The purified DNA was diluted to 2 nM and deposited on APS-functionalized mica before being imaged under ambient conditions using AFM. The DNA contour lengths were then measured as shown in Figure S4; for DNA with three SfiI sites, shown in Figure 1, the contour length distribution (N = 500) produced a single peak at 1037 ± 32 bp (SD); the fit using a Gaussian function is shown in Figure S5. The single SfiI site construct produced a broad distribution approximated by a single peak curve (N = 255) at 1039 ± 29 bp (SD) shown in Figure S6. Both constructs were in good agreement with the designed length and theoretical contour length of 1036 bp.

SfiI−DNA Loop Assembly at Different Temperatures. SfiI−DNA complex assembly was performed in the presence of 2 mM $CaCl_2$, which assembles stable synaptosomes. The complex assembly, dilution, and sample deposition was performed at 4, 25, and 50 °C. Experiments at each temperature were repeated three times. The looped morphology of the DNA is clearly seen in Figure 2. As expected, three different sized DNA loops with SfiI at the DNA intersection were observed. Yields of looped complexes were 38, 58, and 49% for 4, 25, and 50 °C, respectively; the data are shown in Table S1. A total of 1000 particles, with a single tetramer at the DNA intersection, were analyzed for each temperature.

Loop size distributions for each temperature were assembled in histograms and fitted with multiple peak Gaussian functions, presented in Figure 3. Observed loop sizes at 4 °C, Figure 3A, produce three peaks at 261 ± 30 bp (SD), 526 ± 58 bp (SD), and 788 ± 48 bp, respectively. At 25 °C, Figure 3B, the loops are distributed in three peaks at 259 ± 29 bp (SD), 519 ± 44 bp (SD), and 775 ± 52 bp, respectively. At 50 °C, Figure 3C, the three peaks are at 265 ± 33 bp (SD), 539 ± 45 bp (SD), and 807 ± 62 bp, respectively. All loop distribution peaks, regardless of temperature, are close to the expected loop sizes of 254, 532, and 786 bp. Statistical comparison between the different loops at each temperature was performed, and the results are presented in Table S2. Comparison reveals that the distributions of all loop sizes at all temperatures are significantly different, at a minimum of 95% confidence interval and in many cases at 99.9%.

Probability distributions of loops at 4, 25, and 50 °C are shown in Figure S8. The areas under the curve (AUC), corresponding to relative probability for short, medium, and long loop size, at each temperature, are as follows: 0.42, 0.36,

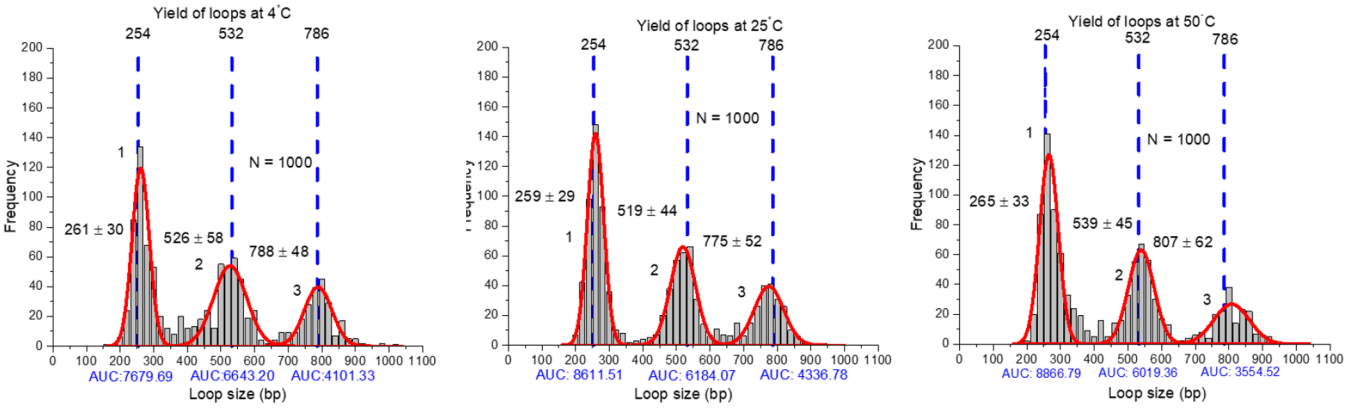


Figure 3. Analysis of SfiI−DNA loop complexes formed at 4, 25, and 50 °C. Multiple peak Gaussian fits of the histograms, with a bin size of 20 bp, produce three peaks at 261 ± 30 bp (SD), 526 ± 58 bp (SD), and 788 ± 48 bp at 4 °C (A); 259 ± 29 bp (SD), 519 ± 44 bp (SD), and 775 ± 52 bp at 25 °C (B); and 265 ± 33 bp (SD), 539 ± 45 bp (SD), and 807 ± 62 bp at 50 °C (C). The blue dotted lines indicate the expected loop size for each loop. The area under the curve (AUC) for each loop is shown under the respective peak.

https://doi.org/10.1021/acs.jpcb.1c00763
J. Phys. Chem. B 2021, 125, 4645−4653

Table 1. Fractions of Looped DNA Conformations from Experimental Measurements[a]

loop size	4 °C	25 °C	50 °C
short	0.42 ± 0.005 (S.E.)	0.44 ± 0.012 (S.E.)	0.48 ± 0.02 (S.E.)
medium	0.36 ± 0.006 (S.E.)	0.34 ± 0.012 (S.E.)	0.33 ± 0.02 (S.E.)
long	0.22 ± 0.012 (S.E.)	0.22 ± 0.012 (S.E.)	0.19 ± 0.00 (S.E.)
persistence length, l_p, in bp[b]	160	144	108

[a]Probabilities are given ± standard error. [b]Experimentally measured DNA persistence lengths obtained from ref 29.

0.22 (4 °C); 0.44, 0.34, 0.22 (25 °C); and 0.48, 0.33, 0.19 (50 °C). These data are shown in Table 1. We assume that these areas are proportional to the relative fractions for the occurrences of the loops. Comparing the relative fractions of different loop sizes shows that the smallest loop, 254 bp, is more probable than other possible loop sizes. The probability of formation of the loops follows the trend 254 bp > 532 bp > 786 bp, across all three temperatures. Furthermore, the relative fraction of 254 bp loops is more probable with increased temperature. These experimental results are further discussed and explained by our theoretical analysis.

Theoretical Analysis of SfiI–DNA Looping and Comparison with the Experiment. Our theoretical approach is based on the application of simple polymer-physics arguments to quantitatively describe the distribution of different loop sizes, obtained at different temperatures. The main assumption is that the system can be described using equilibrium thermodynamics. In other words, it is assumed that experimental measurements have been done for time intervals long enough so that the observed loop sizes reflect the underlying equilibrated free-energy landscapes. Then, the free energy of formation of the DNA loop of size n (in units of kT) due to SfiI protein forming the complex can be approximated as[41]

$$G(n) = \frac{A}{n} + b\ln(n) + E \tag{1}$$

where the first term accounts for the polymer bending energy, the second term describes the entropic cost of loop formation, and the last term is enthalpic due to the chemical interactions between SfiI and DNA at the intersection. The coefficient A is proportional to the bending stiffness, and it is given in terms of the persistence length l_p as

$$A = 2\pi^2 l_p \tag{2}$$

The parameter b is related to the scaling exponent of the radius of gyration; for the ideal polymer chain, it is known that $b = 3/2$. Since dsDNA is not an ideal polymer, larger values of b are expected.

Let us also assume that the enthalpic contributions E are the same for different loop sizes and that they are weakly dependent on the temperature (in the range of experimental measurements from 4 °C until 50 °C). This is a reasonable assumption, since the same chemical interactions are present for all different loop sizes when the SfiI protein associates to DNA at both specific sites. Further support for this assumption comes from the fact that SfiI is a thermophile enzyme and has been shown to cleave in a temperature-independent manner.[42−44] However, the stability of the enzyme is also temperature-dependent and it might influence the formation of the loops. Additionally, the turnover rate does vary with temperature,[42−44] which can contribute to the yield values shown in Table S2.

Now, the only parameters that depend on the temperature are the persistence length l_p and the entropic contribution parameter b. The temperature dependence of the DNA persistence length has recently been accurately measured.[29] In agreement with theoretical expectations, it was found experimentally that increasing the temperature lowers the persistence length.[29] In other words, it is easier to bend the DNA molecule at higher temperatures.

Our theoretical analysis is based on the idea that the relative fractions of different loop sizes, $f(n)$, are proportional to the probabilities of formation of the loops, which are determined by their free energies (more precisely, Boltzmann factors). The lower the free energy of a given conformation, the more probable it is to observe this conformation in experiments. More specifically, the probability of finding the system in the looped state of size n can be estimated from

$$P(n) \sim e^{-G(n)} = \frac{\exp\left(-\frac{2\pi^2 l_p}{n}\right)}{n^b} \tag{3}$$

Table 1 presents the experimental relative fractions of different loop sizes at different temperatures and the temperature-dependent DNA persistent lengths from ref 29. It is expected that the differences in these fractions reflect the differences in the probability of formation of DNA loops. This allows us to employ these data in eqs 1, 2, and 3 in order to quantify what factors are governing the formation of the protein-mediated DNA loops. More specifically, the ratio of relative fractions for any two DNA loop sizes can be associated with the underlying free-energy landscape via

$$\frac{f(n_1)}{f(n_2)} = \frac{P(n_1)}{P(n_2)} = \left(\frac{n_2}{n_1}\right)^b \exp\left[2\pi^2 l_p\left(\frac{1}{n_2} - \frac{1}{n_1}\right)\right] \tag{4}$$

We can now analyze the data using eqs 3 and 4 for every temperature, assuming that the parameter $b(T)$ is unknown and can be used as a fitting parameter. It is shown then that $b \approx 9.0$ at 4 °C, $b \approx 8.5$ for 25 °C (room temperature), and $b \approx 6.3$ for 50 °C. One can see that the entropic contribution starts to decrease with an increase in the temperature. This agrees with expectations that increasing T should bring DNA molecules closer to being ideal polymer chains because they are getting more flexible. Higher temperatures weaken hydrogen and other secondary structure bonds that hold the DNA molecule together. This should also stimulate the appearance of more DNA defects, and it will make DNA more bendable. However, the amplitudes for the parameter b, from the fitting procedures, are higher than expected for nonideal polymer chains ($1.5 < b < 1.8$).

Now let us estimate the free energies of the looped configurations at different temperatures using the expression from eq 1. The results are presented in Table 2. Note that to

https://doi.org/10.1021/acs.jpcb.1c00763
J. Phys. Chem. B 2021, 125, 4645−4653

understand the distribution of loop sizes only the relative values of the free energies are needed.

Table 2. Estimated Free Energies (in *kT* Units) from eq 1

	4 °C	4 °C	25 °C	25 °C	50 °C	50 °C
loop size	bending	entropic	bending	entropic	bending	entropic
short	12.4	49.8	11.2	47.1	8.4	34.6
medium	5.9	56.5	5.3	53.4	4.0	39.2
long	4.0	60.0	3.6	56.7	2.7	41.7

One can see that the entropic contributions dominate the formation of DNA loops at all temperatures. In our theoretical analysis, they are always significantly larger than the bending contributions. The entropic term domination is stronger for longer loops, while for shorter loops the bending contributions start to catch up. For a fixed temperature, increasing the DNA loop size lowers the bending energy but it also increases the entropic term. This result is expected because to create a larger DNA loop requires less bending. An increase in the entropic free-energy term, however, is observed for longer loops because of the stronger decrease in the allowed degrees of freedom for longer loops in comparison with the free polymer. At the same time, for the fixed loop size, the increase in the temperature simultaneously lowers both the bending and the entropic contributions to the overall free energies of the system. For the bending term, it is clearly associated with the decrease in the persistence length, while, for the entropic term, it is related with the fact that the entropy decrease due to looping is smaller at higher temperatures.

Assuming that our approximate description of the free-energy cost of looping (eq 1) provides a reasonable description of the process, using the fitted values of the parameter b we can construct the overall free-energy profiles for the looped DNA configurations. The results are presented in Figure 4 for three different temperatures. We predict that, while the short loops (300−350 bp) are the most probable, the larger loops can still be observed (although with a smaller probability) due to a relatively slow increase in the entropic contribution as a function of the size [$\sim\ln(n)$]. Note also that very short loops (<100 bp) have a much lower probability to be observed because of the strong increase in the bending energy for small n.

Transient SfiI−DNA Looping on Single-SfiI-Site DNA. One of the specific conclusions of our theory is that loops with sizes of 300−350 bp are the most probable species in the looped assemblies under given experimental conditions. To test this theoretical prediction, we investigated the formation of transient loops formed by SfiI bound to a specific and nonspecific site. In these experiments, DNA of the same length but containing a single SfiI recognition site was designed. Our idea is that the loop sizes are determined by bending and entropic contributions to the free energy and they are not dependent on specific enthalpic interactions. This means that the distribution of transiently formed loops with nonspecific protein interactions also should follow the free-energy profiles predicted by our theoretical arguments.

The SfiI site was located at 98 bp from one end of the DNA and 925 bp from the other (Figure S3). SfiI binds this recognition site and probes other parts of the DNA in the site search process, forming transient loops that are unstable due to a low affinity of SfiI to nonspecific sites. To detect such transient states of SfiI stabilized loops, DNA was mixed with SfiI for a short period of time (5 s) and rapidly prepared for the AFM imaging. This procedure allowed us to minimize the formation of stable bimolecular trans complexes held together by SfiI bound to specific sites on two DNA molecules. A representative image is shown in Figure 5A in which the looped structure is indicated with a wide arrow. Linear DNA molecules with SfiI bound are indicated with "1", and the trans bimolecular complex is indicated with "2". A total of 250 looped SfiI−DNA complexes with a single tetramer at the DNA intersection (highlighted in Figure 5A) were analyzed. More examples of the transient loops are shown in Figure S9.

The length distribution for the short flank of the looped structures is narrow with the maximum 101 ± 9 (SD) bp shown in Figure 5C, which coincides with the position of the SfiI binding site, so transient loops are formed between the protein bound to a specific site and another nonspecific DNA sequence. The histogram of the size distribution of the loops is presented in Figure 5B. The distribution is asymmetric with most of the transient loops being between 300 and 400 bp, which is very close to the predicted range of the most favorable loop size of 300−350 (Figure 4). The asymmetry of the histogram with sharp decay toward short loops is also consistent with our theoretical predictions (Figure 4).

DISCUSSION

The direct measurement of DNA loops formed by SfiI restriction enzyme allowed us to characterize the effect of size and temperature on the assembly of SfiI−DNA loops. This provided the information to clarify microscopic features of protein-mediated DNA loop formation. The specific details are outlined in the following paragraphs.

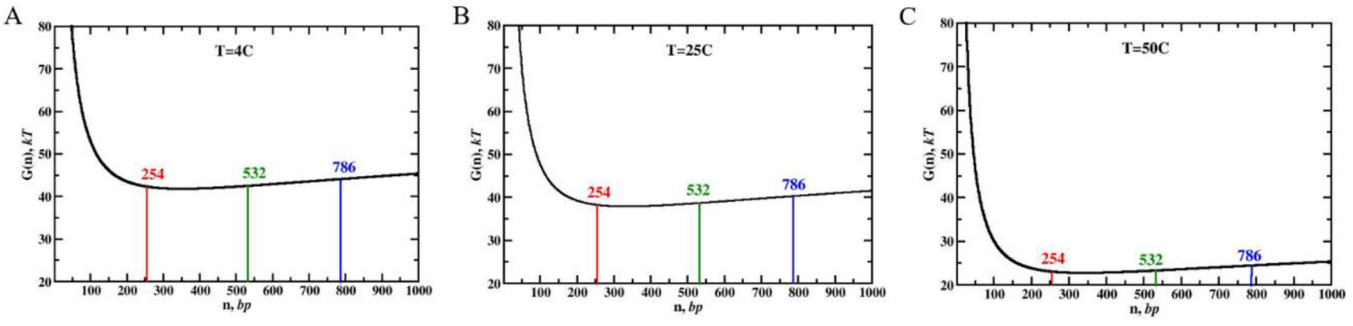


Figure 4. Graphs for the free-energy cost of looping $G(n)$ as a function of n for three different temperatures from eq 1: at 4 °C (A), 25 °C (B), and 50 °C (C). For all calculations, $E = -20$ is assumed.

https://doi.org/10.1021/acs.jpcb.1c00763
J. Phys. Chem. B 2021, 125, 4645−4653

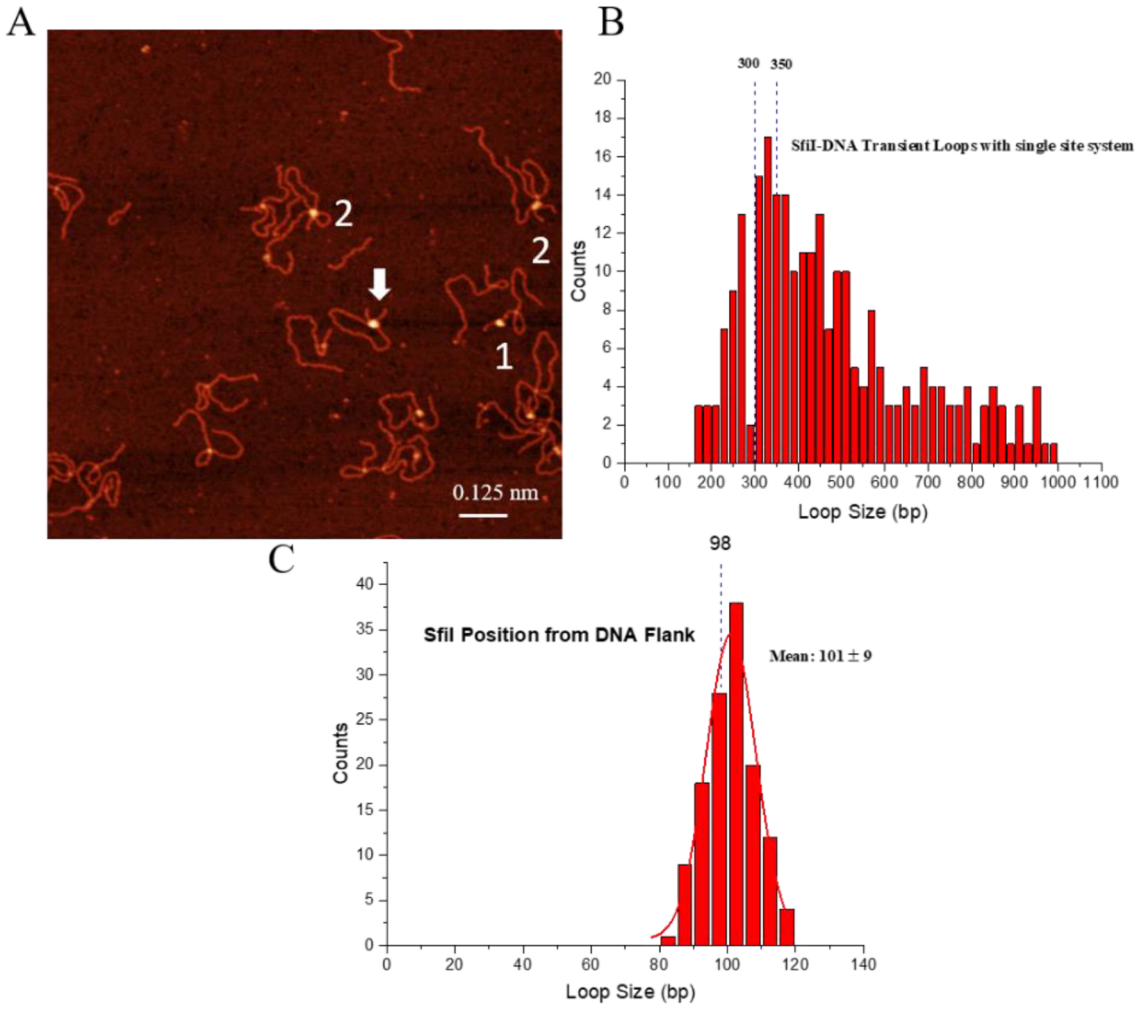


Figure 5. Transient loops for SfiI–DNA complexes at 25 °C. (A) A typical 1 × 1 μm^2 AFM image of transient SfiI–DNA loops assembled using DNA with a single SfiI site. A looped structure is indicated with a wide arrow. Linear DNA molecules with SfiI bound are indicated with "1", and the trans bimolecular complex is indicated with "2". (B) Distribution of loop sizes observed for N = 250 transient loops at 25 °C. The blue dotted lines indicate the region of the predominant loops predicted by the theory. (C) The single peak Gaussian fit of histograms gives a peak at 101 ± 9 bp (SD), indicating the positioning of SfiI is close to 98 bp (indicated by the dotted line), the expected position. Statistical analysis was performed using OriginPro (Origin Laboratories, MA).

Loops of 254 bp Are Predominant at All Temperatures. Our experimental results suggest that the smallest loop (254 bp) is the most probable outcome in the SfiI-mediated DNA loop formation for all temperatures. At first, this is a surprising observation, since this loop size is only slightly larger than the persistence length (l_p = 160 bp at 4 °C) and bending contributions to the free energy are expected to be significant. However, the entropic cost of making short loops is not as large as that for the longer loops. Our theoretical analysis suggests that entropic factors are the biggest contributors in the free energy and they effectively determine the probability of the formation of loops of different sizes. The smallest entropic changes are predicted for short loops, and this fully agrees with our experimental observations. Increasing the DNA loop size, at a given temperature, lowers the bending contribution to the free energy,[45,46] as expected from eq 1, but this effect is fully compensated by the increase in the entropic cost of the loop formation (see Table 2). As a result, our theoretical model predicts that short loops should be observed most frequently under all conditions, and this agrees with experimental observations.

Our data suggest that SfiI prefers the formation of relatively short DNA loops. However, making very short loops (comparable or smaller than the persistence length) will not be a favorable process because the bending energy will be too high. This is illustrated in Figure 4. We predict that there is a specific size of the loop that will be the most favorable from an energetic point of view. For all temperatures, the loops with sizes of 300–350 bp are expected to be the dominant population (see Figure 4). Note also that, although the short loops are preferred, we could simultaneously observe the loops of much larger sizes, although with a smaller frequency. One could speculate that the parameters of this system are optimized in the following way so that the range of possible loop sizes can be reached, and this allows the system to be more flexible. Interestingly, the theoretical prediction of the specific size of loops favorable from an energetic point of view matches with our experimental results with a single SfiI recognition site construct, as shown in Figure 5B. This gives an additional support to our approximate description of the free-energy cost for the formation of DNA loops.

https://doi.org/10.1021/acs.jpcb.1c00763
J. Phys. Chem. B 2021, 125, 4645–4653

Temperature Dependence of the DNA Loop Formation. As shown in Figure 2, three different loop sizes (254 bp, 532 bp, and 784 bp) can be simultaneously observed in our experimental system. However, the frequency of finding the loops of different sizes is not the same because of the different free energies of the protein−DNA looped conformations. As discussed above, the short loops are preferred in the system due to the high entropic cost of making loops of larger sizes. Table 1 quantifies this effect by presenting the relative fractions of different-size loops at various temperatures. At 4 °C, the short loop is the most probable (f = 0.42), and increasing the temperature makes the short loop even more preferred (f = 0.48 at 50 °C). Interestingly, the relative fractions of medium size (532 bp) and long size (784 bp) loops show a nonmonotonic behavior as a function of the temperature. At 4 °C, the relative fraction of the loops with the medium size is f = 0.36, and it drops to f = 0.34 at 25 °C; however, increasing the temperature further decreases the fraction of the medium loops to f = 0.33. The trend is also observed for long loops: starting from f = 0.22 at 4 °C, it remains at f = 0.22 at 25 °C before lowering to f = 0.19 at 50 °C. This is the consequence of the free-energy changes for each loop size when the temperature increases. Lowering of the bending energy for larger loop sizes is not equally compensated by the increase in the entropic cost (see Table 2).

Table 2 presents our theoretical estimates (from eq 1) for the free-energy contributions of the loop formation. One can see that entropic terms are always larger than the bending contributions. For short loops, the entropic term is 4.0−4.1 times larger than the bending energy, while for medium loops this ratio increases to 9.6−9.8 times and becomes even larger for long loops, increasing to 15.0−15.4 times the bending energy. These results clearly show that the entropic cost of creating DNA loops dominates the overall process and determines which loop sizes will be observed in the system. One could also notice that both bending and entropic terms decrease with an increase in temperature, with the effect being slightly stronger for the bending than for the entropic contribution.

Nature of Entropic Contributions. Although the semiphenomenological expression given in eq 1 seems to describe the experimental observations reasonably well, it is important to note that it relies on the simplified assumption that only two types of energies contribute to the free-energy landscape, bending and entropic factors. While the bending energies were estimated from known results about the DNA persistence length, in assigning the energy of the entropic term, the parameter b was fitted. However, the fact that this parameter is of the order 6−9, instead of the expected 1.5−1.8 from polymer physics arguments, suggests a more complex picture of protein−DNA interactions in the system that are not fully accounted for by eq 1. It seems that there are additional stabilizing interactions (beyond the simple loop entropy arguments) when the protein brings together two distinct DNA segments, which are also probably dependent on the loop size. It could be that when the short loops are formed the enthalpic interactions between the protein and DNA are stronger, while for longer loops it might be weaker. Several factors might contribute to this, including steric and electrostatic effects. It is expected that atomic details of the protein molecule forming a loop while associated with two different segments of DNA, which are not yet available, will help to understand the microscopic picture for these additional interactions.

CONCLUSIONS

Interactions between SfiI proteins and DNA are investigated using a combination of experimental and theoretical methods. It was shown that DNA loops of different sizes can be simultaneously observed at several temperatures, although with different probabilities. To quantify the differences in the relative fractions of loop sizes, a polymer-physics-based simple theoretical model was developed. We found that short loops are preferred under all experimental conditions, and these observations are explained by the domination of entropic factors during the loop formation. Increasing the temperature makes short loops even more probable, while the relative contributions of medium and long loops decrease with the temperature. It is argued that making the long loops leads to a larger decrease in the degrees of freedom for the DNA molecules, and this explains the dominating contributions of the entropic factors in the protein-mediated loop formation. We also speculate that the parameters of the system are optimized in such a way so that a large range of loop sizes can be created, allowing the protein to efficiently function under cellular conditions; this assumption was verified using experiments with transiently formed loops assembled by a single-recognition-site construct. At the same time, our analysis suggests that there are additional stabilizing protein−DNA interactions which might depend on the loop size.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c00763.

A detailed description of the restriction digest reaction with SfiI; Sanger sequencing data of the three-SfiI-site construct; schematic of the single-recognition-site construct; description of contour length measurement; contour length histogram for three-site DNA; contour length histogram for single-site DNA; loop size measurement description; transient SfiI−DNA looping on the single-site construct along with a histogram of flanking arm length measurements to determine the SfiI position (PDF)

AUTHOR INFORMATION

Corresponding Authors

Anatoly B. Kolomeisky − *Department of Chemistry-MS60, Rice University, Houston, Texas 77005-1892, United States;* orcid.org/0000-0001-5677-6690; Phone: 713-348-5672; Email: tolya@rice.edu

Yuri L. Lyubchenko − *Department of Pharmaceutical Sciences, College of Pharmacy, University of Nebraska Medical Center, Omaha, Nebraska 68198-6025, United States;* orcid.org/0000-0001-9721-8302; Phone: 402-559-1971; Email: ylyubchenko@unmc.edu

Authors

Sridhar Vemulapalli − *Department of Pharmaceutical Sciences, College of Pharmacy, University of Nebraska Medical Center, Omaha, Nebraska 68198-6025, United States*

https://doi.org/10.1021/acs.jpcb.1c00763
J. Phys. Chem. B 2021, 125, 4645−4653

Mohtadin Hashemi − *Department of Pharmaceutical Sciences, College of Pharmacy, University of Nebraska Medical Center, Omaha, Nebraska 68198-6025, United States;* orcid.org/0000-0003-2698-9761

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c00763

Author Contributions

Y.L.L. and A.B.K. designed the project. S.V. and M.H. performed the AFM experiments and data analysis. A.B.K. provided the theoretical model. All authors contributed to writing the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by NSF grant MCB- 1941049/1941106 to Y.L.L. A.B.K. acknowledges the support from the Welch Foundation (C-1559), from the NSF (CHE-1953453 and MCB-1941106), and from the Center for Theoretical Biological Physics sponsored by the NSF (PHY-2019745). We thank L. Shlyakhtenko (University of Nebraska Medical Center) for useful insights and all of the Y.L.L lab members for their fruitful discussions. The authors thank Tommy D. Stormberg (University of Nebraska Medical Center) for proofreading the manuscript. The UNMC Genomics Core Facility receives partial support from the NIGMS INBRE - P20GM103427-19, as well as the NCI center grant for Fred & Pamela Buffett Cancer Center - P30CA036727.

■ REFERENCES

(1) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular biology of the cell: Reference edition*; Garland Science: 2007; Vol. *5*.

(2) Kumar, D. T.; Mendonca, E.; Christy, J. P.; Doss, C. G. P.; Zayed, H. A computational model to predict the structural and functional consequences of missense mutations in O6-methylguanine DNA methyltransferase. In *Advances in Protein Chemistry and Structural Biology*; Elsevier: 2019; Vol. *115*, pp 351−369.

(3) Lodish, H.; Berk, A.; Kaiser, C. A.; Kaiser, C.; Krieger, M.; Scott, M. P.; Bretscher, A.; Ploegh, H.; Matsudaira, P. *Molecular cell biology*; Macmillan: 2008.

(4) Vetcher, A. A.; Lushnikov, A. Y.; Navarra-Madsen, J.; Scharein, R. G.; Lyubchenko, Y. L.; Darcy, I. K.; Levene, S. D. DNA Topology and Geometry in Flp and Cre Recombination. *J. Mol. Biol.* **2006**, *357* (4), 1089−1104.

(5) Lushnikov, A. Y.; Potaman, V. N.; Oussatcheva, E. A.; Sinden, R. R.; Lyubchenko, Y. L. DNA Strand Arrangement within the SfiI−DNA Complex: Atomic Force Microscopy Analysis. *Biochemistry* **2006**, *45* (1), 152−158.

(6) Shlyakhtenko, L. S.; Gilmore, J.; Portillo, A.; Tamulaitis, G.; Siksnys, V.; Lyubchenko, Y. L. Direct visualization of the EcoRII−DNA triple synaptic complex by atomic force microscopy. *Biochemistry* **2007**, *46* (39), 11128−11136.

(7) Grindley, N. D. F.; Whiteson, K. L.; Rice, P. A. Mechanisms of Site-Specific Recombination. *Annu. Rev. Biochem.* **2006**, *75* (1), 567−605.

(8) Brissett, N. C.; Pitcher, R. S.; Juarez, R.; Picher, A. J.; Green, A. J.; Dafforn, T. R.; Fox, G. C.; Blanco, L.; Doherty, A. J. Structure of a NHEJ polymerase-mediated DNA synaptic complex. *Science* **2007**, *318* (5849), 456−459.

(9) Hoverter, N. P.; Waterman, M. L. A Wnt-fall for gene regulation: repression. *Sci. Signaling* **2008**, *1* (39), pe43−pe43.

(10) Mouw, K. W.; Rowland, S.-J.; Gajjar, M. M.; Boocock, M. R.; Stark, W. M.; Rice, P. A. Architecture of a serine recombinase-DNA regulatory complex. *Mol. Cell* **2008**, *30* (2), 145−155.

(11) Swanson, P. C. The bounty of RAGs: recombination signal complexes and reaction outcomes. *Immunol. Rev.* **2004**, *200* (1), 90−114.

(12) Moens, P. B.; Marcon, E.; Shore, J. S.; Kochakpour, N.; Spyropoulos, B. Initiation and resolution of interhomolog connections: crossover and non-crossover sites along mouse synaptonemal complexes. *J. Cell Sci.* **2007**, *120* (6), 1017−1027.

(13) Vaezeslami, S.; Sterling, R.; Reznikoff, W. S. Site-directed mutagenesis studies of tn5 transposase residues involved in synaptic complex formation. *J. Bacteriol.* **2007**, *189* (20), 7436−7441.

(14) Weterings, E.; Chen, D. J. The endless tale of non-homologous end-joining. *Cell Res.* **2008**, *18* (1), 114−124.

(15) Kuznetsov, S. V.; Sugimura, S.; Vivas, P.; Crothers, D. M.; Ansari, A. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (49), 18515−18520.

(16) Ross, E. D.; Hardwidge, P. R.; Maher, L. J. HMG proteins and DNA flexibility in transcription activation. *Mol. Cell. Biol.* **2001**, *21* (19), 6598−6605.

(17) Watson, M. A.; Gowers, D. M.; Halford, S. E. Alternative geometries of DNA looping: an analysis using the SfiI endonuclease. *J. Mol. Biol.* **2000**, *298* (3), 461−475.

(18) Halford, S. Hopping, jumping and looping by restriction enzymes. *Biochem. Soc. Trans.* **2001**, *29* (4), 363−373.

(19) Halford, S. E.; Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **2004**, *32* (10), 3040−3052.

(20) Shvets, A. A.; Kolomeisky, A. B. The role of DNA looping in the search for specific targets on DNA by multisite proteins. *J. Phys. Chem. Lett.* **2016**, *7* (24), 5022−5027.

(21) Cournac, A.; Plumbridge, J. DNA Looping in Prokaryotes: Experimental and Theoretical Approaches. *J. Bacteriol.* **2013**, *195* (6), 1109.

(22) Kolomeisky, A. B. Protein-Assisted DNA Looping: A Delicate Balance among Interactions, Mechanics, and Entropy. *Biophys. J.* **2015**, *109* (3), 459−460.

(23) Matthews, K. S. DNA looping. *Microbiol. Rev.* **1992**, *56* (1), 123−136.

(24) Priest, D. G.; Kumar, S.; Yan, Y.; Dunlap, D. D.; Dodd, I. B.; Shearwin, K. E. Quantitation of interactions between two DNA loops demonstrates loop domain insulation in cells. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (42), E4449.

(25) Peters, J. P., 3rd; Maher, L. J. DNA curvature and flexibility in vitro and in vivo. *Q. Rev. Biophys.* **2010**, *43* (1), 23−63.

(26) Privalov, P. L.; Dragan, A. I.; Crane-Robinson, C. The cost of DNA bending. *Trends Biochem. Sci.* **2009**, *34* (9), 464−470.

(27) Rubinstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press: New York: 2003; Vol. *23*.

(28) Driessen, R. P. C.; Sitters, G.; Laurens, N.; Moolenaar, G. F.; Wuite, G. J. L.; Goosen, N.; Dame, R. T. Effect of temperature on the intrinsic flexibility of DNA and its interaction with architectural proteins. *Biochemistry* **2014**, *53* (41), 6430−6438.

(29) Geggier, S.; Kotlyar, A.; Vologodskii, A. Temperature dependence of DNA persistence length. *Nucleic Acids Res.* **2011**, *39* (4), 1419−1426.

(30) Geggier, S.; Vologodskii, A. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (35), 15421−15426.

(31) Hagerman, P. J. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17* (1), 265−286.

(32) Vologodskii, A.; Frank-Kamenetskii, M. D. Strong bending of the DNA double helix. *Nucleic Acids Res.* **2013**, *41* (14), 6785−6792.

(33) Wiggins, P. A.; Phillips, R.; Nelson, P. C. Exact theory of kinkable elastic polymers. *Phys. Rev. E* **2005**, *71* (2), 021909.

(34) Krasnoslobodtsev, A. V.; Shlyakhtenko, L. S.; Lyubchenko, Y. L. Probing interactions within the synaptic DNA-SfiI complex by AFM force spectroscopy. *J. Mol. Biol.* **2007**, *365* (5), 1407.
(35) Williams, S. A.; Halford, S. E. Sfi I endonuclease activity is strongly influenced by the non-specific sequence in the middle of its recognition site. *Nucleic Acids Res.* **2001**, *29* (7), 1476−1483.
(36) Suzuki, Y.; Gilmore, J. L.; Yoshimura, S. H.; Henderson, R. M.; Lyubchenko, Y. L.; Takeyasu, K. Visual analysis of concerted cleavage by type IIF restriction enzyme SfiI in subsecond time region. *Biophys. J.* **2011**, *101* (12), 2992−2998.
(37) Wentzell, L. M.; Halford, S. E. DNA looping by the SfiI restriction endonuclease. *J. Mol. Biol.* **1998**, *281* (3), 433−444.
(38) Lyubchenko, Y. L.; Shlyakhtenko, L. S. AFM for analysis of structure and dynamics of DNA and protein-DNA complexes. *Methods* **2009**, *47* (3), 206−213.
(39) Shlyakhtenko, L. S.; Gall, A. A.; Filonov, A.; Cerovac, Z.; Lushnikov, A.; Lyubchenko, Y. L. Silatrane-based surface chemistry for immobilization of DNA, protein-DNA complexes and other biological materials. *Ultramicroscopy* **2003**, *97* (1−4), 279−287.
(40) Kolmogorov−Smirnov Test. *The Concise Encyclopedia of Statistics*; Springer: New York, 2008; pp 283−287.
(41) Shin, J.; Kolomeisky, A. B. Facilitation of DNA loop formation by protein−DNA non-specific interactions. *Soft Matter* **2019**, *15* (26), 5255−5263.
(42) Embleton, M. L.; Vologodskii, A. V.; Halford, S. E. Dynamics of DNA loop capture by the SfiI restriction endonuclease on supercoiled and relaxed DNA. *J. Mol. Biol.* **2004**, *339* (1), 53−66.
(43) Nobbs, T. J.; Szczelkun, M. D.; Wentzell, L. M.; Halford, S. E. DNA excision by the Sfi I restriction endonuclease. *J. Mol. Biol.* **1998**, *281* (3), 419−32.
(44) Qiang, B. Q.; Schildkraut, I. A type II restriction endonuclease with an eight nucleotide specificity from Streptomyces fimbriatus. *Nucleic Acids Res.* **1984**, *12* (11), 4507−16.
(45) Mulligan, P. J.; Chen, Y.-J.; Phillips, R.; Spakowitz, A. J. Interplay of protein binding interactions, DNA mechanics, and entropy in DNA looping kinetics. *Biophys. J.* **2015**, *109* (3), 618−629.
(46) Levene, S. D.; Giovan, S. M.; Hanke, A.; Shoura, M. J. *The thermodynamics of DNA loop formation, from J. to Z.*; Portland Press Ltd.: 2013.

https://doi.org/10.1021/acs.jpcb.1c00763
J. Phys. Chem. B 2021, 125, 4645−4653