

Manuscript submitted to *Biophysical Journal*

Article

DNA looping by protamine follows a nonuniform spatial distribution

Ryan B. McMillan¹, Victoria D. Kuntz¹, Luka M. Devenica¹, Hilary Bediako¹, and Ashley R. Carter^{1*}

¹Department of Physics, Amherst College, Amherst, MA 01002, USA

ABSTRACT DNA looping plays an important role in cells in both regulating and protecting the genome. Often, studies of looping focus on looping by prokaryotic transcription factors like *lac* repressor or by structural maintenance of chromosomes (SMC) proteins such as condensin. Here, however, we are interested in a different looping method whereby multivalent cations (charge $\geq +3$), such as protamine proteins, neutralize the DNA, causing it to form loops and toroids. We considered two previously proposed mechanisms for DNA looping by protamine. In the first mechanism, protamine stabilizes spontaneous DNA fluctuations, forming randomly distributed loops along the DNA. In the second mechanism, protamine binds and bends the DNA to form a loop, creating a distribution of loops that is biased by protamine binding. To differentiate between these mechanisms, we imaged both spontaneous and protamine-induced loops on short-length ($\leq 1\mu\text{m}$) DNA fragments using atomic force microscopy (AFM). We then compared the spatial distribution of the loops to several model distributions. A random looping model, which describes the mechanism of spontaneous DNA folding, fit the distribution of spontaneous loops, but it did not fit the distribution of protamine-induced loops. Specifically, it overestimated the number of loops that form at the ends of the molecule and failed to predict a peak in the spatial distribution of loops at an intermediate location along the DNA. An electrostatic multibinding model, which was created to mimic the bind-and-bend mechanism of protamine, was a better fit of the distribution of protamine-induced loops. In this model, multiple protamines bind to the DNA electrostatically within a particular region along the DNA to coordinate the formation of a loop. We speculate that these findings will impact our understanding of protamine's *in vivo* role for looping DNA into toroids and the mechanism of DNA condensation by multivalent cations more broadly.

SIGNIFICANCE DNA looping is important in a variety of both *in vivo* functions (e.g. gene regulation) and *in vitro* applications (e.g. DNA origami). Here, we sought a mechanistic understanding of DNA looping by multivalent cations ($\geq +3$), which condense DNA into loops and toroids. One such multivalent cation is the protein protamine, which condenses DNA in sperm. We investigated the mechanism for loop formation by protamine and found that the experimental data was consistent with an electrostatic multibinding model in which two protamines bind electrostatically to the DNA within a 50-nm region to form a loop. This model is likely general to all multivalent cations and may be helpful in applications involving toroid formation or DNA nanoengineering.

INTRODUCTION

DNA looping plays a set of diverse and critical roles. In cells, DNA loops can activate or repress genes in prokaryotes (1–4), organize and compact the genome in eukaryotes (5–9), or compact the entire genome in a sequence-independent manner in sperm and bacteria (10–13). In DNA nanoengineering, synthetic looping proteins have allowed for self assembly of DNA-protein nanostructures (14).

There are different methods of DNA looping, each of which involves a different mechanism. One method of loop formation, spontaneous looping, occurs when thermal fluctuations cause two distal DNA segments to come together in the absence of proteins, creating a transient spontaneous loop (15, 16). An experimental realization of spontaneous loops is when DNA is kinetically trapped on a surface (17).

A second method of loop formation occurs when a protein leverages the thermal fluctuations in the DNA to form loops of a specific size. For example, the prokaryotic transcription factors *lac* repressor and AraC bind to one region of the DNA in a sequence-specific manner and then wait until a thermal fluctuation of the DNA brings a second site in contact with the transcription factor (1–4). Another well-studied looping method is loop extrusion (5–8, 18, 19). In this method, proteins like condensin hydrolyze ATP in order to unidirectionally translocate along DNA, processively enlarging a loop as they move (5, 8).

In this study, we focus on the less-well-understood looping mechanism of multivalent cations, which are also called DNA condensing agents (20, 21). Some common examples of DNA condensing agents are cobalt (III) hexaammine (22,

McMillan *et al.*

23), spermine (24–26), spermidine (26–29), and protamine (16, 17, 30), although any cation with a charge of at least +3 is thought to function similarly (20), and some divalent cations have also been shown to condense DNA under certain conditions (20). DNA condensing agents are known to bind to DNA nonspecifically (11, 20) and form loops (16, 27, 31), as well as toroids (20, 32, 33). Toroid sizes vary, but toroids generally have an outer diameter of about 100 nm (17, 32) and can contain up to 50 kbp of DNA (30) in concentrically wound, hexagonally packed loops (22). Understanding the mechanism of loop formation by condensing agents would provide insight into DNA toroid formation, as well as looping mechanisms more broadly.

Here, our goal is to understand how condensing agents loop DNA. We will focus on the protein protamine. Protamines are a family of small (~50-amino-acid), arginine-rich, positively charged proteins (11), which bind and neutralize the negatively charged DNA before folding the DNA into a loop or toroid (34). To fold the DNA into a loop, a previous model suggested that protamine stabilizes spontaneous loops (15, 32). Recent evidence, however, suggests that protamine instead forms loops via a bind-and-bend mechanism (16) in which each protamine binds the DNA and induces a small (~20°) bend.

We can differentiate between these two models by examining the spatial distribution of loops. For example, if loop formation is spontaneous, then loops will be equally likely to initiate at any point along the length of the molecule. If loop formation occurs via a bind-and-bend mechanism, then biases in protamine binding will affect the distribution of loops.

To measure the spatial distribution of loops, we used atomic force microscopy (AFM) to image short-length ($\leq 1\mu\text{m}$) DNA fragments with and without protamine (16, 35). We then compared the experimentally observed spatial distribution of loops to the predictions of three models: (i) a random looping model that assumes that loop formation is unbiased and spontaneous, (ii) an electrostatic binding model that assumes that loop formation occurs when electrostatic interactions cause a single protamine to bind to the DNA, and (iii) an electrostatic multibinding model based on the bind-and-bend mechanism that assumes that loop formation occurs when electrostatic interactions cause multiple protamines to bind to a single DNA region.

We found that the random looping model fit the spatial distribution of spontaneous loops (average residual of 0.016 ± 0.002 , comparable to the measurement error of 0.018 ± 0.002), but not the distribution of protamine-induced loops (average residual of 0.04 ± 0.01 for 398-nm-length DNA, a factor of ~2 higher than the measurement error of 0.028 ± 0.005). In particular, the distribution of protamine-induced loops contained a peak in the data at a fractional DNA length of 0.1–0.3. The electrostatic multibinding model was able to predict the location of a peak in the data (0.2 ± 0.1 in 398-nm-length DNA). Thus, our data is consistent with the electrostatic multibinding model that mimics the bind-and-bend mechanism, but not with the random looping model that

assumes that loop initiation is unbiased and uniform.

MATERIALS AND METHODS

Preparing DNA constructs and protamine

We generated DNA of lengths 217 nm (639 bp), 398 nm (1170 bp), and 1023 nm (3008 bp) using PCR (36). Specifically, we used bacteriophage lambda DNA (New England Biolabs N3011; Ipswich, MA) as a template, custom oligonucleotide primers (Integrated DNA Technologies; Coralville, IA), and an LA *Taq* DNA polymerase (TaKaRa Bio RR004; Kusatsu, Japan). We verified that products had amplified correctly using gel electrophoresis and then extracted the DNA using a commercial kit (Qiagen QIAquick Gel Extraction Kit, 28704; Hilden, Germany). Finally, we measured the concentration and purity using a spectrophotometer (ThermoFisher NanoDrop Lite; Waltham, MA). Samples with A260/A280 ratios of less than 1.7 were discarded.

We purchased protamine from salmon (Sigma-Aldrich P4005; Saint Louis, MO), diluted it in deionized water, and stored 30 μM aliquots at -20°C .

Preparing AFM Slides

AFM slides were prepared by affixing 10-mm-diameter ruby muscovite mica slides (Ted Pella grade V1; Redding, CA) to metallic disks. To create a clean surface, we used tape to remove the top layer of the mica. We prepared a 20 μL solution without protamine that contained 1.0 $\text{ng}/\mu\text{L}$ DNA and 2.0 mM magnesium acetate. After pipetting this solution onto the surface, we waited ~30 seconds and then washed the mica with 1 mL of deionized water and dried with nitrogen.

For sample preparation with protamine, we used a different procedure to reduce DNA aggregation (16). Specifically, we prepared 20 μL solutions of 0.2 $\text{ng}/\mu\text{L}$ DNA, 2.0 mM magnesium acetate, and protamine concentrations of either 0.2 μM , 0.6 μM , 2.0 μM , 3.5 μM , or 5.0 μM . We pipetted this solution onto the surface of the mica and then immediately (~2 s wait) washed with 1 mL of deionized water and dried with nitrogen. We then repeated this procedure until there were a total of 5 depositions on the mica. All samples were stored in a desiccator.

Imaging AFM slides

The AFM images were captured using a Dimension 3000 AFM with a Nanoscope IIIa controller (Digital Instruments; Tonawanda, NY). AFM tips (PPP-XYNCSTR-model, Nanosensors; Neuchatel, Switzerland; Parameters: resonant frequency = 150 kHz, force constant = 7.4 N/m, length = 150 μm , tip radius <7 nm) were used in tapping mode. We took images using a scan rate of either 2 Hz or 4 Hz. The image size was either 2 μm x 2 μm (512 x 512 pixels) or 1 μm x 1 μm (256 x 256 pixels).

Analyzing AFM slides

Image processing of AFM slides was done using Gwyddion (37). Images were corrected using three steps. First, we aligned rows using a 5th degree polynomial. Second, we removed high-frequency oscillations using an FFT filter. Third, we removed scars. After we corrected images, we identified DNA singlets as molecules that were lying flat on the surface with at least 1 pixel of separation between other molecules. The DNA contour length L_C had to be within 20% of the nominal length for that construct. About 70% of molecules passed this cut on L_C . Finally, we cropped and saved square images of singlets (Fig. S1-S2, Supplemental Information). Images were 200 x 200 nm for 217-nm-length DNA, 400 x 400 nm for 398-nm-length DNA, and 400 x 400 nm or larger as needed for 1023-nm-length DNA.

We examined these singlets to identify loops and flowers. A loop had to completely enclose one region of bound area. A flower had one or more loops and a central point at which they all come together. Flowers were subclassified by the number of loops.

We measured three quantities for each singlet (Fig. S3). First, we took two perpendicular diameter measurements and then averaged them together to compute the diameter d . Second, we measured the start site s_s as the arc length from the closer DNA end to the crossover point of the loop. Third, we measured the contour length L_C of the DNA. The measurement error for all three of these variables was 3 nm, or <1 pixel. For flowers, we measured d for each loop individually, and s_s was defined as the arc length from the closer DNA end to the crossover point of the flower.

Simulating loop formation with a random looping model

We wrote functions to simulate loop formation in MATLAB (MathWorks; Natick, MA). All simulations discretized the DNA into segments of length 1 nm. Simulations compared to experimental data ran 100,000 iterations to produce 100,000 loop start site s_s values, while simulations used to explore the parameter space used 10,000 iterations.

The first model we developed was the random looping model (Fig. S4), which simulates random, unbiased loop formation. It does not consider any protamine-protamine interactions or properties of the DNA. In this model, we consider a polymer of contour length L_C . To simulate the process of loop formation, we perform the following steps to produced a single s_s value:

1. Output a loop initiation site s_i from a discrete uniform random distribution.
2. Output a loop circumference C by drawing from a Gamma distribution of the experimentally observed loop circumferences (see Fig. S5 and Modeling loop formation using Gamma distributions, Supplemental Information).

3. Compute two candidate s_s values as $s_i - C/2$ and $L_C - (s_i + C/2)$. s_s is the smaller of these two. If s_s is less than 0, then we assume that the loop reaches the end of the polymer and record $s_s = 0$.

Simulating loop formation with an electrostatic binding model

We developed a second model that would incorporate the effects of the electrostatic interactions between protamine and DNA. In this model, we treat the DNA as a line of charge with uniform charge density $-\lambda$ and the protamine as a point charge $+q$ located a distance d away from the DNA. Protamine is a distance of a from the left end of the DNA (measured along the DNA) and a distance of b from the right end of the DNA. Choosing $d = \infty$ as our reference point, the potential of this geometry is:

$$V = -k\lambda \ln \left(\frac{(b^2 + d^2)^{1/2} + b}{(a^2 + d^2)^{1/2} - a} \right). \quad (1)$$

It is useful to recast this equation in terms of the total DNA contour length L_C and loop initiation site s_i . We make the substitutions $a = s_i$, $b = L_C - s_i$ to find that

$$V = -k\lambda \ln \left(\frac{((L_C - s_i)^2 + d^2)^{1/2} + L_C - s_i}{(s_i^2 + d^2)^{1/2} - s_i} \right). \quad (2)$$

We want to use this potential to derive the probability distribution function for protamine binding. If we assume that the temperature of the system is fixed, then we can apply Boltzmann statistics to find the binding probability. The probability distribution function as a function of s_i is then:

$$f(s_i) = \frac{\exp\left(-\frac{qV}{k_B T}\right)}{Z}, \quad (3)$$

where Z is the partition function of the system and V is given by Eq. 2. Substituting this expression into the probability distribution function gives:

$$f(s_i) = \frac{1}{Z} \exp \left[\frac{qk\lambda}{k_B T} \ln \left(\frac{((L_C - s_i)^2 + d^2)^{1/2} + L_C - s_i}{(s_i^2 + d^2)^{1/2} - s_i} \right) \right]. \quad (4)$$

We can rewrite this equation as:

$$f(s_i) = \frac{1}{Z} \left(\frac{((L_C - s_i)^2 + d^2)^{1/2} + L_C - s_i}{(s_i^2 + d^2)^{1/2} - s_i} \right)^{qk\lambda/k_B T}. \quad (5)$$

Observe that $\frac{qk\lambda}{k_B T}$ represents the order of magnitude of the electrostatic force relative to the thermal fluctuations. We will define this as the variable θ , such that:

$$\theta = \frac{qk\lambda}{k_B T}. \quad (6)$$

In the limit of very high charges, $\theta \rightarrow \infty$, while in the limit of low charges, $\theta \rightarrow 0$. At $\theta = 1$, the electrostatic potential

McMillan *et al.*

energy is of the same order of magnitude as the thermal energy.

We then compute s_s values for fixed L_C and θ using the following steps:

1. Generate loop initiation sites along the DNA using the method of rejection sampling on Eq. 5, which has been previously implemented in MATLAB (38).
2. Use steps 2-3 of the random looping model to generate s_s .

Simulating loop formation with an electrostatic multibinding model

Our third model is based on the bind-and-bend mechanism of protamine-induced DNA folding. In this model, we build upon the electrostatic binding model and allow multiple protamines to bind the same DNA molecule. We consider a discretized polymer of contour length L_C , just as we did for the random looping and electrostatic binding models. We use the following steps to generate s_s values:

1. Use step 1 of the electrostatic binding model to generate a candidate s_i value.
2. Repeat step 1 for n protamines, updating the probability distribution after every iteration by neutralizing the charge in a 10-nm region centered around the prior protamine's binding site (Fig. S6).
3. Once all n protamines have been placed, check that the two outermost binding sites are no more than the maximum distance m apart. If they are too far apart, then reset the probability distribution and return to step 1.
4. Select s_i from the n candidate s_i values as the s_i of the outermost protamine binding site.
5. Use steps 2-3 of the random looping model to generate s_s .

Statistics for loop formation histograms

Experimental and simulated s_s data were imported into Igor (WaveMetrics; Portland, OR). s_s data were plotted in histograms with a binwidth of 3 pixels, or 11.7 nm, which is about 4x measurement error. The x-axis of the histogram was divided by L_C to create a histogram of fractional DNA length. The height of each bin in the histogram was normalized such that all bins summed to one. Residuals in the height of each bin (experiment-simulation) were also computed and displayed for each distribution. To compute the error on the height of each bin in the experimental data, we used Poisson statistics (39). Specifically, the error on a bin with N_{bin} observations is

$N_{bin}^{1/2}$. Because bins were normalized by the total number of observations in the histogram N , the reported error is:

$$\text{Error} = \frac{N_{bin}^{1/2}}{N}. \quad (7)$$

RESULTS

Spontaneous loops are fit by a random looping model

Before describing protamine-induced loop formation, we wanted to develop a model that accurately describes spontaneous loop formation. Spontaneous loops occur when random thermal fluctuations cause the DNA to bend and transiently overlap. Since the mechanism of spontaneous loop formation is random, we hypothesized that spontaneous loop formation should follow a random looping model (see Materials and Methods) in which loops are equally likely to initiate at any point along the DNA.

To test this model, we used DNA kinetically trapped on a surface as our source of spontaneous loops. In this setup, spontaneous looping is more likely to occur for DNA contour lengths, L_C , that are much greater than the DNA persistence length, L_P . The L_P of the DNA (50 nm (40, 41)) is the length over which the tangent vector to the DNA remains correlated (42) and is essentially the length that the DNA is fairly stiff and straight. If L_C is much longer than L_P , then the DNA can bend due to thermal fluctuations, creating a spontaneous loop. We found that only 6% of molecules in 217- and 398-nm-length DNA had a spontaneous loop, compared to 85% of molecules in 1023-nm-length DNA.

Thus, to measure the spatial distribution of spontaneous loops, we immobilized long DNA molecules ($L_C = 1023$ nm) on a 2D surface in the absence of folding agents (Fig. 1A). As the molecules adhered to the surface, random thermal bending of the DNA created spontaneous loops. We then imaged the DNA on the surface with an AFM (Fig. 1B), which captured the structure of the DNA and allowed for the visualization of loops. For each molecule with a single loop ($N=44$), we measured the start site of the loop, s_s , as the length along the contour of the DNA from the closest DNA end to the DNA crossover point.

To generate the spatial distribution of simulated loops using the random looping model, we randomly chose sites along the DNA to initiate a loop and calculated s_s . Specifically, the program first chose a loop initiation site s_i using a uniform probability distribution (Fig. 1C). Second, the program measured the loop start site s_s as the distance from the closest DNA end to the location that is half of the loop circumference, $C/2$, from the initiation site. If the initiation site was $<C/2$ from the DNA end, then the s_s was set to zero. Finally, we plotted the simulated s_s/L_C , along with the measured s_s/L_C , in a histogram and calculated the residuals (experiment-simulation) between the two data sets (Fig. 1D).

The two distributions were very similar. The residuals had

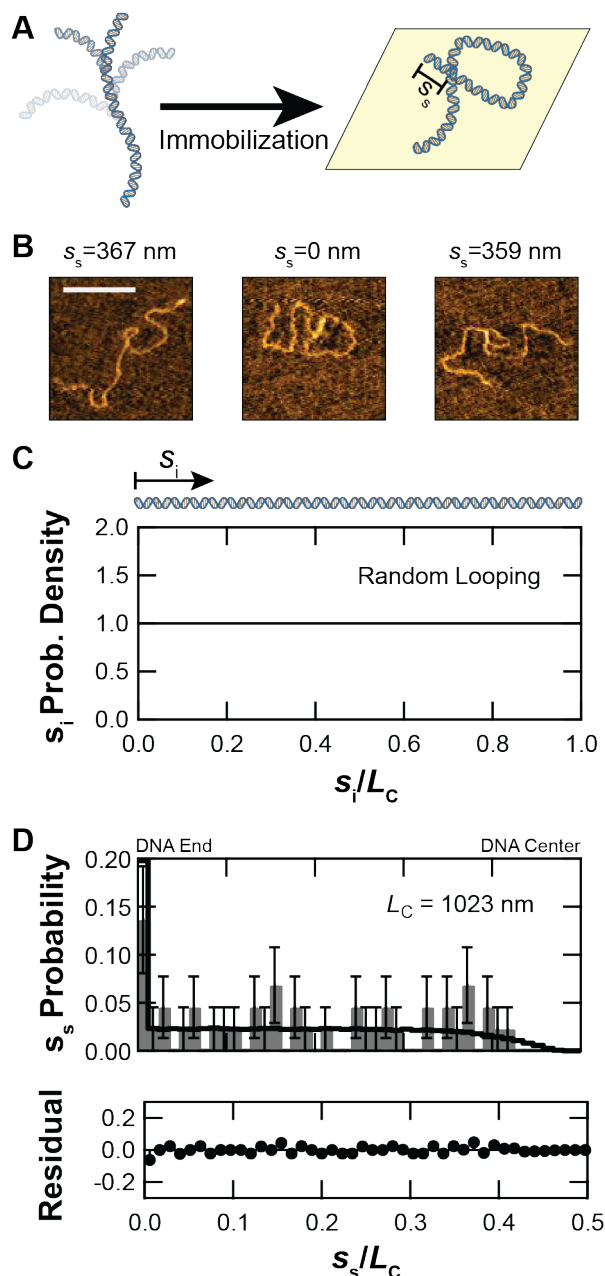


Figure 1: Spontaneous loops form according to a random looping model. A) To measure spontaneous loop formation, we immobilized DNA of contour length $L_C = 1023$ nm on the surface of an AFM slide without protamine. We determined the start site $s_s \in [0, L_C/2]$ as the arc length from the closest DNA end to the DNA crossover location. B) Three sample loops with their measured start sites are shown. Scale bar is 200 nm. C) To simulate spontaneous loop formation, we assume that the initiation site for the loop $s_i \in [0, L_C]$ is distributed uniformly and generate s_s . D) We plot the simulated (black) and experimental (gray) start site distributions. Experimental data have a surplus at the predicted theoretical location of $s_s/L_C = 0$, and residuals are ≤ 0.06 for all subsequent bins up to the predicted falloff at $s_s/L_C \approx 0.45$.

an average of 0.016 ± 0.002 (mean \pm standard error), while the average measurement error was 0.018 ± 0.002 (Table S1), indicating that any deviation between the two distributions is likely attributable to measurement error. In addition, the model captures the flat distribution of equal probability across most of the fractional DNA length, and it captures the behavior at the end and center of the DNA. Specifically, in the first bin of the distribution, which corresponds to the end of the DNA, there is an increased probability of loops (bin height of 0.14 ± 0.06 for $s_s/L_C = 0-0.011$). This increased probability is an end effect that occurs because any loop that forms within a distance of $C/2$ of either end of the DNA will have an apparent start site of 0. Due to this end effect, the first bin is higher than the second by a factor of

$$1 + \frac{C}{2L_C * (\text{binwidth}/L_C)}. \quad (8)$$

Given that the peak C value is 103 nm (Fig. S5), this factor is about 5 for this distribution, which matches the experimental value of 6 ± 3 . Interestingly, our binwidth is normalized by L_C , so L_C has no effect on the relative height of the first bin. At the other end of the distribution, which corresponds to the center of the DNA, there are no start sites predicted by the model past the falloff of

$$0.5 - \frac{C}{2L_C}. \quad (9)$$

This is because the s_i that produces the largest start site is $L_C/2$, which creates a DNA crossover point at a distance of $C/2$ from this location. Again using $C = 103$ nm (Fig. S5), this creates a falloff point at $s_s/L_C = 0.45$, slightly larger than the measured falloff in the experimental distribution at 0.42 ± 0.01 .

The random looping model is therefore able to accurately describe the spatial distribution of spontaneous loops, indicating that spontaneous loops have loop initiation sites that are given by a uniform random distribution.

Protamine-induced loops do not follow the random looping model

Having confirmed that spontaneous loops are consistent with our random looping model, we next asked whether this model is also accurate for protamine-induced loops. Recently, we found that protamine-induced loops do not form in a single step (16). Instead, these loops form in multiple steps, with each step thought to correspond to one or more protamine molecules that bind the DNA and bend it into a particular radius of curvature (~ 10 nm) (16). Multiple folding events, rather than just one event, are then needed to bend the DNA into a loop. However, there is evidence that protamine bending and spontaneous thermal fluctuations might work together to form loops (16). In addition, if protamine binding follows a uniform distribution, then we might expect the random

McMillan *et al.*

looping model to describe protamine-induced loops as well as spontaneous loops.

To test the hypothesis that the spatial distribution of protamine-induced loops follows the random looping model, we measured the start sites for protamine-induced loops and compared this experimental data to the simulated start sites produced by the random looping model (Fig. 2). Specifically, we immobilized 217-nm-length ($N=77$ loops) and 398-nm-length ($N=59$ loops) DNA to a surface in the presence of 0.2-5.0 μM protamine (Fig. 2A). The shorter DNA lengths (217-398 nm) are used in this experiment as they are likely to form single loops, rather than toroids. We then imaged the DNA with an AFM to visualize the single, protamine-induced loops and measured s_s (Fig. 2B).

A histogram of the fractional start site s_s/L_C (Fig. 2C) shows poor agreement between the experimental data and the random looping model. The average residual for 217-nm-length and 398-nm-length DNA was 0.07 ± 0.03 and 0.04 ± 0.01 which is higher than the average error of 0.04 ± 0.01 and 0.028 ± 0.005 , respectively (Table S1). These residuals are also a factor of 2-3.5 higher than the residuals for the spontaneous looping data. More importantly, the random looping model does not predict the shape of the distribution. The first bin of the histogram, which represents the end of the DNA, has a measured probability density that is a factor of ~ 2 lower than the model would predict. Also, between the fractional start sites of 0.15-0.3, which corresponds to about a quarter of the way along the DNA, the model predicts a flat distribution, when in fact there is a peak in the experimental data. The random looping model did predict the end behavior of the distribution and measured a falloff at 0.38 ± 0.05 for 217-nm-length (predicted value of 0.33 for $C = 72$ nm, Fig. S5) and 0.41 ± 0.03 for 398-nm-length (predicted value of 0.38 for $C = 94$ nm, Fig. S5).

We also noticed that for one isolated bin ($s_s/L_C = 0.09$ -0.12 in the 398-nm-length DNA) there were no data points, indicating loops did not start at that location. Upon further inspection, the local AT content ($\sim 50\%$) is decreased in this region as compared to the rest of the DNA ($\sim 70\%$), suggesting that sequence-dependent effects might be responsible for the decreased looping probability (see Fig. S7 and Local DNA sequence variations, Supplemental Information). The random looping model does not account for local variations in DNA flexibility and curvature since it assumes that the DNA is uniformly flexible along its length. While DNA sequence is known to affect looping in general (43), we did not study this effect further here.

Thus, we conclude that protamine-induced loops are not formed uniformly along the DNA length as in the random looping model. Instead, protamine-induced loops have a lower bias for formation at the end of the DNA, and a higher bias for formation at about a quarter of the way along the DNA. This means that there must be some physical mechanism besides random looping that creates this bias. One possible physical effect is the electrostatic binding between the positively charged

protamine and the negatively charged DNA (20, 21), which should bias protamine binding and loop initiation toward the center of the DNA.

Protamine-induced loops are biased by electrostatic interactions

In an attempt to fit the spatial distribution of protamine-induced loops, we created the electrostatic binding model (see Materials and Methods). In this model, we assume that the protamine is a positive point charge and that the DNA is a negative line charge. We then calculate the electrostatic potential energy for protamine binding given this assumption. Electrostatic binding of protamine would bias loop formation away from the DNA end and could be a physical effect that creates the peak in the experimental spatial distribution a quarter of the way along the DNA.

The result of these assumptions is that the loop initiation site s_i follows an inverse parabolic distribution (Fig. 2D) rather than the uniform distribution assumed in the random looping model (Fig. 1C). The height of the distribution, or how strong the binding probability at the center is compared to the DNA end, is set by the parameter θ (Eq. 6). This parameter is inversely proportional with the level of thermal noise in the system and proportional to the magnitude of the electrostatic potential energy between the protamine and DNA. Thus, we found that increasing θ shifts the s_s/L_C distribution towards the center of the DNA (Fig. S8). The optimal value for θ was one, meaning that the thermal fluctuations and electrostatic interactions are of the same magnitude. This value matched our order-of-magnitude estimate of θ (see An order-of-magnitude estimate of θ , Supplemental Information).

We used the electrostatic binding model to generate s_s and plotted the spatial distribution of the simulated loops (Fig. 2C). This simulated distribution was a better fit to the experimental distribution of loops than the random looping model distribution. The electrostatic binding model predicts within error the height of the first bin, and the average residuals (0.04 ± 0.01 for $L_C = 217$ nm, 0.03 ± 0.01 for $L_C = 398$ nm) were comparable to the average experimental error (0.04 ± 0.01 for $L_C = 217$ nm and 0.028 ± 0.005 for $L_C = 398$ nm) (Table S1). However, the model predicts a steady increase in the probability over the interval $s_s/L_C = 0.1$ -0.4, rather than the peak seen in the experimental data.

We conclude that the electrostatic binding model is a better model of protamine-induced looping than the random looping model, but that it does not fully describe the mechanics of loop formation. If electrostatics is biasing protamine binding and therefore DNA looping toward the center of the DNA, then there must be some other physical effect biasing loop formation away from the center to create a peak in the spatial distribution of loops at an intermediate value.

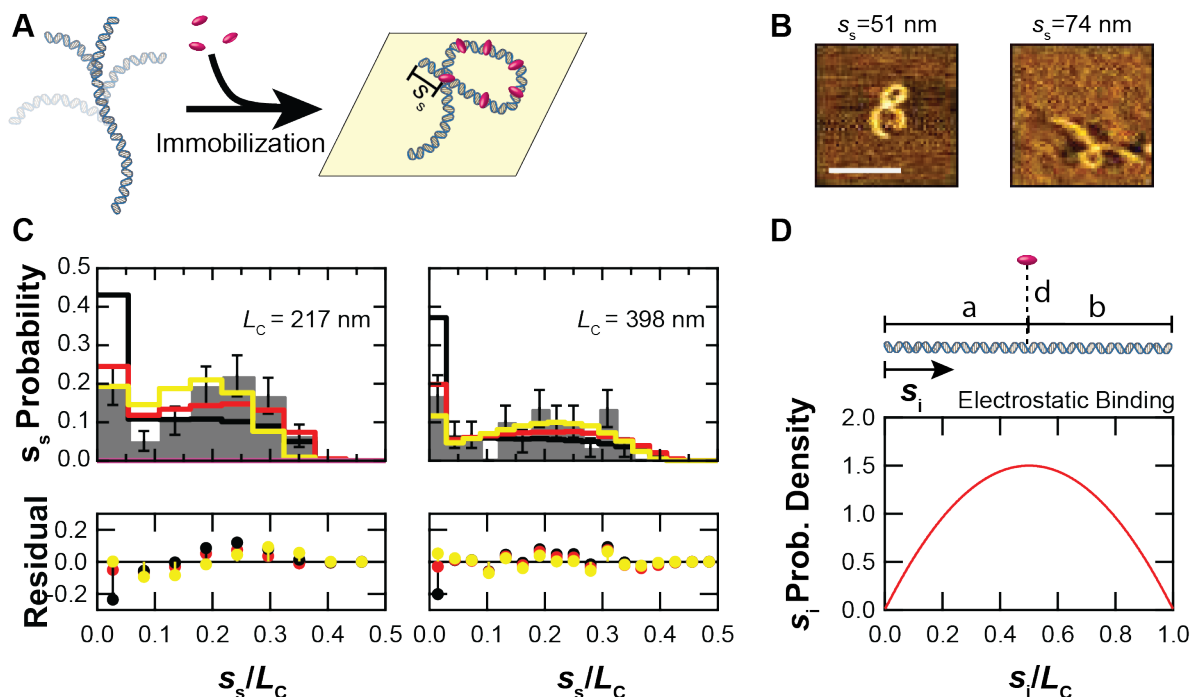


Figure 2: Single loops are best described by an electrostatic multibinding model. A) Experimental setup for protamine-induced loops. DNA in solution is bent by protamine and then immobilized on a surface so that we can measure s_s . B) Sample AFM images of loops in 217-nm-length DNA are shown. Scale bar is 100 nm. C) We plot a histogram of the fractional start site location s_s/L_C for the experimental data (*gray*) as well as the simulations for the random looping model (*black*), the electrostatic binding model (*red*), and the electrostatic multibinding model (*yellow*) for single loops in both 217-nm-length (*left*) and 398-nm-length (*right*) DNA. Residuals are experiment-simulation. D) Setup and probability density for the electrostatic binding model. Protamine is located a distance d above the DNA molecule, which is perfectly linear. Its projection onto the DNA is a distance a from the left end of the molecule and a distance b from the right end of the molecule. In the diagram, d is drawn as if it is comparable in magnitude to a and b for clarity, but in practice, $d \ll a, b$.

Protamine-induced loops follow the electrostatic multibinding model

Thus, to further update our model, we considered how the bind-and-bend mechanism of protamine-induced looping might create a peak in the spatial distribution of loops a quarter of the way along the DNA. In the bind-and-bend mechanism, loop formation involves multiple steps of protamine binding and bending (16). To account for this coordinated binding of multiple protamine molecules, we developed the electrostatic multibinding model (see Materials and Methods). In this model, n protamines bind the DNA using the probability distribution that was developed for the electrostatic binding model (Fig. 2D), except as each protamine binds, the electrostatic potential is updated to account for the bound protamines (Fig. S6). In addition, all protamines that bind the DNA must bind within a particular distance m of each other to cooperatively bend the DNA into a loop. Finally, the model assumes that the binding location of the outermost protamine is the initiation site (Figs. S9), which leads to a bias in loop formation towards the DNA end (Fig. S10).

To set the three parameters in the electrostatic multibinding model, we ran the simulation under different conditions. We then plotted the spatial distributions of the simulated loops (Figs. S11-S13) and optimized the parameters. We first optimized the parameter θ , which describes the strength of the electrostatic effect relative to thermal noise (Fig. S11). Increasing θ pushes the start site distribution towards the center of the DNA. We found an optimal value of 1.5, which is on the same order of magnitude as the predicted value (see An order-of-magnitude estimate of θ , Supplemental Information). We also examined the effect of varying the parameter for the number of protamines n (Fig. S12). Increasing n shifts the distribution towards the DNA end, presumably due to the higher likelihood that at least one protamine will bind away from the DNA center. We found that 2 "molecules" (which could really be 2 groups of molecules if there is cooperative binding) is the best fit to the experimental data. Decreasing the parameter for the maximum distance between the molecules m (Fig. S13) shifts the distribution towards the center of the molecule. The simulation that produced the best results had $m = 50$ nm, which happens to be the persistence

McMillan *et al.*

length of DNA (41). This observation is interesting physically because it suggests that binding within an L_P produces a loop. This would be the case if loop formation requires correlated bending, which might not occur over length scales much longer than L_P .

After selecting the model parameters ($\theta = 1.5$, $n = 2$ molecules, $m = 50$ nm), we then compared our simulated distribution to the experimental data (Fig. 2C). For both DNA lengths, the electrostatic multibinding model predicted the height of the first bin within error. In addition, the model predicted peaks at an s_s/L_C of 0.18 ± 0.06 (mean \pm standard deviation of Gaussian fit to data) and 0.2 ± 0.1 , which agreed with the peaks in the experimental data of 0.23 ± 0.06 and 0.22 ± 0.04 in the 217-nm-length and 398-nm-length data, respectively.

Thus, there seem to be three effects that create a peak a quarter of the way along the DNA: the electrostatic interactions which bias loop formation towards the center of the DNA, the coordinated binding of multiple protamines within a persistence length of each other that bias loop formation towards the center of the DNA, and the fact that loop initiation is set to be the position of the outermost protamine, which biases loop formation towards the end of the DNA.

Protamine-induced flowers follow the electrostatic multibinding model

To create another test of the electrostatic multibinding model, we wondered if our models of loop formation would generalize to the formation of DNA flowers (27, 44). DNA flowers are multi-looped DNA structures that form in the presence of protamine (44) or other folding agents (27, 45) and look flower-like when immobilized on a surface, since all of the loops overlap each other at a central location. Flowers are thought to be an intermediate step in toroid formation (44) and likely form when multiple protamines bind and bend the DNA into several loops. Here, we assume that the initiation mechanism of the flower is the same as that of the loop and that the multiple loops in the flower can be modeled as a single loop with a larger circumference.

To measure the experimental s_s distribution, we immobilized DNA ($L_C = 398$ nm) on the surface in the presence of protamine (0.2-5 μ M), as before, and used an AFM to image 2-looped (N=78) and 3-looped (N=42) flowers (Fig. 3A). The start site s_s was measured as the distance along the DNA from the closest DNA end to the location where all the loops overlap each other. We then histogrammed the fractional start site locations. We found that the shape of this spatial distribution was similar to the shape of the spatial distribution for single protamine-induced loops and contained a surplus in the first bin and a second peak at an intermediate location along the DNA.

We compared this experimental spatial distribution to the predicted distributions output by our three models—random looping, electrostatic binding, and electrostatic multibinding

(Fig. 3B-C). We did not vary the parameters m , θ , and n from their previous values. The only change we made was to update the loop circumference C (see Fig. S5 and Modeling loop formation using Gamma distributions, Supplemental Information). For both 2-loop and 3-loop flowers, the electrostatic binding and electrostatic multibinding models predicted the height of the first bin within error, whereas the random looping model did not. The residuals for the electrostatic binding and electrostatic multibinding models were also both within the measurement error of the dataset (Table S2). However, only the electrostatic multibinding model predicted the position of the peak in the experimental data (experimental peak was at 0.14 ± 0.03 and 0.12 ± 0.03 and the predicted peak was at 0.1 ± 0.1 and 0.1 ± 0.1 for 2-loop and 3-loop flowers, respectively). We note that this model did underestimate the height of the peaks.

We thus find that the electrostatic multibinding model fits both the spatial distribution of loops and flowers. This suggests that both are consistent with the bind-and-bend mechanism of protamine looping rather than the a mechanism that depends on spontaneous looping.

DISCUSSION

Here, our goal was to characterize the mechanism behind protamine-induced loop formation. We used AFM to image DNA that had formed spontaneous loops, protamine-induced loops, and protamine-induced flowers. We combined this experimental data with computational modeling to compare our results to predicted outcomes from three different models of loop formation: random looping, electrostatic binding, and electrostatic multibinding. Using this approach, we found that the random looping model that has a uniform probability for the loop initiation site describes the spatial distribution of spontaneous loops (average residual = 0.016 ± 0.002) within the experimental error (0.018 ± 0.002), but not protamine-induced loops or flowers. The distributions of protamine-induced loops and flowers have fewer loops in the first bin of the histogram than the random looping model would predict, and the distributions have an additional peak at a fractional DNA length of 0.1-0.3. The electrostatic multibinding model is consistent with the data (average residuals within experimental error, see Tables S1-S2) and predicts the location of the peak in the spatial distribution at a fractional location of 0.1-0.3 (Table S3). Thus, the spatial distribution of protamine-induced loops is explained by three physical effects in the electrostatic multibinding model: electrostatic protamine-DNA interactions, the coordinated binding of multiple protamines within a ~ 50 nm region, and the fact that the protamine nearest the DNA end sets the initiation site of the loop.

There are a few limitations to these conclusions. First, we assumed that the flexibility along the DNA is constant. This is not true, as local DNA sequence variations can play an important role in setting the DNA flexibility (43, 46), and are thought to cause the lack of loop start sites in the 398-

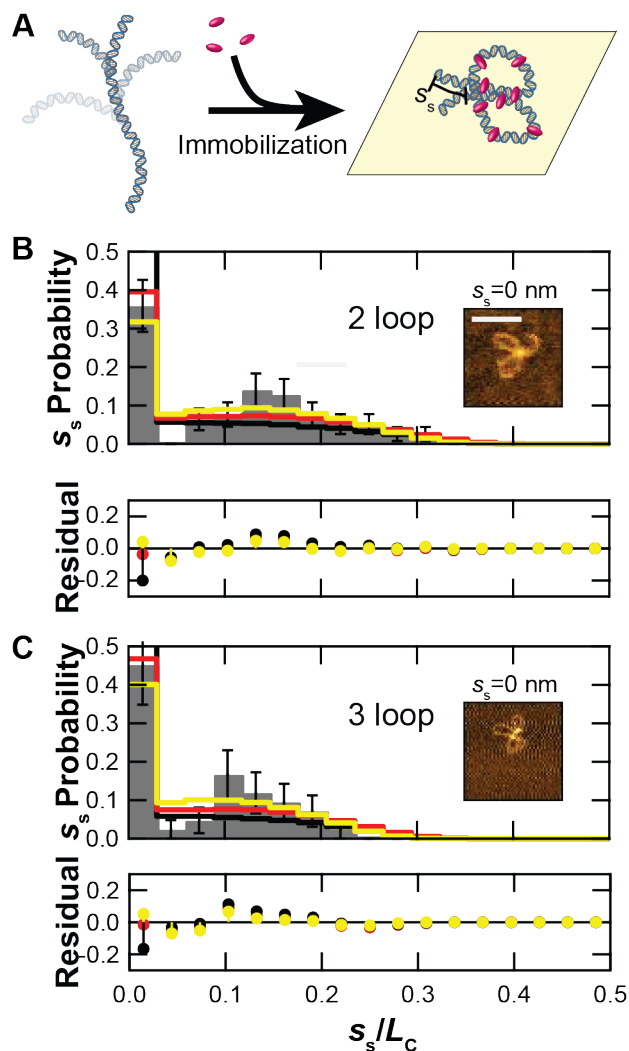


Figure 3: Flowers also follow an electrostatic multibinding model. A) To study flower formation, we immobilized DNA of contour length $L_C = 398$ nm on the surface of an AFM slide in the presence of protamine. Now the start site $s_s \in [0, L_C/2]$ is the arc length to the flower central point. B-C) We collected and plotted the probability at each fractional start site for 2-loop (B) and 3-loop (C) flowers (gray). We also simulated distributions for the random looping model (black), the electrostatic binding model (red), and the electrostatic multibinding model (yellow). *Inset:* 2-loop and 3-loop flowers extracted from AFM images. Scale bar is 100 nm.

nm-length DNA at a bin value of $s_s/L_C = 0.09 - 0.12$ (Fig. 2C, Fig. S7). Second, we assumed that the DNA is fairly stiff over its length since our experimental data has $L_C < 10L_P$ to prevent toroid formation. Longer molecules ($>10L_P$) would be floppier and have more spontaneous looping. Spontaneous looping might cause the spatial distribution of loops to look more like the one predicted by the random looping model. Third, performing this experiment *in vivo* with phosphorylated protamine and DNA with bound proteins might affect binding probabilities and the spatial distribution of loops. Finally, we found that the electrostatic multibinding model is consistent with the experimental data, but other models could also fit the data. More in-depth studies would be needed to determine the effects of these limitations.

Given these limitations, we hypothesize that the electrostatic multibinding model will shed insight on the toroid formation process (32). This is because two early steps in toroid formation—loop formation and flower formation—appear to follow a similar spatial distribution. This makes it likely that later steps in the toroid formation process, including the formation of toroids themselves, might also follow this spatial distribution. Second, we find two protamines (or two groups of protamines with cooperative binding) are needed to form a loop. This matches a prior study (47) that found that toroids folded by spermine have two interactions/loop. If one molecule or group of molecules facilitates one interaction, then this data would be consistent with the electrostatic multibinding model.

We also note that our results point towards a general mechanism of looping by multivalent cations (20, 21, 32). Future work could investigate how binding probability, charge, or concentration for different multivalent cations might affect the electrostatic multibinding model.

Future work might also examine how looping by protamine using the electrostatic multibinding model compares to looping by other proteins, such as *lac* repressor and condensin. Protamine could also be compared to DNA bridging proteins such as the histone-like nucleoid-structuring (H-NS) protein in bacteria (10, 13). H-NS also uses multiple molecules to nonspecifically bind and compact the DNA.

Finally, we speculate that our results may aid in the design of DNA nanostructures (48, 49). Synthetic looping proteins have been used previously to form specific DNA contacts (50) which have aided in the assembly of DNA nanostructures (14). Here, we speculate that multivalent cations with nonspecific contacts could also be used to bend, condense, or stabilize engineered DNA constructs.

AUTHOR CONTRIBUTIONS

RBM and ARC designed the research, took the data, ran the analysis, and wrote the article. VDK analyzed the data. LMD took the data and analyzed the data. HB took the data and assisted in writing the article.

McMillan *et al.*

ACKNOWLEDGMENTS

This work was supported by a Cottrell Science Award from the Research Corporation for Scientific Advancement (ARC, Project #23239), a CAREER award from the National Science Foundation (ARC, Project #1653501), and Amherst College. RBM was partially funded by a Barry M. Goldwater Scholarship.

The authors declare no conflicts of interest.

SUPPORTING CITATIONS

References (51–53) appear in the Supplemental Information.

REFERENCES

- Schleif, R., 1992. Dna Looping. *Annual Review of Biochemistry* 61:199–223. <https://doi.org/10.1146/annurev.bi.61.070192.001215>, eprint: <https://doi.org/10.1146/annurev.bi.61.070192.001215>.
- Matthews, K. S., 1992. DNA looping. *Microbiology and Molecular Biology Reviews* 56:123–136. <https://mmbr.asm.org/content/56/1/123>, publisher: American Society for Microbiology Journals Section: Research Article.
- Martin, K., L. Huo, and R. F. Schleif, 1986. The DNA loop model for ara repression: AraC protein occupies the proposed loop sites in vivo and repression-negative mutations lie in these same sites. *Proceedings of the National Academy of Sciences* 83:3654–3658. <https://www.pnas.org/content/83/11/3654>, publisher: National Academy of Sciences Section: Research Article.
- Cournac, A., and J. Plumbridge, 2013. DNA Looping in Prokaryotes: Experimental and Theoretical Approaches. *Journal of Bacteriology* 195:1109–1119. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3591992/>.
- Ruiten, M. S. v., and B. D. Rowland, 2018. SMC Complexes: Universal DNA Looping Machines with Distinct Regulators. *Trends in Genetics* 34:477–487. [https://www.cell.com/trends/genetics/abstract/S0168-9525\(18\)30056-8](https://www.cell.com/trends/genetics/abstract/S0168-9525(18)30056-8), publisher: Elsevier.
- Hassler, M., I. A. Shaltiel, and C. H. Haering, 2018. Towards a unified Model of SMC Complex Function. *Current biology : CB* 28:R1266–R1281. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6850909/>.
- Yuen, K. C., and J. L. Gerton, 2018. Taking cohesin and condensin in context. *PLoS Genetics* 14. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5784890/>.
- Terakawa, T., S. Bisht, J. M. Eeftens, C. Dekker, C. H. Haering, and E. C. Greene, 2017. The condensin complex is a mechanochemical motor that translocates along DNA. *Science (New York, N.Y.)* 358:672–676.
- Krepel, D., R. R. Cheng, M. D. Pierro, and J. N. Onuchic, 2018. Deciphering the structure of the condensin protein complex. *Proceedings of the National Academy of Sciences* 115:11911–11916. <https://www.pnas.org/content/115/47/11911>, publisher: National Academy of Sciences Section: Physical Sciences.
- Rimsky, S., 2004. Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. *Current Opinion in Microbiology* 7:109–114. <http://www.sciencedirect.com/science/article/pii/S1369527404000141>.
- Balhorn, R., 2007. The protamine family of sperm nuclear proteins. *Genome Biology* 8.
- Dame, R. T., M. C. Noom, and G. J. L. Wuite, 2006. Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature* 444:387–390.
- Dame, R. T., M. S. Luijsterburg, E. Krin, P. N. Bertin, R. Wagner, and G. J. L. Wuite, 2005. DNA Bridging: a Property Shared among H-NS-Like Proteins. *Journal of Bacteriology* 187:1845–1848. <https://jb.asm.org/content/187/5/1845>, publisher: American Society for Microbiology Journals Section: STRUCTURAL BIOLOGY.
- Praetorius, F., and H. Dietz, 2017. Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes. *Science* 355. <https://science.sciencemag.org/content/355/6331/eaam5488>, publisher: American Association for the Advancement of Science Section: Research Article.
- Hud, N. V., K. H. Downing, and R. Balhorn, 1995. A constant radius of curvature model for the organization of DNA in toroidal condensates. *Proceedings of the National Academy of Sciences of the United States of America* 92:3581–3585.
- Ukogu, O. A., A. D. Smith, L. M. Devenica, H. Bediako, R. B. McMillan, Y. Ma, A. Balaji, R. D. Schwab, S. Anwar, M. Dasgupta, and A. R. Carter, 2020. Protamine loops DNA in multiple steps. *Nucleic Acids Research* 48:6108–6119. <https://academic.oup.com/nar/article/48/11/6108/5835822>, publisher: Oxford Academic.
- Allen, M. J., E. M. Bradbury, and R. Balhorn, 1997. AFM analysis of DNA-protamine complexes bound to mica. *Nucleic Acids Research* 25:2221–2226.
- Ganji, M., I. A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C. H. Haering, and C. Dekker, 2018. Real-time imaging of DNA loop extrusion by condensin. *Science* 360:102–105. <https://science.sciencemag.org/content/>

- 360/6384/102, publisher: American Association for the Advancement of Science Section: Report.
19. Davidson, I. F., D. Goetz, M. P. Zaczek, M. I. Molodtsov, P. J. Huis in 't Veld, F. Weissmann, G. Litos, D. A. Cisneros, M. Ocampo-Hafalla, R. Ladurner, F. Uhlmann, A. Vaziri, and J. Peters, 2016. Rapid movement and transcriptional re-localization of human cohesin on DNA. *The EMBO Journal* 35:2671–2685. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5167347/>.
 20. Bloomfield, V. A., 1997. DNA condensation by multivalent cations. *Biopolymers* 44:269–282.
 21. Teif, V. B., and K. Bohinc, 2011. Condensed DNA: Condensing the concepts. *Progress in Biophysics and Molecular Biology* 105:208–222. <http://www.sciencedirect.com/science/article/pii/S0079610710000659>.
 22. Hud, N. V., and K. H. Downing, 2001. Cryoelectron microscopy of lambda phage DNA condensates in vitreous ice: the fine structure of DNA toroids. *Proceedings of the National Academy of Sciences of the United States of America* 98:14925–14930.
 23. Schnell, J. R., J. Berman, and V. A. Bloomfield, 1998. Insertion of telomere repeat sequence decreases plasmid DNA condensation by cobalt (III) hexaammine. *Biophysical Journal* 74:1484–1491. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1299494/>.
 24. Leforestier, A., and F. Livolant, 2009. Structure of toroidal DNA collapsed inside the phage capsid. *Proceedings of the National Academy of Sciences* 106:9157–9162. <https://www.pnas.org/content/106/23/9157>, publisher: National Academy of Sciences Section: Physical Sciences.
 25. Leforestier, A., A. Šiber, F. Livolant, and R. Podgornik, 2011. Protein-DNA Interactions Determine the Shapes of DNA Toroids Condensed in Virus Capsids. *Biophysical Journal* 100:2209–2216. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3149266/>.
 26. Takahashi, M., K. Yoshikawa, V. V. Vasilevskaya, and A. R. Khokhlov, 1997. Discrete CoilGlobule Transition of Single Duplex DNAs Induced by Polyamines. *The Journal of Physical Chemistry B* 101:9396–9401. <https://doi.org/10.1021/jp9716391>, publisher: American Chemical Society.
 27. Fang, Y., and J. H. Hoh, 1998. Early Intermediates in Spermidine-Induced DNA Condensation on the Surface of Mica. *Journal of the American Chemical Society* 120:8903–8909. <https://doi.org/10.1021/ja981332v>.
 28. Marx, K. A., and G. C. Ruben, 1983. Evidence for hydrated spermidine-calf thymus DNA toruses organized by circumferential DNA wrapping. *Nucleic Acids Research* 11:1839–1854.
 29. Murayama, Y., Y. Sakamaki, and M. Sano, 2003. Elastic response of single DNA molecules exhibits a reentrant collapsing transition. *Physical Review Letters* 90:018102.
 30. Brewer, L. R., M. Corzett, and R. Balhorn, 1999. Protamine-induced condensation and decondensation of the same DNA molecule. *Science (New York, N.Y.)* 286:120–123.
 31. Smith, A. D., O. A. Ukogu, L. M. Devenica, E. D. White, and A. R. Carter, 2017. Optical methods for measuring DNA folding. *Modern Physics Letters B* 31:1730001. <https://www.worldscientific.com/doi/abs/10.1142/S0217984917300010>.
 32. Hud, N. V., and I. D. Vilfan, 2005. Toroidal DNA condensates: unraveling the fine structure and the role of nucleation in determining size. *Annual Review of Biophysics and Biomolecular Structure* 34:295–318.
 33. Brewer, L. R., 2011. Deciphering the structure of DNA toroids. *Integrative Biology: Quantitative Biosciences from Nano to Macro* 3:540–547.
 34. Hud, N. V., M. J. Allen, K. H. Downing, J. Lee, and R. Balhorn, 1993. Identification of the elemental packing unit of DNA in mammalian sperm cells by atomic force microscopy. *Biochemical and Biophysical Research Communications* 193:1347–1354.
 35. Binnig, G., C. F. Quate, and C. Gerber, 1986. Atomic Force Microscope. *Physical Review Letters* 56:930–933. <https://link.aps.org/doi/10.1103/PhysRevLett.56.930>.
 36. Lorenz, T. C., 2012. Polymerase Chain Reaction: Basic Protocol Plus Troubleshooting and Optimization Strategies. *Journal of Visualized Experiments : JoVE* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4846334/>.
 37. Nečas, D., and P. Klapetek, 2012. Gwyddion: an open-source software for SPM data analysis. *Central European Journal of Physics* 10:181–188. <https://doi.org/10.2478/s11534-011-0096-2>.
 38. Savransky, D., 2020. Simple Rejection Sampling. <https://www.mathworks.com/matlabcentral/fileexchange/27590-simple-rejection-sampling>.
 39. Dobrow, R. P., 2014. Probability with applications and R. John Wiley & Sons, Inc, Hoboken, New Jersey. OCLC: ocn755700207.

McMillan *et al.*

40. Devenica, L. M., C. Contee, R. Cabrejo, M. Kurek, E. F. Deveney, and A. R. Carter, 2016. Biophysical measurements of cells, microtubules, and DNA with an atomic force microscope. *American Journal of Physics* 84:301–310. <https://aapt.scitacion.org/doi/10.1119/1.4941048>.
41. Manning, G. S., 2006. The Persistence Length of DNA Is Reached from the Persistence Length of Its Null Isomer through an Internal Electrostatic Stretching Force. *Biophysical Journal* 91:3607–3616. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1630458/>.
42. Liverpool, T. B., and S. F. Edwards, 1995. Probability distribution of wormlike polymer loops. *The Journal of Chemical Physics* 103:6716–6719. <https://aip.scitacion.org/doi/abs/10.1063/1.470350>, publisher: American Institute of Physics.
43. Shen, M. R., K. H. Downing, R. Balhorn, and N. V. Hud, 2000. Nucleation of DNA Condensation by Static Loops: Formation of DNA Toroids with Reduced Dimensions. *Journal of the American Chemical Society* 122:4833–4834. <https://doi.org/10.1021/ja000200a>.
44. McMillan, R. B., H. A. Bediako, L. M. Devenica, Y. E. Ma, D. M. Roscoe, and A. R. Carter. DNA toroids form via a flower intermediate. *In preparation*.
45. Li, Y., U. H. Yildiz, K. Müllen, and F. Gröhn, 2009. Association of DNA with Multivalent Organic Counterions: From Flowers to Rods and Toroids. *Biomacromolecules* 10:530–540. <https://doi.org/10.1021/bm8011852>.
46. Mitchell, J. S., J. Glowacki, A. E. Grandchamp, R. S. Manning, and J. H. Maddocks, 2017. Sequence-Dependent Persistence Lengths of DNA. *Journal of Chemical Theory and Computation* 13:1539–1555. <https://doi.org/10.1021/acs.jctc.6b00904>, publisher: American Chemical Society.
47. van den Broek, B., M. C. Noom, J. van Mameren, C. Battle, F. C. MacKintosh, and G. J. Wuite, 2010. Visualizing the Formation and Collapse of DNA Toroids. *Biophysical Journal* 98:1902–1910. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2862208/>.
48. Hong, F., F. Zhang, Y. Liu, and H. Yan, 2017. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chemical Reviews* 117:12584–12640. <https://doi.org/10.1021/acs.chemrev.6b00825>.
49. Rothmund, P. W. K., 2006. Folding DNA to create nanoscale shapes and patterns. *Nature* 440:297. <https://www.nature.com/articles/nature04586>.
50. Gowetski, D. B., E. J. Kodis, and J. D. Kahn, 2013. Rationally designed coiled-coil DNA looping peptides control DNA topology. *Nucleic Acids Research* 41:8253–8265. <https://doi.org/10.1093/nar/gkt553>.
51. Manning, G. S., 2007. Counterion Condensation on Charged Spheres, Cylinders, and Planes [†]. *The Journal of Physical Chemistry B* 111:8554–8559. <https://pubs.acs.org/doi/10.1021/jp0670844>.
52. Archer, D. G., and P. Wang, 1990. The Dielectric Constant of Water and Debye-Hückel Limiting Law Slopes. *Journal of Physical and Chemical Reference Data* 19:371–411. <https://aip.scitacion.org/doi/10.1063/1.555853>, publisher: American Institute of Physics.
53. Peters, J. P., and L. J. Maher, 2010. DNA curvature and flexibility in vitro and in vivo. *Quarterly reviews of biophysics* 43:23–63. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4190679/>.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at <http://www.biophysj.org>. MATLAB code is available at Github at <http://doi.org/10.5281/zenodo.4321605>.