Biophysical Journal
Article

Biophysical Society

DNA looping by protamine follows a nonuniform spatial distribution

Ryan B. McMillan,[1] Victoria D. Kuntz,[1] Luka M. Devenica,[1] Hilary Bediako,[1] and Ashley R. Carter[1,*]
[1]Department of Physics, Amherst College, Amherst, Massachusetts

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 proteins such as condensin. Here, however, we are interested in a different looping method whereby condensing agents (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 (≤ 1 μm) DNA fragments using atomic force microscopy. 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 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 condensing agents 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 condensing agents (with charge $\geq +3$), which condense DNA into loops and toroids. One such condensing agent is the protein protamine, which condenses DNA in sperm. We investigated the mechanism for loop formation by protamine and found that the experimental data were 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 condensing agents 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). 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, prokaryotic transcription factors like *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 (5–8,17,18). In this method, proteins like condensin hydrolyze ATP to unidirectionally translocate along DNA, enlarging a loop as they move (5,8).

Submitted December 16, 2020, and accepted for publication April 19, 2021.
*Correspondence: acarter@amherst.edu
Editor: Wilma Olson.
https://doi.org/10.1016/j.bpj.2021.04.022

In this study, we focus on the less-well-understood looping mechanism of multivalent cations and other large molecules that condense DNA (19,20). Some common examples of DNA condensing agents are cobalt (III) hexaammine (21,22), spermine (23–25), spermidine (25–28), and protamine (16,29,30), although any molecule with a charge of at least +3 is thought to function similarly (19), and some divalent cations have also been shown to condense DNA under certain conditions (19). DNA condensing agents are known to bind to DNA nonspecifically (11,19) and form loops (16,26,31). This loop formation occurs as part of a pathway to form DNA toroids (19,32,33). Toroid sizes vary, but toroids generally have an outer diameter of ~100 nm (29,32) and can contain up to 50 kbp of DNA (30) in concentrically wound, hexagonally packed loops (21). 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 that are found in sperm cells (11) that bind and neutralize the negatively charged DNA (34). To fold the DNA into a loop, a previous mechanism suggested that protamine stabilizes spontaneous loops (15,32). Specifically, a thermal fluctuation in the DNA would create a spontaneous loop and protamine would stabilize this loop by anchoring the DNA contacts at the point where the DNA crosses over itself. Recent evidence, however, suggests that protamine instead forms loops via a bind-and-bend mechanism (16). In this mechanism, protamine takes a more active role, binding and bending the DNA by a small amount (~20° per protamine) to create the loop. Still, even in the bind-and-bend mechanism, protamine would also serve to anchor the DNA contacts at the DNA crossover point.

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 the conformations of short-length (≤ 1 μm) DNA fragments with and without protamine (16,35). We then compared the experimentally observed spatial distribution of loops to the predictions of three models: 1) a random looping model that assumes that loop formation is unbiased and spontaneous, 2) an electrostatic binding model that assumes that loop formation occurs when electrostatic interactions cause a single protamine to bind to the DNA, and 3) 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 was able to describe the spatial distribution of loops formed without protamine, but not the distribution of protamine-induced loops. In particular, the distribution of protamine-induced loops had a peak of loops forming at about a quarter of the way along the DNA (fractional DNA length of 0.1–0.3). This means that loop formation is biased rather than uniform. The electrostatic multibinding model was able to predict the location of this peak. Thus, our data are consistent with a looping mechanism for protamine that mimics the bind-and-bend mechanism, but not with a spontaneous looping mechanism that assumes that loop initiation is 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 polymerase chain reaction (PCR) (36). Specifically, we used bacteriophage lambda DNA (N3011; New England Biolabs, Ipswich, MA) as a template, custom oligonucleotide primers (Integrated DNA Technologies, Coralville, IA), and an LA *Taq* DNA polymerase (RR004; TaKaRa Bio, Kusatsu, Japan). We verified that products had amplified correctly using gel electrophoresis and then extracted the DNA using a commercial kit (QIAquick Gel Extraction Kit, 28704; Qiagen, Hilden, Germany). Finally, we measured the concentration and purity using a spectrophotometer (NanoDrop Lite; Thermo Fisher Scientific, Waltham, MA). Samples with A260/A280 ratios of less than 1.7 were discarded.

We chose the longest DNA length (1023 nm) for spontaneous looping experiments because looping is more likely to occur for DNA lengths, L_C (contour length or length along the DNA contour), that are much greater than the DNA persistence length, L_P. The L_P of the DNA (50 nm (37,38)) is the length over which the tangent vector to the DNA remains correlated (39) 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 because of 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. We therefore used 1023-nm-length molecules in measurements for which we wanted to measure spontaneous looping, and we used 217- and 398-nm-length molecules in measurements for which we wanted to measure protamine-induced looping.

We purchased protamine from salmon (P4005; Sigma-Aldrich, Saint Louis, MO), diluted it in deionized water, and stored 30 μM aliquots at −20°C. We check for protamine monodispersion using an AFM to image protamine directly adhered to the mica surface at a concentration of 2 μM. We see single molecules of protamine. Occasionally, we see larger aggregates of protamine. These aggregates bind to multiple DNA molecules and are not studied.

Preparing AFM slides

AFM slides were prepared by adhering DNA to a mica surface. DNA adhered to the surface in this way (40) or in the presence of protamine (29) is known to equilibrate on the surface; it is not kinetically trapped in 3D. Specifically, we affixed 10-mm-diameter ruby muscovite mica slides (grade V1; Ted Pella, Redding, CA) to metallic disks. Then, to create a clean surface, we used tape to remove the top layer of the mica. Next, we created the DNA solution using a procedure to reduce DNA aggregation in solution (16). Specifically, we prepared 20 μL solutions of 0.2 ng/μL DNA, 2.0 mM magnesium acetate, and protamine concentrations of either 0 μM (for samples without protamine) or 0.2–5.0 μM (for samples with

protamine). 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. Waiting longer times produced more DNA toroids and caused more DNA aggregation because protamine binding is rather fast (the on rate for bull protamine P1 is ~2000 molecules/s · μm (41)). We then repeated this procedure until there were a total of three to five depositions on the mica. All data shown in the work are taken from samples with multiple depositions.

However, as a control we also prepared samples with a 20 μL solution without protamine that contained 1.0 ng/μL DNA and 2.0 mM magnesium acetate. After pipetting this solution onto the surface, we waited ~30 s and then washed the mica with 1 mL of deionized water and dried with nitrogen. Multiple depositions did not change the height of the DNA (~0.5 nm consistent with previous data (42)), the length of the DNA (200 ± 40 nm (mean ± standard deviation), 400 ± 30 nm, and 1000 ± 200 nm for nominal DNA lengths of 217, 398, and 1023 nm), or the persistence length of the DNA (50 nm, consistent with previous measurements (40)). We thus conclude that the DNA in our multideposition samples is equilibrating on the surface as is seen in single deposition experiments (40). 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 14 Hz. The image size was either 2 μm × 2 μm (512 × 512 pixels) or 1 μm × 1 μm (256 × 256 pixels).

Analyzing AFM slides

Image processing of AFM slides was done using Gwyddion (43). Images were corrected using three steps. First, we aligned rows using a fifth-degree polynomial. Second, we removed high-frequency oscillations using an fast Fourier transform (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 one 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 (Figs. S1 and S2). Images were 200 × 200 nm for 217-nm-length DNA, 400 × 400 nm for 398-nm-length DNA, and 400 × 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

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 γ distribution of the experimentally observed loop circumferences (see Fig. S5 and "Modeling loop formation using gamma distributions," Supporting materials and methods).
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

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

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

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

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

$$f(s_i) = \frac{\exp\left(-\frac{qV}{k_BT}\right)}{Z}, \tag{3}$$

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

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

We can rewrite this equation as

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

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

$$\theta = \frac{qk\lambda}{k_BT} \tag{6}$$

In the limit of very high charges, $\theta \rightarrow \infty$, whereas in the limit of low charges, $\theta \rightarrow 0$. At $\theta = 1$, the electrostatic potential 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 (44).
2) Use steps 2 and 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 maximal 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 (Fig. S7).
5) Use steps 2 and 3 of the random looping model to generate s_s.

Simulation parameters

We wrote functions to simulate the random looping model, electrostatic binding model, and electrostatic multibinding model in MATLAB (The MathWorks, Natick, MA).

All simulations discretized the DNA into segments of length 1 nm, and this was inputted into the program as the number of segments (e.g., 217-nm-length DNA was input as 217 segments). To define the loop circumference C in the simulations, we inputted the fit parameters from a γ distribution that defines the experimentally determined distribution of loop circumferences. These parameters are α (which was always set to 2), β, and d_0 (see Fig. S5 and "Modeling loop formation using gamma distributions," Supporting materials and methods). All simulations input the number of s_s-values to calculate as the number of trials (100,000 for simulation in the main text and 10,000 for simulations in the Supporting materials and methods). Finally, we input the number of loops to simulate—either 1 for modeling single loops or 2 and 3 for modeling flowers with two and three loops.

For the electrostatic binding model and the electrostatic multibinding model, we also input the parameter θ given by Eq. 6 and the number of protamines n (one for the electrostatic binding model and more than one for the electrostatic multibinding model). In the electrostatic multibinding model, we also input the maximal distance m that the multiple protamines could bind from each other. These three parameters (θ, n, and m) are optimized to fit each data set. We display how the simulation changes with each parameter in Figs. S8–S11. Specific parameters for each simulation are listed in Table S1.

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 three pixels, or 11.7 nm, which is ~4× 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 minus 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 (45). 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

$$\text{Error} = \frac{N_{bin}^{1/2}}{N} \tag{7}$$

RESULTS

Spontaneous loops are fitted 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. Because the mechanism of spontaneous loop formation is random, we hypothesize that spontaneous loop formation should follow a model in which loops are equally likely to initiate at any point along the DNA. We then measured the spatial distribution of spontaneous loops and checked to see whether the spatial distribution was described by our model.

To produce spontaneous loops, we immobilized long DNA molecules (contour length, $L_C = 1023$ nm) on a mica surface in the absence of folding agents (Fig. 1 *A*). As the molecules equilibriated on the surface, random thermal bending of the DNA created spontaneous loops. We then imaged the DNA on the surface with an AFM (Fig. 1 *B*), 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. We also measured the diameter d for each loop and found a peak at 33 nm (Fig. S5), meaning that the peak circumference C is 103 nm (303 bp). Interestingly, this value is above 150 bp. Previous experiments on DNA cyclization show that the efficiency of DNA cyclization decreases below 150 bp (46–48).

To generate simulated loops using the random looping model, we randomly chose sites along the DNA to initiate

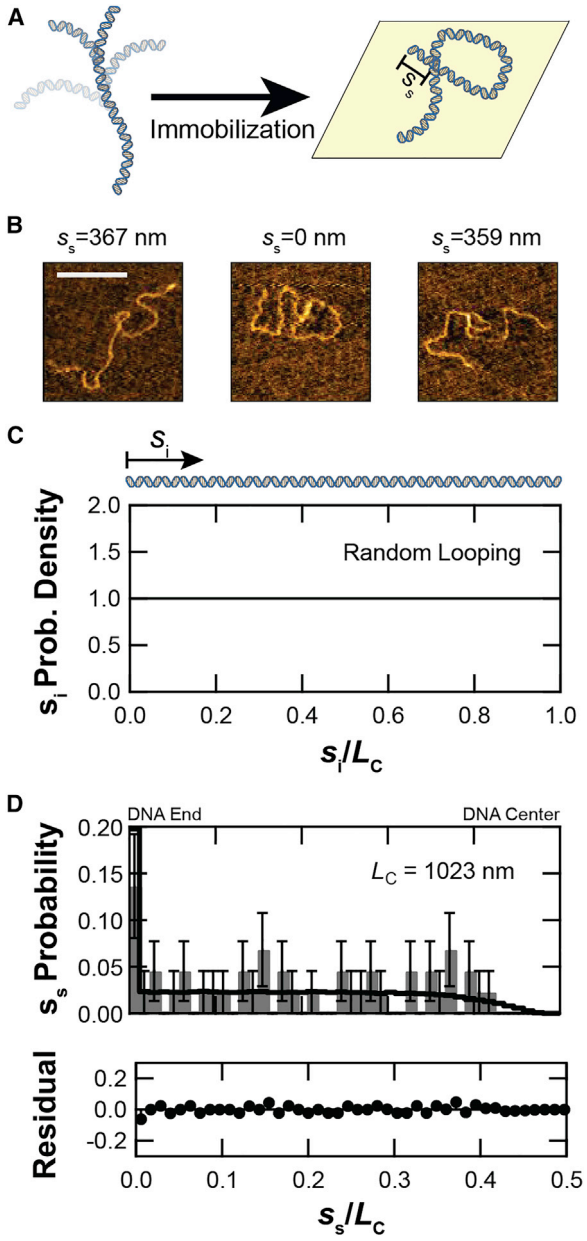

FIGURE 1 Spontaneous loops form according to a random looping model. (*A*) To measure spontaneous loop formation, we immobilized DNA of contour length L_C = 1023 nm on the surface of an AFM slide without protamine. We determined the start site $s_s \in [0, L_C/2]$ as the arc length from the closest DNA end to the DNA crossover location. (*B*) Three sample loops with their measured start sites are shown. Scale bars, 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 $\leq$0.06 for all subsequent bins up to the predicted falloff at $s_s/L_C \approx 0.45$. Error bars were calculated according to Eq. 7. To see this figure in color, go online.

a loop and calculated s_s (more information in the Materials and methods). Specifically, the program first chose a loop initiation site s_i using a uniform probability distribution (Fig. 1 *C*). 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 0. Finally, we plotted the simulated s_s/L_C, along with the measured s_s/L_C, in a histogram and calculated the residuals (height of the bin for the measurement minus the height of the bin for the simulation) between the two data sets (Fig. 1 *D*).

The two distributions are very similar. To see this clearly, notice that the height of the bin for the measurement is the percentage of the total molecules ($N = 44$) that fell in that bin. The average measurement error was 0.018 $\pm$ 0.002 (Table S2), which corresponds to an error that is less than one molecule per bin (44 $\times$ 0.018 = 0.79 molecules per bin). Now notice that the residuals had an average of 0.016 $\pm$ 0.002 (mean $\pm$ standard error), meaning that there was also less than one molecule per bin (44 $\times$ 0.016 = 0.70 molecules per bin) difference. This indicates 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 of the DNA ($s_s/L_C = 0$) and center of the DNA ($s_s/L_C = 0$).

We can be more quantitative about what happens at the DNA ends and the center of the DNA. At the DNA end, in the first bin of the distribution, there is an increased probability of loops (bin height of 0.14 $\pm$ 0.06 for s_s/L_C = 0–0.011) that is predicted by the simulation. 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. Because of this end effect, the first bin is higher than the second by a factor of

$$1 + \frac{C}{2L_C(\text{binwidth}/L_C)}, \tag{8}$$

where binwidth is the width of the bin in nanometers. Given that the peak C-value is 103 nm (Fig. S5), this corresponds to a factor of ~5 for the simulated distribution, which matches the experimentally measured factor of 6 $\pm$ 3. Interestingly, if binwidth is normalized by L_C, then 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 fractional DNA length of

$$0.5 - \frac{C}{2L_C} \tag{9}$$

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 a fractional DNA location of 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 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. Previously, protamine was thought to stabilize spontaneous DNA loops (15,32). If this model is correct, then we might expect the random looping model to describe protamine-induced loops as well as spontaneous loops. However, we found recently 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. If this model is correct, then we would not expect the random looping model to fit the data.

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 these experimental data to the simulated start sites produced by the random looping model (Fig. 2). To produce protamine-induced loops, 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. 2 *A*). The shorter DNA lengths (217–398 nm) are used in this experiment because 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. 2 *B*) and d (Fig. S5). In general, protamine-induced loops were smaller than spontaneous loops, having peak diameters of 23 nm for the 217-nm-length DNA. This loop size is comparable to loops created by bacterial transcription factors (49), which can be as small as 14 nm in diameter (50). To create simulated data, we ran our simulation for the random loop model again, but this time for a DNA length of 217 nm.

This time, there is poor agreement between the experimental data and the simulated data. A histogram of the fractional start site s_s/L_C (Fig. 2 *C*) showed that the average residual between the experimental data and the random looping model was 0.07 ± 0.03 and 0.04 ± 0.01 (for 217-nm-length and 398-nm-length DNA, respectively). This means that the simulation is off by ~five molecules per bin in the case of the data for the 217-nm-length DNA. This is almost two times higher than the average error per bin of 0.04 ± 0.01 and 0.028 ± 0.005, respectively (Table S2). The reason for this difference is that the random looping model does not predict the shape of the distribution. Between the fractional start sites of 0.1–0.3, which correspond 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. In addition, 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. This is not due to resolution limitations, as the binwidth for the graph is three pixels (or 11.7 nm) divided by L_C, and 11.7 nm is ~4 times the measurement error of 3 nm. The random looping model was

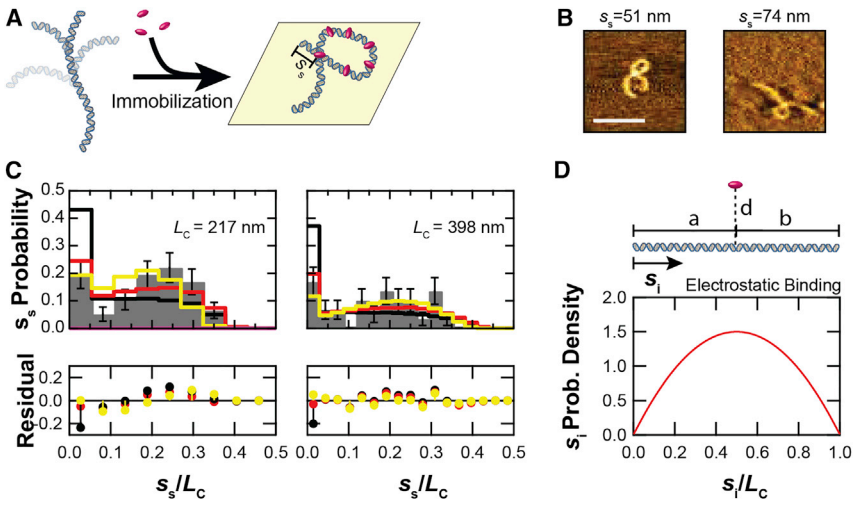

FIGURE 2 Single loops are best described by an electrostatic multibinding model. (*A*) Experimental setup for protamine-induced loops. DNA in solution is bent by protamine and then immobilized on a surface so that we can measure s_s. (*B*) Sample AFM images of loops in 217-nm-length DNA are shown. Scale bars, 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 minus 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, $d << a, b$. To see this figure in color, go online.

able to predict the end behavior of the distribution because this is just due to loop geometry. The measured falloff was 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 (~50%) is decreased in this region as compared to the rest of the DNA (~70%), suggesting that sequence-dependent effects might be responsible for the decreased looping probability (see Fig. S12 and "Local DNA sequence variations," Supporting materials and methods). The random looping model does not account for local variations in DNA flexibility and curvature because it assumes that the DNA is uniformly flexible along its length. Although DNA sequence is known to affect looping in general (51,52), 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 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 (19,20), 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. In this model, we assume that the protamine is a positive point charge and that the DNA is a finite, negative line charge. We then calculate the electrostatic potential energy for protamine binding given this assumption. This electrostatic potential energy biases loop formation away from the DNA end and could be the physical effect that creates the peak in the experimental spatial distribution a quarter of the way along the DNA.

To simulate start sites using the electrostatic binding model, we select the loop initiation site s_i, assuming that the probability density for the loop initiation site s_i follows an inverse parabolic distribution (Fig. 2 *D*). This probability density is the inverse of the electrostatic potential energy. The height of the probability density, 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, increasing θ shifts the s_s/L_C distribution toward the center of the DNA (Fig. S8). The optimal value for θ was 1, 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," Supporting materials and methods).

Once we determined the loop initiation sites, we calculated the loop start sites s_s as before. We then plotted the spatial distribution of the simulated loop start sites for the electrostatic binding model (Fig. 2 *C*). 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.

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 density that was developed for the electrostatic binding model (Fig. 2 *D*). The height of this probability density is determined by the parameter θ. To make sure multiple protamines do not bind in the same location, the probability density is updated to account for any bound protamines (Fig. S6). In addition, we defined a particular distance m that protamines must bind within to create a DNA loop. This seems reasonable, as molecules on opposite ends of the DNA would not be able to work together to bend the DNA. Finally, the model assumes that the binding location of the outermost protamine is the initiation site (Fig. S7).

To produce simulated start sites with the electrostatic multibinding model, we ran the simulation under different conditions for the three parameters (n, θ, m). We then plotted

the spatial distributions of the simulated loop start sites (Figs. S9–S11) and optimized the parameters. We first optimized the parameter θ, which describes the strength of the electrostatic effect relative to thermal noise (Fig. S9). Increasing θ pushes the start site distribution toward 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," Supporting materials and methods). We also examined the effect of varying the parameter for the number of protamines n (Fig. S10). Increasing n shifts the distribution toward the DNA end, presumably because of the higher likelihood that at least one protamine will bind away from the DNA center. We found that two "molecules" (which could really be two groups of molecules) is the best fit to the experimental data. Decreasing the parameter for the maximal distance between the molecules m (Fig. S11) shifts the distribution toward the center of the molecule. The simulation that produced the best results had $m = 50$ nm, which happens to be the persistence length of the DNA (38). This observation is interesting physically because it suggests that binding within a persistence length produces a loop. This would be the case if loop formation requires correlated bending, which might not occur over length scales much longer than a persistence length.

After selecting the model parameters ($\theta = 1.5$, $n = 2$ molecules, $m = 50$ nm), we then compared our simulated distribution with the experimental data (Fig. 2 *C*). 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 $\pm$ 0.06 (mean $\pm \pm$ standard deviation of Gaussian fit to data) and 0.2 $\pm$ 0.1, which agreed with the peaks in the experimental data of 0.23 $\pm$ 0.06 and 0.22 $\pm$ 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 that bias loop formation toward the center of the DNA, the coordinated binding of multiple protamines within a persistence length of each other that bias loop formation toward 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 toward the end of the DNA.

Protamine-induced flowers follow the electrostatic multibinding model

To create another test of the electrostatic multibinding model, we wondered whether our models of loop formation would generalize to the formation of DNA flowers (26,53). DNA flowers are multilooped DNA structures that form in the presence of protamine (53) or other folding agents (26,54) and look flower-like when immobilized on a surface, as all of the loops overlap each other at a central location. Flowers are thought to be an intermediate step in toroid formation (53) and likely form when multiple protamines bind and bend the DNA into several loops. If the initiation mechanism of the DNA flower is the same as that of the loop, then we might expect the electrostatic multibinding model to also describe the spatial distribution of DNA flowers.

To measure the experimental s_s distribution of DNA flowers, we immobilized DNA ($L_C = 398$ nm) on the surface in the presence of protamine (0.2–5 μM), as before. However, instead of selecting molecules that folded into a single loop, we selected molecules that folded into a flower. DNA flowers were more likely to form at the higher protamine concentrations (0.6–5.0 μM). We used an AFM to image two-looped ($N = 78$) and three-looped ($N = 42$) flowers (Fig. 3 *A*). 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 (Fig. 3, *B* and *C*). 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 then created spatial distributions of the start sites output by our three models: random looping, electrostatic binding, and electrostatic multibinding (Fig. 3, *B* and *C*). The parameters for the three models are listed in Table S1. Interestingly, we did not vary the parameters m, θ, and n from their previous values with single loops. The only change we made in implementing these simulations was to update the parameter for the number of loops.

Finally, we compared our experimental data to the three simulations: random looping model, electrostatic binding, and electrostatic multibinding (Fig. 3, *B* and *C*). For both two-loop and three-loop flowers, the electrostatic multibinding model predicted the height of the first bin within error, whereas the random looping model did not. The residuals for the electrostatic multibinding model were also both within the measurement error of the data set (Table S3). However, only the electrostatic multibinding model predicted the position of the peak in the experimental data (experimental peak was at 0.14 $\pm$ 0.03 and 0.12 $\pm$ 0.03 and the predicted peak was at 0.1 $\pm$ 0.1 and 0.1 $\pm$ 0.1 for two-loop and three-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 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

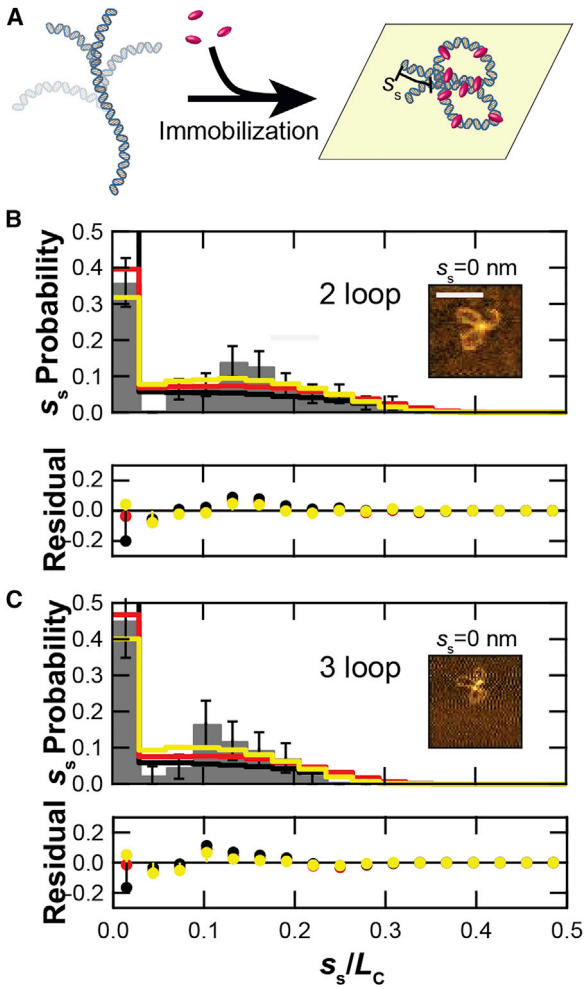

FIGURE 3 Flowers also follow an electrostatic multibinding model. (*A*) To study flower formation, we immobilized DNA of contour length L_C = 398 nm on the surface of an AFM slide in the presence of protamine. Now the start site $s_s \in [0, L_C/2]$ is the arc length to the flower central point. (*B* and *C*) We collected and plotted the probability at each fractional start site for two-loop (*B*) and three-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: two-loop and three-loop flowers extracted from AFM images. Scale bars, 100 nm. To see this figure in color, go online.

DNA that had formed spontaneous loops, protamine-induced loops, and protamine-induced flowers. We combined these 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, but not protamine-induced loops or flowers. The distributions of protamine-induced loops and flowers are peaked at a fractional DNA length of 0.1–0.3, which is about a quarter of the way along the DNA. This peak indicates that the spatial distribution is not uniform. Instead, it is biased. The electrostatic multibinding model predicts the location of this peak (Table S4), giving further evidence that protamine uses a bind-and-bend mechanism to loop the DNA instead of stabilizing a spontaneous 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 (51,55) and are thought to cause the lack of loop start sites in the 398-nm-length DNA at a bin value of s_s/L_C = 0.09–0.12 (Fig. 2 *C*; Fig. S12). DNA sequence variation can also lead to static curvature in the DNA. For example, a series of A-tracts of the form d(AAAAAA) can be phased with the helical spacing of the DNA to create a static curve (52). This DNA sequence variation generates smaller diameter toroids and, given our results with DNA sequence variation, would likely bias the binding or bending of condensing agents. Second, we assumed that the DNA is fairly stiff over its length (217–398 nm). Longer molecules would be floppier and have more spontaneous looping. Increased spontaneous looping would 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.

We hypothesize that the electrostatic multibinding model may shed some insight on the toroid formation process (32). First, because toroids rely on looping of the DNA to form and both loop formation and flower formation appear to follow a spatial distribution depicted by the electrostatic multibinding model, 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 (56) that found that toroids folded by spermine have two interactions/loop. If one molecule or group of molecules facilitates one interaction, then these data would be consistent with the electrostatic multibinding model.

We also note that our results indicate that the charge of the molecule, and the electrostatic interaction in particular, is the driver of loop formation. This points toward a general mechanism of looping by condensing agents, including spermine, cobalt (III) hexaammine, and others (19,20,32). Future work could investigate how binding probability, charge, or concentration for different condensing agents 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 protein in bacteria (10,13). Histone-like nucleoid-structuring protein 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 (57,58). Synthetic looping proteins have been used previously to form specific DNA contacts (59) that have aided in the assembly of DNA nanostructures (14). Here, we speculate that condensing agents with nonspecific contacts could also be used to bend, condense, or stabilize engineered DNA constructs.

MATLAB code is available at Github at http://doi.org/10.5281/zenodo.4321605 http://doi.org/10.5281/zenodo.4321604.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2021.04.022.

AUTHOR CONTRIBUTIONS

R.B.M. and A.R.C. designed the research, took the data, ran the analysis, and wrote the article. V.D.K. analyzed the data. L.M.D. took the data and analyzed the data. H.B. took the data and assisted in writing the article.

ACKNOWLEDGMENTS

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

SUPPORTING CITATIONS

References (60–62) appear in the Supporting material.

REFERENCES

1. Schleif, R. 1992. DNA looping. *Annu. Rev. Biochem.* 61:199–223.
2. Matthews, K. S. 1992. DNA looping. *Microbiol. Rev.* 56:123–136.
3. 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. *Proc. Natl. Acad. Sci. USA.* 83:3654–3658.
4. Cournac, A., and J. Plumbridge. 2013. DNA looping in prokaryotes: experimental and theoretical approaches. *J. Bacteriol.* 195:1109–1119.
5. van Ruiten, M. S., and B. D. Rowland. 2018. SMC complexes: universal DNA looping machines with distinct regulators. *Trends Genet.* 34:477–487.
6. Hassler, M., I. A. Shaltiel, and C. H. Haering. 2018. Towards a unified model of SMC complex function. *Curr. Biol.* 28:R1266–R1281.
7. Yuen, K. C., and J. L. Gerton. 2018. Taking cohesin and condensin in context. *PLoS Genet.* 14:e1007118.
8. Terakawa, T., S. Bisht, …, E. C. Greene. 2017. The condensin complex is a mechanochemical motor that translocates along DNA. *Science.* 358:672–676.
9. Krepel, D., R. R. Cheng, …, J. N. Onuchic. 2018. Deciphering the structure of the condensin protein complex. *Proc. Natl. Acad. Sci. USA.* 115:11911–11916.
10. Rimsky, S. 2004. Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. *Curr. Opin. Microbiol.* 7:109–114.
11. Balhorn, R. 2007. The protamine family of sperm nuclear proteins. *Genome Biol.* 8:227.
12. 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.
13. Dame, R. T., M. S. Luijsterburg, …, G. J. L. Wuite. 2005. DNA bridging: a property shared among H-NS-like proteins. *J. Bacteriol.* 187:1845–1848.
14. Praetorius, F., and H. Dietz. 2017. Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes. *Science.* 355:eaam5488.
15. Hud, N. V., K. H. Downing, and R. Balhorn. 1995. A constant radius of curvature model for the organization of DNA in toroidal condensates. *Proc. Natl. Acad. Sci. USA.* 92:3581–3585.
16. Ukogu, O. A., A. D. Smith, …, A. R. Carter. 2020. Protamine loops DNA in multiple steps. *Nucleic Acids Res.* 48:6108–6119.
17. Ganji, M., I. A. Shaltiel, …, C. Dekker. 2018. Real-time imaging of DNA loop extrusion by condensin. *Science.* 360:102–105.
18. Davidson, I. F., D. Goetz, …, J. M. Peters. 2016. Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J.* 35:2671–2685.
19. Bloomfield, V. A. 1997. DNA condensation by multivalent cations. *Biopolymers.* 44:269–282.
20. Teif, V. B., and K. Bohinc. 2011. Condensed DNA: condensing the concepts. *Prog. Biophys. Mol. Biol.* 105:208–222.
21. Hud, N. V., and K. H. Downing. 2001. Cryoelectron microscopy of lambda phage DNA condensates in vitreous ice: the fine structure of DNA toroids. *Proc. Natl. Acad. Sci. USA.* 98:14925–14930.
22. Schnell, J. R., J. Berman, and V. A. Bloomfield. 1998. Insertion of telomere repeat sequence decreases plasmid DNA condensation by cobalt (III) hexaammine. *Biophys. J.* 74:1484–1491.
23. Leforestier, A., and F. Livolant. 2009. Structure of toroidal DNA collapsed inside the phage capsid. *Proc. Natl. Acad. Sci. USA.* 106:9157–9162.
24. Leforestier, A., A. Šiber, …, R. Podgornik. 2011. Protein-DNA interactions determine the shapes of DNA toroids condensed in virus capsids. *Biophys. J.* 100:2209–2216.
25. Takahashi, M., K. Yoshikawa, …, A. R. Khokhlov. 1997. Discrete coil globule transition of single duplex DNAs induced by polyamines. *J. Phys. Chem. B.* 101:9396–9401.
26. Fang, Y., and J. H. Hoh. 1998. Early intermediates in spermidine-induced DNA condensation on the surface of mica. *J. Am. Chem. Soc.* 120:8903–8909.
27. Marx, K. A., and G. C. Ruben. 1983. Evidence for hydrated spermidine-calf thymus DNA toruses organized by circumferential DNA wrapping. *Nucleic Acids Res.* 11:1839–1854.
28. Murayama, Y., Y. Sakamaki, and M. Sano. 2003. Elastic response of single DNA molecules exhibits a reentrant collapsing transition. *Phys. Rev. Lett.* 90:018102.
29. Allen, M. J., E. M. Bradbury, and R. Balhorn. 1997. AFM analysis of DNA-protamine complexes bound to mica. *Nucleic Acids Res.* 25:2221–2226.
30. Brewer, L. R., M. Corzett, and R. Balhorn. 1999. Protamine-induced condensation and decondensation of the same DNA molecule. *Science.* 286:120–123.

31. Smith, A. D., O. A. Ukogu, …, A. R. Carter. 2017. Optical methods for measuring DNA folding. *Mod. Phys. Lett. B.* 31:1730001.
32. Hud, N. V., and I. D. Vilfan. 2005. Toroidal DNA condensates: unraveling the fine structure and the role of nucleation in determining size. *Annu. Rev. Biophys. Biomol. Struct.* 34:295–318.
33. Brewer, L. R. 2011. Deciphering the structure of DNA toroids. *Integr. Biol.* 3:540–547.
34. Hud, N. V., M. J. Allen, …, R. Balhorn. 1993. Identification of the elemental packing unit of DNA in mammalian sperm cells by atomic force microscopy. *Biochem. Biophys. Res. Commun.* 193:1347–1354.
35. Binnig, G., C. F. Quate, and C. Gerber. 1986. Atomic force microscope. *Phys. Rev. Lett.* 56:930–933.
36. Lorenz, T. C. 2012. Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *J. Vis. Exp.* 63:e3998.
37. Devenica, L. M., C. Contee, …, A. R. Carter. 2016. Biophysical measurements of cells, microtubules, and DNA with an atomic force microscope. *Am. J. Phys.* 84:301–310.
38. 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. *Biophys. J.* 91:3607–3616.
39. Liverpool, T. B., and S. F. Edwards. 1995. Probability distribution of wormlike polymer loops. *J. Chem. Phys.* 103:6716–6719.
40. Rivetti, C., M. Guthold, and C. Bustamante. 1996. Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.* 264:919–932.
41. Brewer, L., M. Corzett, …, R. Balhorn. 2003. Dynamics of protamine 1 binding to single DNA molecules. *J. Biol. Chem.* 278:42403–42408.
42. Hansma, H. G., I. Revenko, …, D. E. Laney. 1996. Atomic force microscopy of long and short double-stranded, single-stranded and triple-stranded nucleic acids. *Nucleic Acids Res.* 24:713–720.
43. Nečas, D., and P. Klapetek. 2012. Gwyddion: an open-source software for SPM data analysis. *Cent. Eur. J. Phys.* 10:181–188.
44. Savransky, D. 2020. Simple rejection sampling https://www.mathworks.com/matlabcentral/fileexchange/27590-simple-rejection-sampling.
45. Dobrow, R. P. 2014. Probability with Applications and R. John Wiley & Sons, Inc, Hoboken, NJ.
46. Geggier, S., and A. Vologodskii. 2010. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. USA.* 107:15421–15426.
47. Crothers, D. M., J. Drak, …, S. D. Levene. 1992. DNA bending, flexibility, and helical repeat by cyclization kinetics. *Methods Enzymol.* 212:3–29.
48. Shore, D., J. Langowski, and R. L. Baldwin. 1981. DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. USA.* 78:4833–4837.
49. Johnson, S., Y.-J. Chen, and R. Phillips. 2013. Poly(dA:dT)-rich DNAs are highly flexible in the context of DNA looping. *PLoS One.* 8:e75799.
50. Becker, N. A., J. P. Peters, …, T. A. Lionberger. 2013. Mechanism of promoter repression by Lac repressor-DNA loops. *Nucleic Acids Res.* 41:156–166.
51. Shen, M. R., K. H. Downing, …, N. V. Hud. 2000. Nucleation of DNA condensation by static loops: formation of DNA toroids with reduced dimensions. *J. Am. Chem. Soc.* 122:4833–4834.
52. Conwell, C. C., I. D. Vilfan, and N. V. Hud. 2003. Controlling the size of nanoscale toroidal DNA condensates with static curvature and ionic strength. *Proc. Natl. Acad. Sci. USA.* 100:9296–9301.
53. McMillan, R. B., H. A. Bediako, …, A. R. Carter. 2021. DNA toroids form via a flower intermediate. *Biophys. J.* 120 (*Suppl 1*):34A.
54. Li, Y., U. H. Yildiz, …, F. Gröhn. 2009. Association of DNA with multivalent organic counterions: from flowers to rods and toroids. *Biomacromolecules.* 10:530–540.
55. Mitchell, J. S., J. Glowacki, …, J. H. Maddocks. 2017. Sequence-dependent persistence lengths of DNA. *J. Chem. Theory Comput.* 13:1539–1555.
56. van den Broek, B., M. C. Noom, …, G. J. Wuite. 2010. Visualizing the formation and collapse of DNA toroids. *Biophys. J.* 98:1902–1910.
57. Hong, F., F. Zhang, …, H. Yan. 2017. DNA origami: scaffolds for creating higher order structures. *Chem. Rev.* 117:12584–12640.
58. Rothemund, P. W. K. 2006. Folding DNA to create nanoscale shapes and patterns. *Nature.* 440:297–302.
59. Gowetski, D. B., E. J. Kodis, and J. D. Kahn. 2013. Rationally designed coiled-coil DNA looping peptides control DNA topology. *Nucleic Acids Res.* 41:8253–8265.
60. Manning, G. S. 2007. Counterion condensation on charged spheres, cylinders, and planes. *J. Phys. Chem. B.* 111:8554–8559.
61. Archer, D. G., and P. Wang. 1990. The dielectric constant of water and Debye-Hückel limiting law slopes. *J. Phys. Chem. Ref. Data.* 19:371–411.
62. Peters, J. P., III, and L. J. Maher. 2010. DNA curvature and flexibility in vitro and in vivo. *Q. Rev. Biophys.* 43:23–63.