Published online 19 July 2018

Nucleic Acids Research, 2018, Vol. 46, No. 18 ***9401–9413***
doi: 10.1093/nar/gky639

Sequence-dependent DNA condensation as a driving force of DNA phase separation

Hyunju Kang[1,†], Jejoong Yoo[2,3,†], Byeong-Kwon Sohn[1], Seung-Won Lee[1], Hong Soo Lee[1], Wenjie Ma[4], Jung-Min Kee[4], Aleksei Aksimentiev[2,5,*] and Hajin Kim[1,6,*]

[1]School of Life Sciences, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea, [2]Department of Physics and the Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, [3]Center for Self-assembly and Complexity, Institute for Basic Science, Pohang, Republic of Korea, [4]Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea, [5]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA and [6]Center for Genomic Integrity, Institute for Basic Science, Ulsan, Republic of Korea

Received February 17, 2018; Revised June 06, 2018; Editorial Decision July 02, 2018; Accepted July 04, 2018

ABSTRACT

The physical properties of DNA have been suggested to play a central role in spatio-temporal organization of eukaryotic chromosomes. Experimental correlations have been established between the local nucleotide content of DNA and the frequency of inter- and intra-chromosomal contacts but the underlying physical mechanism remains unknown. Here, we combine fluorescence resonance energy transfer (FRET) measurements, precipitation assays, and molecular dynamics simulations to characterize the effect of DNA nucleotide content, sequence, and methylation on inter-DNA association and its correlation with DNA looping. First, we show that the strength of DNA condensation mediated by polylysine peptides as a reduced model of histone tails depends on the DNA's global nucleotide content but also on the local nucleotide sequence, which turns out to be qualitatively same as the condensation by spermine. Next, we show that the presence and spatial arrangement of C5 methyl groups determines the strength of inter-DNA attraction, partially explaining why RNA resists condensation. Interestingly, multicolor single molecule FRET measurements reveal strong anti-correlation between DNA looping and DNA–DNA association, suggesting that a common biophysical mechanism underlies them. We propose that the differential affinity between DNA regions of varying sequence pattern may drive the phase separation of chromatin into chromosomal subdomains.

INTRODUCTION

The spatio-temporal organization of the chromosomes is central to regulation of gene expression and genome maintenance in eukaryotic cells (1–4). A necessary step toward understanding the organization of chromosomes at the molecular level is characterization of the physical properties of bare DNA, including the effects of nucleotide sequence and chemical modifications. Indeed, the physical properties of DNA have long been implicated in chromosome organization, from DNA flexibility governing nucleosome positioning (5) to DNA–DNA interaction influencing the structure of chromatin at genome scales (6). Being negatively charged, DNA molecules repel one another in water but condense in the presence of polycations such as polyamines or lysine- or arginine-rich peptides (7–13). At the same time, these biogenic polycations are known to participate in regulatory mechanisms that affect chromatin condensation, DNA replication, transcription, and translation (14).

Much of what is known about the polycation-induced DNA–DNA interactions has come from *in vitro* experiments that examined DNA condensation through osmotic stress-induced compaction (10,15,16), X-ray scattering (17–20), precipitation (9), magnetic tweezers (16), or single-molecule fluorescence resonance energy transfer (smFRET) experiments (6). Those experiments revealed the universal hexagonal packing pattern regardless of the type of condensing agents and the energetics of the packing as a function of packing density. However, the spatial and temporal limitations of the experimental approaches precluded determination of the exact location of polycations within the DNA condensates and hence the molecular mechanism that gives rise to the DNA condensation phenomenon. Recent theoretical and computational studies suggest that DNA

*To whom correspondence should be addressed. Tel: +82 52 217 2557; Email: hajinkim@unist.ac.kr
Correspondence may also be addressed to Aleksei Aksimentiev. Email: aksiment@illinois.edu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

condensation originates from specific placement of polycations between the DNA molecules (6,21), which is commonly referred to as the bridging or charge–density wave model (22,23).

The bridging model has been successful in predicting the sequence- and chemical modification-dependent strength of DNA condensation. Specifically, it was shown that the strength of attractive interaction between two DNA molecules mediated by spermine increases with the TA content of the DNA molecule and with the methylation of cytosine nucleotides (6). An independent set of experiments has found DNA molecules to be more amenable to condensation by polycations than double-stranded RNA molecules of similar length (24,25). DNA condensation can also be controlled by DNA flexibility—a key determinant of the chromatin architecture (5,26,27). For example, long DNA molecules have been shown to form toroidal structures in the presence of polycations and the DNA looping to control nucleation of such toroidal structures (28). Direct observation of real-time DNA dynamics can reveal how DNA looping and condensation are coordinated.

In this work, we probe the mechanisms of polycation-mediated nucleic acid condensation through a combination of all-atom molecular dynamics (MD) simulations and DNA precipitation and FRET measurements. First, we show that DNA condensation mediated by different types of polycations is governed by the same methylation-dependent mechanism. By using constructs that combine elements of DNA and RNA structures, we establish the dominant role of methylation in determining the strength of DNA condensation. We also show that the strength of DNA condensation depends not only on nucleotide composition but also on their sequence arrangement. Finally, our three-color smFRET measurements revealed strong anti-correlation between DNA binding and DNA looping dynamics, suggesting that the two phenomena might be governed by a common physical mechanism.

MATERIALS AND METHODS

DNA design and synthesis

Labeled and unlabeled dsDNAs were made using commercial kits for PCR (Phusion High-Fidelity PCR Master Mix, New England Biolabs) and PCR purification (Genet Bio) and used for FRET (ensemble FRET, smFRET) and DNA precipitation measurements, respectively (Supplementary Figure S1). For the labeled dsDNAs, one of the primers was fluorescently labeled at the 5′-end amine modification. Precipitation measurements were performed with dsDNAs without the amine modification. Supplementary Table S1 lists the design of DNA templates used in this study. 120 bp templates with TA, CG, or ^{m}CG-rich sequence were designed and prepared as described earlier (6). UA-rich DNA was custom-synthesized, replacing all the deoxythymines in the central 80 bp region of TA-rich DNA with deoxyuracils, followed by thermal annealing of equimolar amounts of complementary strands. For consistency, the annealed dsDNAs were purified using the same PCR purification kit. For the control of the methyl group arrangement, we designed additional 120 bp DNAs with more consecutive thymine or cytosine residues on one side (TT-rich and CC-rich DNAs; Supplementary Table S1). In case of CC-rich DNA, as the consecutive guanines seemed to hinder efficient PCR amplification, we custom-synthesized both DNA strands and annealed them as we did for UA-rich DNA. For DNA length dependence, we designed additional 60, 80 and 100 bp TA-rich DNAs by reducing the size of the middle region (Supplementary Table S1), followed by the same PCR and purification steps. In case of 60 bp TA-rich DNA, we used Prime Taq Premix kit following standard protocol for PCR (Jenet bio) as the Phusion Master Mix resulted in dsDNAs with erroneous mixture of lengths. All PCR products were checked for size and purity. For ^{m}CG-rich DNA, CpG methylation by the methyltransferase M.SssI was checked by BstUI digestion of unmethylated sites, which showed saturation in within our reaction time of 8 hrs (Supplementary Figure S1). All DNA sequences were designed to avoid unwanted hairpin structure and PCR primers were designed to have proper ΔG values for efficient amplification.

Ensemble FRET measurement

For each sequence construct, we mixed two kinds of dsDNAs labeled with Cy3 and Cy5, 1 μM each, in buffer solution containing 20 mM Tris–HCl pH 8.0, 1 mM EDTA, 50 mM NaCl. After adding varying concentration of spermine·4HCl (Sigma Aldrich) or Lys_6·6HCl (Genscript), the solution was incubated for 5 minutes before measuring fluorescence with spectrofluorophotometer (RF-5310PC, Shimadzu). The change of pH was checked for each mixture, which remained less than 0.2 in all cases. Upon excitation at 550 nm, Cy3 and Cy5 emission peaks were observed at 568 and 668 nm, respectively (Supplementary Figure S2). Fluorescence spectrum of the buffer solution was subtracted from each spectrum. In order not to be affected by the different quantum yield of Cy3 and Cy5 and also the wavelength-dependent detection efficiency of the spectrofluorophotometer, we calculated FRET efficiency as $E = 1 - I_{DA}/I_D$, where I_{DA} and I_D are the donor fluorescence peak intensities with and without the acceptor, respectively. At each condition, FRET efficiency was averaged from triplicate measurements and the error bars represent the S.E.M. between the measurements.

DNA precipitation

Each kind of DNA was added at 0.6 μM concentration to buffer solution containing 20 mM Tris–HCl pH 8.0, 1 mM EDTA, 50 mM NaCl, and varying amount of spermine·4HCl/Lys_6·6HCl to total volume of 50 μl. After 5 min incubation at room temperature, the sample was centrifuged for 5 min at 21 000 × g. Concentration of supernatant DNA was measured from 1 μl of supernatant carefully taken, using spectrophotometer (NanoDrop 2000C, Thermo Scientific). Precipitation occurred fast, showing no further difference in supernatant concentration after 5 min incubation. At each condition, data were averaged from triplicate measurements using independently prepared DNAs and buffer. The error bars represent the S.E.M. between the measurements.

Three-color smFRET measurement

Vesicle encapsulation. We encapsulated two kinds of TA-rich DNAs, one labeled with Cy3 and the other labeled with Cy5/Cy7 at both ends, in lipid vesicles as described earlier (6). Briefly, lipid films were prepared by mixing 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine-*N*-(Cap Biotinyl) (0.05 mg/ml) with dimyristoyl phosphatidylcholine (5 mg/ml) (Avanti Polar Lipids) in chloroform and gently drying with nitrogen gas in glass bottle. The lipid film was hydrated with 100 μl buffer solution containing 100 mM NaCl and 25 mM Tris–HCl pH 8.0, followed by repeated freeze-thaw cycles to homogenize the suspension. Two kinds of DNAs were added to make total volume of 200 μl and concentration of 400 nM for each, which corresponds to one molecule in a spherical volume of 200 nm diameter. The solution was extruded 11 times through membrane filter with 200 nm track-etched pores (Whatman, 800281) to create uniformly sized vesicles. The encapsulation efficiency of 15 ± 3% was estimated as the ratio of the number of vesicles containing two dyes (two Cy3's, two Cy5's, or one of each Cy3 and Cy5) to the number of vesicles containing one dye (Cy3 or Cy5).

smFRET measurement. Using the biotin attached to one kind of lipid, the lipid vesicles were immobilized on a PEG-coated surface through biotin-neutravidin interaction (Figure 4A). Fluorescence signals from individual vesicles were collected by custom-built total internal reflection microscopy. Imaging solution contained 1 mg/ml glucose oxidase, 0.04 mg/ml catalase, 0.6% glucose, saturated Trolox (∼3 mM), 25 mM Tris–HCl pH 8.0, 25 mM NaCl and varying amount of polyamine. By alternating excitation of Cy3 (532 nm laser) and Cy5 (647 nm laser) per each camera frame, all three pairwise FRET efficiencies were pseudo-simultaneously measured.

Analysis of single molecule traces. Single molecule traces exhibited a variety of behaviors, including parallel binding, anti-parallel binding, bending, or combinations of these events (Supplementary Figure S9). At first, we selected the traces containing only a single copy of Cy3, Cy5 and Cy7 by examining their photobleaching steps and signal intensities. We categorized the traces depending on whether they show clear events of binding, bending, or both. The criteria we used for each event was that the FRET efficiency jumped over 0.5 or maintained FRET level of 0.25 or higher in Cy3–Cy5 (parallel binding), Cy3–Cy7 (anti-parallel binding), or Cy5–Cy7 (looping) FRET signals. The number and fraction of each kind of traces are shown in Supplementary Figure S10 for three randomly chosen movies. Cross correlation between Cy3–Cy5 and Cy5–Cy7 FRET signals and scatter plot of average FRET efficiencies were drawn for 100 collected traces exhibiting both binding and bending events and lasted over 30 seconds.

Molecular dynamics simulations

Force field. All simulations were carried out using the AMBER ff99-derivative force fields (57); bsc0 for DNA (58), SB-ILDN-PHI for protein (59–61), Joung and Cheatham for ions (62), Dai *et al.* for spermine (63), and the TIP3P water model (64). Custom pair-specific van der Waals parameters (CUFIX) were used to describe non-bonded interactions between amine and carboxylate/phosphate groups, between sodium ion and phosphate/carboxylate groups, between aliphatic carbon atoms (CT atom types in AMBER), and between sodium and chloride ions (65,66). The complete set of force field used in this report is available online: http://bionano.physics.illinois.edu/CUFIX.

All-atom MD protocols. All MD simulations were carried out using the Gromacs 4.5.5 package (67). Integration time step was 2 fs. The temperature was kept constant at 300 K using the Nosé–Hoover scheme (68,69). The pressure in the *xy* plane (normal to DNA) was kept constant at 1 bar using the Parrinello–Rahman scheme (70) while the length of the box in the *z* direction was kept constant. Van der Waals forces were computed using a 7–10-Å switching scheme. For the computation of electrostatic forces, we employed the particle-Mesh Ewald summation scheme (71) with a 1.5-Å Fourier-space grid and a 12-Å cutoff for the real-space Coulomb interaction. Covalent bonds involving hydrogen were constrained using SETTLE (72) and LINCS (73) algorithms for water and non-water molecules, respectively.

All-atom potential of mean force calculations. The potential of mean force (PMF) as a function of inter-DNA distance in the presence of Lys_6 peptides was computed using a previously described method (6). A 21-bp duplex DNA was prepared using the nucleic acid builder of the AMBER package (74). The sequences of duplex DNA helices were either $(CG)_{10}C{\cdot}(GC)_{10}G$ or $(TA)_{10}T{\cdot}(AT)_{10}A$. A pair of duplex DNA molecules parallel to the z axis were placed in a hexagonal box that measured ∼140 Å within the *xy* plane and 71.4 Å along the *z* axis. Each DNA strand was effectively infinite due to a covalent bond between the first and the last nucleotides under the periodic boundary conditions. Note that only the orientation of DNA molecule was constrained, but DNA molecules were free to rotate with respect to the z axis and free to translate in all directions. Fourteen Lys_6 molecules were randomly added. Because the charges of Lys_6 molecules cancel out the charges of DNA, the concentration of Lys_6 was approximately at sub-mM level. The simulation cell was solvated using the explicit 150 mM NaCl solution that contains 100 sodium and chloride ions and ∼35 000 water molecules. Each system was equilibrated for at least 50 ns. Using the last frame of each equilibration trajectory, umbrella sampling simulations were performed. The reaction coordinate of the umbrella sampling was defined as the distance between the centers of mass of the two DNA molecules projected onto the *xy* plane. The harmonic force constant for the umbrella sampling simulations was 2000 kJ mol^{-1} nm^{-2}. The reaction coordinate was varied from 24 to 45 Å in 1-Å increments. The umbrella sampling simulations were ∼100 ns in duration in each sampling window. The time trajectory along the reaction coordinate was recorded every 2 ps, and the first 20 ns of each trajectory was discarded. By analyzing the umbrella sampling trajectories using the weighted histogram analysis method implemented in the Gromacs package, the interaction free energy as a function of inter-DNA distance was

computed for two different sequences: $(CG)_{10}C{\cdot}(GC)_{10}G$ or $(TA)_{10}T{\cdot}(AT)_{10}A$. To estimate the error, we divided the umbrella sampling trajectory in each window into four non-overlapping subset trajectories and computed four interaction free energy profiles using the subset trajectories. The error bars to the PMF plots were determined as the standard errors of the four free energy profiles. To validate the convergence, an additional umbrella sampling simulation was performed with coordinate swapping. For example, the last frame in each window of $(CG)_{10}C{\cdot}(GC)_{10}G$ simulations was taken, and the sequence of DNA was modified to $(TA)_{10}T{\cdot}(AT)_{10}A$. Using those modified coordinates as the first frame in each corresponding window, 100-ns umbrella sampling simulations and the WHAM analysis were repeated.

Coarse-grained simulations. A custom coarse-grained model of dsDNA was developed to reproduce the sequence-dependent DNA–DNA interactions observed in all-atom simulations. The parameterization was done against the all-atom results for the spermine solution systems where the total charge of spermine completely compensated the total charge of DNA (Supplementary Figure S4A). In our beads-on-a-string model of dsDNA, each bead represented a 10 base-pair dsDNA segment (50,000 atomic mass units); two neighboring beads were connected by a harmonic spring of the spring constant $k_{1-2} = 1000$ kJ/mol·nm^2 and the equilibrium length 3.4 nm. Considering the strong screening by spermine, the DNA's persistence length was set to 30 nm by adding harmonic bonds to all 1–3 pairs; the force constant $k_{1-3} = 50$ kJ/ mol·nm^2 and the equilibrium length was 6.8 nm (Supplementary Figure S11A). Lennard–Jones (LJ) non-bonded interaction potential was applied to all pairs of beads excluding beads within the same molecules separated by 3 or less beads. Two different LJ bead types corresponding to TA- and CG-repeat segments were defined. The interaction parameters were obtained by computing the free energy profile for a pair of effectively infinite parallel coarse-grained DNA molecules and comparing it to the all-atom data. Using LJ $\varepsilon = 6$ and 4 kJ/mol for TA and CG beads, respectively, and $\sigma = 2.7$ nm for both, the coarse-grained free energy profiles matched the all-atom targets (Supplementary Figure S11B). The coarse-grained simulations (stochastic dynamics) were performed using the Gromacs 4.5.5 package. Temperature and the inverse friction constant (tau_t) were set to 300 K and 100 ps, respectively. Time step was 0.2 ps and the cutoff for the LJ interactions was 4 nm.

RESULTS

Polylysine and spermine both condense TA-rich DNA more strongly than CG-rich DNA

To measure the effect of TA content on the condensation strength, we designed two 120 bp double-stranded DNA (dsDNA) constructs to have their central 80 bp region made up entirely of TA bp (TA-rich) or contain only 25% of randomly distributed TA bp (CG-rich) (see Figure 1A, Supplementary Figure S1, and Supplementary Table S1). Both constructs had common 20 bp primer regions flanking the central region and Cy3 or Cy5 fluorophores attached to one of the primers for FRET measurements. The association of DNA constructs was studied using two types of polycations: lysine hexamer (Lys_6) and spermine (Figure 1B). We used Lys_6 as a mimic of highly charged, intrinsically disordered lysine-rich regions, such as those found in histone tails which are central to the epigenetic control of gene expression. Spermine is a tetravalent biogenic polycation widely used for *in vitro* DNA condensation experiments and suggested to control key cellular processes (14,29).

Our ensemble FRET measurements show that both Lys_6 and spermine condense the TA-rich DNA construct more strongly than the CG-rich one (Figure 1C, D and Supplementary Figure S2). In a typical measurement, we mixed equimolar amounts of Cy3- and Cy5-labeled dsDNA constructs of the same nucleotide sequence and measured the FRET efficiency as a function of the polycation concentration (see Materials and Methods for details). Condensation of the DNA constructs places the fluorophores within short distance of one another, increasing the FRET efficiency. In all experiments, the FRET signal was found to rise abruptly above the background level as the concentration of either Lys_6 or spermine increased. However, the abrupt increase of FRET occurred at a significantly lower polycation concentration for the TA-rich constructs, indicating their higher propensity to form condensates. At high polycation concentrations, the FRET signal from the TA-rich constructs saturated at higher values than that from the CG-rich ones, suggesting that TA-rich DNA forms more compact condensate.

Precipitation assay measurements, Figure 1E and F reaffirmed the conclusions drawn from the ensemble FRET data. In a precipitation assay, solution concentration of DNA (supernatant concentration) was measured after incubation of the DNA constructs in the presence of polycations and spinning down the precipitate (Materials and Methods) (22). As with the FRET data, the supernatant concentration of DNA dropped abruptly at a threshold concentration of polycation indicating DNA condensation; the threshold concentration was lower for the TA-rich constructs in both Lys_6 and spermine solutions. The threshold concentration of spermine was in the 0.1–1 mM range, which overlaps with its physiological concentration (30). The threshold concentration for Lys_6 was about 10 times lower than that of spermine, presumably due to the higher charge valency of Lys_6. The threshold concentration was found to depend on the length of the DNA constructs, with shorter constructs requiring higher polyamine concentration to initiate DNA condensation (Supplementary Figure S3), consistent with earlier observations (8,9).

MD simulations determined that the phenomena of DNA condensation by Lys_6 and spermine share common molecular mechanism including the dependence on the TA content. In a typical simulation setup containing Lys_6, two effectively infinite parallel dsDNA molecules of either TA-repeat, $(TA)_{10}T{\cdot}(AT)_{10}A$, or CG-repeat, $(CG)_{10}C{\cdot}(GC)_{10}G$, sequence were submerged in an aqueous solution containing 14 Lys_6 molecules and 150 mM NaCl (Figure 1G). The inter-DNA distance, d, was restrained to a prescribed value using a harmonic spring potential. Twenty two replicas of each system were simulated for 100 ns each probing the inter-DNA distance range from 24 to 45 Å in 1 Å incre-

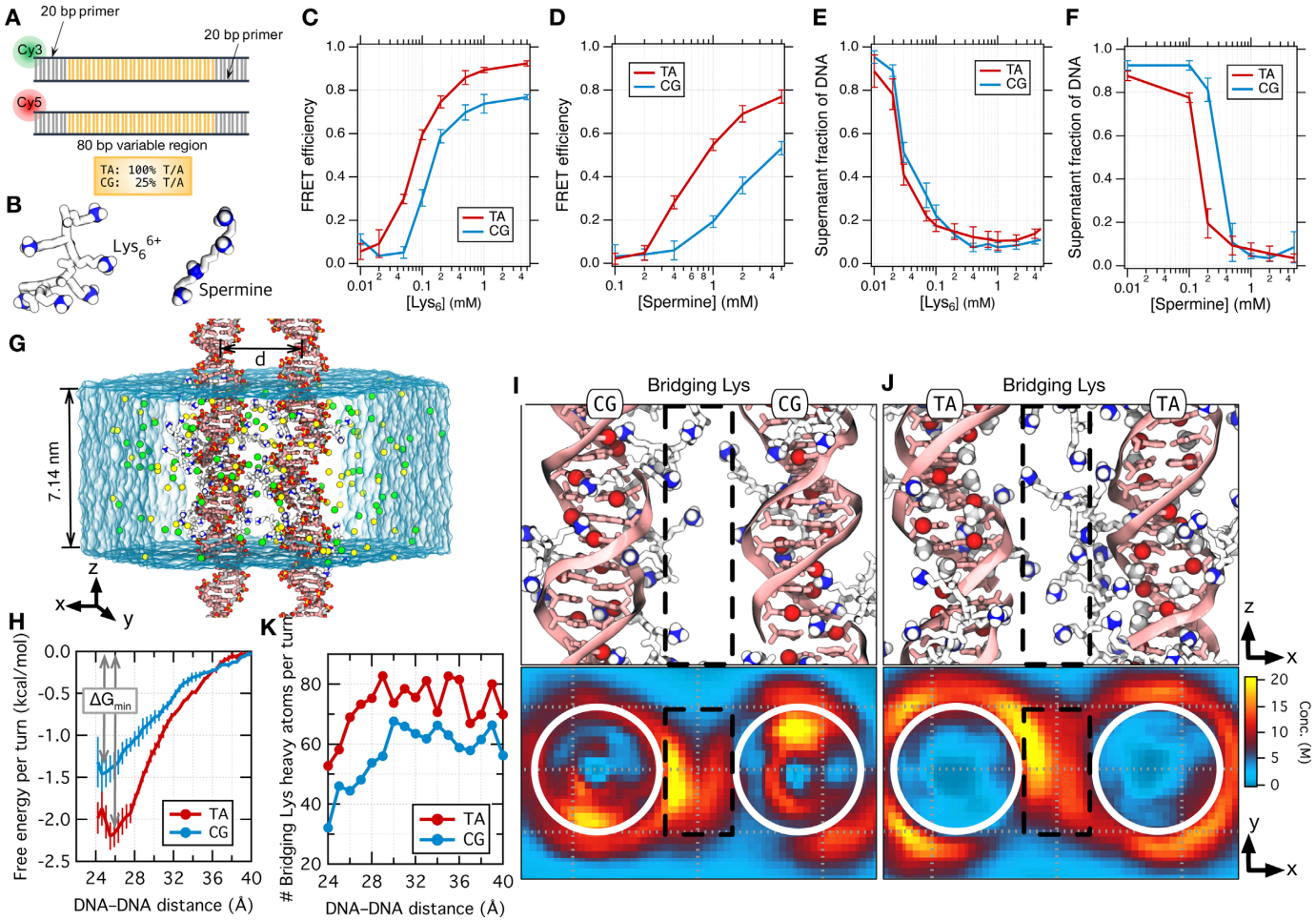

Figure 1. Polycations condense TA-rich DNA more strongly than CG one. (**A**) Design of TA-rich and CG-rich dsDNA constructs. For FRET measurements, the fluorophores were placed at one end of the constructs. The exact sequences can be found in Materials and Methods. (**B**) Chemical structure of the hexavalent Lys_6 and tetravalent spermine. (**C**, **D**) Titration curves of DNA condensation by Lys_6 and spermine from the ensemble FRET measurements. (**E**, **F**) Titration curves of DNA condensation by Lys_6 and spermine from the precipitation measurements. All measurements were done in the presence of 50 mM NaCl and at 25°C. All error bars represent S.E.M. from triplicate independent measurements. (**G**) Setup of MD simulations with a pair of parallel 21-bp dsDNAs that are effectively infinite under periodic boundary conditions in an aqueous solution (semi-transparent surface) containing 150 mM NaCl and 14 Lys_6 molecules. The inter-DNA distance, d, is defined as the distance between center-of-masses of two dsDNAs projected on the xy plane. (**H**) Comparison of the interaction free energy, computed as a function of d using the umbrella sampling method (Materials and Methods). Inter-dsDNA binding strength was determined by the well depth of the free energy curves, ΔG_{min}. (**I**, **J**) Snapshots of the simulations showing the positions of Lys_6 (upper panels) and the projected density maps of Lys heavy atoms (lower panels) averaged over the simulated time for CG and TA DNA pairs, respectively, taken at $d = 32$ Å. White circles of 1 nm radius in the lower panels indicate the positions of DNA helices. See Supplementary Figure S5 for the projected density maps at a range of d from 28 to 40 Å. (**K**) The count of bridging Lys heavy atoms per pair of turn, computed by integrating the local concentration heatmaps over $x = [-d/2+10$ Å, $d/2-10$ Å$]$ and $y = [-10$ Å, 10 Å$]$, as a function of d. Black dashed boxes in panels H and I indicate the integration domain at $d = 32$ Å.

ments (Materials and Methods). Using the weighted histogram analysis method (31,32), we obtained the free energy profile of the pairs of the DNA molecules as a function of the inter-DNA distance (Figure 1H). Both free energy profiles exhibit an attractive potential well, $\Delta G_{min} < 0$. The depth of the attractive potential is about 0.8 kcal/mol per helical turn deeper for the TA-repeat DNA pair in comparison to that of the CG-repeat pair, indicating stronger attraction in the former system, in qualitative agreement with ensemble FRET and precipitation measurements. Reducing polyamine concentration in MD simulations reduced ΔG_{min} values (Supplementary Figure S4), indicating weaker attraction in less concentrated solutions, in accordance with FRET and precipitation data (Figure 1C–F). Interestingly, when the charge of the polyamine molecules compensates only 50% of the total DNA charge, the sequence-dependence of DNA–DNA interaction produces qualitatively different forces: attraction is observed for AT-rich dsDNA whereas CG-rich dsDNA repel one another, (Supplementary Figure S4E).

The sequence of the DNA constructs affected the distribution of Lys_6 near DNA, Figure 1I and J. In the case of the CG-repeat DNA, Lys_6 molecules were found both in the major groove of DNA (inside the white circles) and at the periphery of the DNA molecule (outside the white circles) (Figure 1I, Supplementary Figure S5A). By contrast, Lys_6 molecules were found to reside exclusively at the periphery of the TA-repeat DNA, avoiding the DNA's ma-

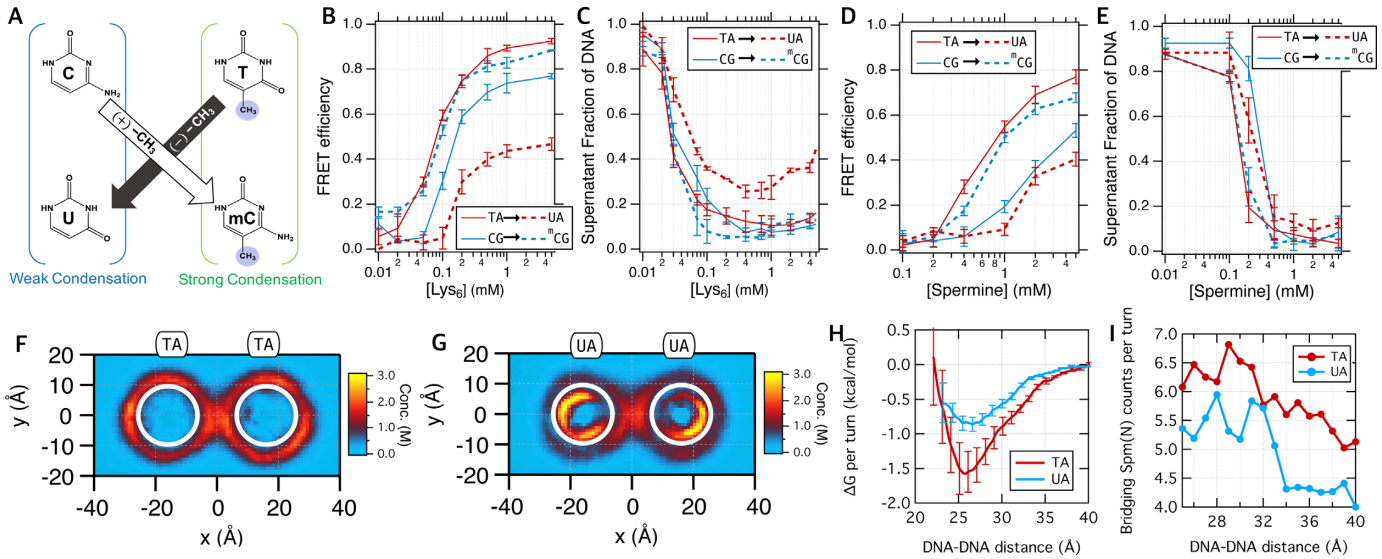

Figure 2. Abundance of C5 methyl group controls the condensation strength. (**A**) Replacing cytosine with 5-methylcytosine adds a C5 methyl group, while replacing thymine with uracil effectively subtracts a C5 methyl group. Details of the construct design can be found in Materials and Methods. (**B–E**) Methylation-dependent changes of the DNA condensation by Lys_6 (B, C) and spermine (D, E) from ensemble FRET and precipitation measurements. All error bars represent S.E.M. from triplicate independent measurements. (**F**, **G**) The density maps of amine nitrogen atoms in spermine averaged over the simulated time for TA- (F) and UA-repeat (G) DNAs, respectively, taken at the inter-dsDNA distance of 32 Å. (**H**) Comparison of the inter-dsDNA interaction free energy, computed as in Figure 1H. **(I)** The count of amine nitrogen atoms in bridging spermine per pair of turn as a function of the distance, computed as in Figure 1K. Note that simulation data for TA-repeat DNA is taken from (6).

jor groove (Figure 1J and Supplementary Figure S5B). The nucleotide composition has a similar effect on the distribution of spermine molecules regardless of the spermine concentration (Supplementary Figures S6 and S7) (6). The number of Lys_6 or spermine molecules located between the two DNA molecules (black rectangles in Figure 1I and J) was higher for the TA-repeat DNA in comparison to the CG-repeat one at all inter-DNA distances (Figure 1K and Supplementary Figure S4B, D, F). According to the bridging model of DNA attraction (9,33,34), a higher density of such bridging polycations should generate stronger attraction between DNA molecules, which is exactly what we observe in our MD simulations and experiments.

C5 methyl groups promote DNA condensation

Previous MD simulations found the presence of C5 methyl groups in the major grooves of DNA molecules to determine the strength of spermine-mediated DNA attraction (6). To test if this mechanism also applies to the inter-DNA interaction mediated by other biogenic polycations (such as cationic peptides) and if it can explain weaker condensation propensity of RNA molecules, we repeated our ensemble FRET and precipitation experiments using two additional DNA constructs, Figure 2A. First, we modified the central 80 bp of the CG-rich DNA construct so that all CG nucleotides form CpG sites and methylated all the CpG cytosines, making ^{m}CG-rich DNA. Second, we removed C5 methyl groups from the central 80 bp region of the TA-rich DNA by replacing all the deoxythymines with deoxyuracils, making UA-rich DNA. In both ensemble FRET and precipitation measurements, and for both Lys_6- and spermine-containing electrolytes, the ^{m}CG-rich DNA exhibited stronger condensation than the CG-rich DNA, almost as strong as the TA-rich DNA (see dashed blue curves in Figure 2B–E). On the other hand, the UA-rich DNA showed a weaker condensation propensity than the TA-rich DNA, which, in case of Lys_6, was even weaker than that of the CG-rich DNA (see dashed red curves in Figure 2B–E). These results confirm that the sequence-dependence of the condensation strength arises mainly from the difference in the C5 methyl group content within the DNA. Furthermore, the weaker condensation of the UA-rich DNA is consistent with the results of recent studies that found RNA condensation to be generally weaker than that of DNA (24,25). Noticing that unmodified RNA does not contain C5 methyl groups, our results suggest that the lack of such methyl groups might be the factor causing weaker RNA condensation, which may promote expulsion of dsRNA molecules from a dsRNA/dsDNA mixture.

Repeating MD simulations for a pair of UA-repeat DNA constructs in the presence of spermine polycations reaffirmed the molecular origin of the condensation strength regulation, Figure 2F–I. In contrast to the TA-repeat DNA, where spermine is mostly found at the periphery and in-between of the DNA molecules (Figure 2F), spermine is found to enter the grooves of the UA-repeat DNA (Figure 2G). As a result, the well depth in the free energy profile of the UA-repeat DNA pair is 0.8 kcal/mol per helical turn shallower than that of the TA-repeat DNA pair (Figure 2H). Thus, removing the C5 methyl group makes DNA condensation weaker. In accordance with the bridging model of DNA condensation, we find the number of spermine molecules residing between DNA molecules to be smaller for the UA-repeat DNA pair in comparison to the TA-repeat pair at all inter-DNA distances (Figure 2I).

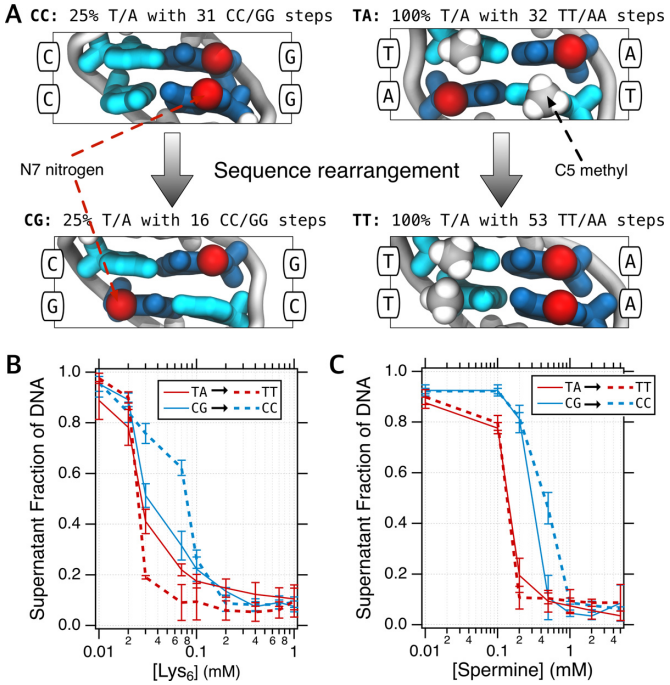

Figure 3. Condensation strength depends on local nucleotide sequence. (**A**) Schematics depicting different sequence arrangements. For the same sequence compositions of high or low TA content, we designed additional sequences having more consecutive nucleotides of the same kind on the same strand (TT-rich and CC-rich DNAs). (**B**, **C**) Titration curves of DNA condensation by Lys_6 and spermine, respectively, from the precipitation measurements. Condensation of TT-rich DNA was stronger than or similar to that of TA-rich DNA while the condensation of CC-rich DNA was significantly weaker than that of CG-rich one. All error bars represent S.E.M. from triplicate independent measurements.

Condensation strength depends on local DNA sequence

We have already established the presence of the C5 methyl groups in the major groove of DNA to be the key factor determining the strength of polycation-induced DNA condensation. To determine if the arrangement of the methyl groups within the major groove—a proxy for DNA sequence—can influence DNA condensation, we used DNA constructs that had the same overall nucleotide composition as the TA- and CG-repeat DNA but different arrangement of the nucleotides within the strands. Specifically, we designed TT-rich and CC-rich DNA constructs that had more pyrimidines of the same kind arranged consecutively along the same strand (Figure 3A, see Supplementary Table S1 for the exact sequence). The introduction of the pyrimidine dinucleotide steps exacerbated the difference between the TA- and CG-rich constructs in the precipitation measurements carried out using either Lys_6 or spermine electrolytes, Figure 3B and C. Whereas precipitation of the TT-rich and TA-rich DNA had similar dependence on polycation concentration, precipitation behavior of the CC-rich DNA indicated weaker attraction in comparison to CG-rich DNA, i.e. a larger concentration of polyamines was required to initiate a condensation transition. These results are consistent with the predicted dependence of the DNA condensation strength on the DNA sequence obtained using the MD method (Supplementary Figure S4A, C, E) (6).

Figure 4 summarizes the dependence of the condensation strength on the type and arrangement of basepairs within a DNA molecule. From experimental measurements, the DNA condensation strength decreases in the following order: TT ≥ TA ≈ ^{m}CG > UA ≥ CG > CC (Figure 4A–D). Results of our MD simulations show similar dependence of the attractive interaction strength on DNA sequence: TT ≥ TA ≈ ^{m}CG > UA ≈ CG > CC (Figure 4E–F). The linear correlation between the number of bridging polycations and $|\Delta G_{min}|$ clearly suggests that the bridging polycations are the major determinant of DNA condensation (Figure 4G). In summary, both the presence of C5 methyl groups and the sequence of DNA affect the population of polycations in the bridging region between the two DNA molecules and, consequently, the population of the bridging polycations determines the DNA condensation strength (Figure 4H).

Molecular mechanism of the sequence- and methylation-dependent binding of polycations to major and minor grooves

To determine how the DNA sequence and the arrangement of the C5 methyl groups (Figure 5A) control the binding of polycations, we analyzed our MD trajectories to count the average number of polycations bound to the major and minor grooves of dsDNA. Figure 5B shows representative configurations of spermine molecules near dsDNA. Our analysis revealed that the presence of electronegative sites (e.g. N7 and N3 nitrogen atoms), bulky C5 methyl groups, and the width of the minor groove are the three major determinants of the DNA attraction strength.

For CG or CC-rich dsDNA, electronegative N7 nitrogen atoms strongly attract spermine to the major groove of dsDNA whereas electropositive N2 atoms repel spermine from the minor groove (13). Thus, although the minor groove of the CG or CC-rich dsDNA is relatively wide (Figure 5C), spermine was predominantly bound to the major groove (Figure 5D and E). More spermine molecules were bound to the CC-repeat dsDNA than the CG-repeat dsDNA (2 and 1.5 molecules per turn on average), because a continuous tract of N7 atoms enhances the binding. Conversely, ^{m}CG repeats had only 0.7 spermine molecules per turn bound to the major grooves because the C5 methyl groups in the vicinity of N7 atoms sterically block spermine binding, Figure 5A.

For TT and TA repeats, less than half spermine molecules per turn were bound to the major grooves (Figure 5D), partially because of the presence of the C5 methyl groups of thymine nucleotides. For UA repeats that lack C5 methyl groups, 0.7 spermine molecules per turn were bound to the major grooves, comparable to ^{m}CG repeats and only half compared to CG repeats. These observations reaffirm the fact that the major groove of guanine nucleotides is generally more favorable to cation binding than that of adenine nucleotides (35). Interestingly, the binding of spermine to the minor grooves of TT, TA, and UA repeats shows a significant correlation with the minor groove width (Figure 5E). Thus, TT-repeat dsDNA, which has the narrowest minor groove, also has the least number of spermine molecules in minor grooves (~0.2 per turn). The minor groove of the

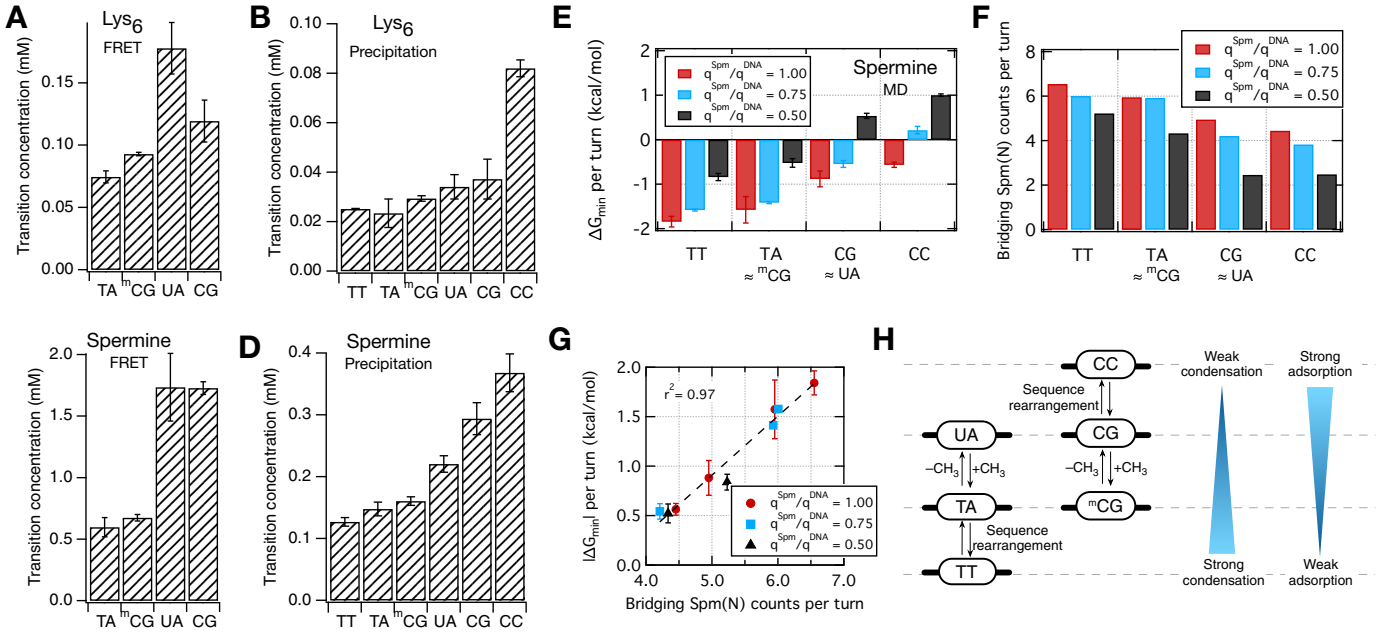

Figure 4. Summary of the dependence of DNA condensation strength on the sequence and methylation of DNA. (**A**, **B**) Transition concentration of Lys_6 observed in the ensemble FRET and precipitation measurements by fitting to sigmoid curves, shown for each DNA construct. Error bars represent the error from fitting. (**C**, **D**) Same as in panels A and B but for spermine solutions. (**E**) Minima of the pairwise DNA–DNA free energy profiles, ΔG_{min}, measured in MD simulations performed in the presence of spermine. **(F)** The number of amine nitrogen atoms of bridging spermine molecules per pair of turn averaged over $d = [25\ Å, 40\ Å]$, shown for each DNA construct. (**G**) The linear correlation between the number of bridging spermine nitrogen atoms and the well depth of the attractive interaction free energy curves ($\Delta G_{min} < 0$). (**H**) Schematic representation of the dependence of DNA condensation strength on the methylation and sequence arrangement, interpreted as the likelihood of polycations adsorption on the DNA grooves.

UA-repeat construct is 2.5 Å wider and accommodates, on average, about 0.7 spermine molecules per turn (Figure 5E). This observation suggests that bulky C5 methyl groups can indirectly affect binding of spermine to the minor grooves of dsDNA by changing the minor groove width.

Overall, the number of spermine molecules bound to the grooves of dsDNA is inversely correlated with the number of bridging polycations and $|\Delta G_{min}|$ (compare Figure 5F with Figure 4E and F). Similar groove-binding trends were observed in the case of lysine peptides (Figure 5G). However, the difference in the chemical structure of the polycations (Figure 1B) modulates their binding affinity to the grooves. For both TA and CG repeats, lysine peptides bind to the minor groove of DNA more strongly than spermine molecules do (compare Figure 5F and G), possibly because the long branched side chains of polylysine can penetrate into the narrow minor grooves, as seen in the crystal structures, for example, of nucleosomes (36).

Similar analysis was performed to investigate the roles of sodium and water molecules bound to the grooves of dsDNA (Supplementary Figure S8). The number of sodium ions bound to major and minor grooves were less than one per turn in all cases despite orders of magnitude higher concentration in bulk solution in comparison to spermine (Supplementary Figure S8A–D. More importantly, no bridging sodium ions were observed, suggesting that the sodium's contribution to condensation is insignificant (Supplementary Figure S8A and B). Hydration of the grooves is highly correlated with the groove's geometry (37,38), including formation of a crystallographically resolved chain of water molecules (the 'water spine') in the minor groove of TA-rich dsDNA (37). In our MD simulations, 35 water molecules are found in every turn of the minor groove of TA-rich dsDNAs, which is about 30% less than the number of water molecules in the minor grooves of CG-rich dsDNAs (Supplementary Figure S8F). This observation suggests that hydration may affect polycation binding to the minor grooves of dsDNA by controlling the width of the grooves.

DNA binding is temporally anti-correlated with DNA looping

The above experimental approaches do not probe the internal deformation of DNA molecules whereas the MD simulations impose parallel arrangement of the DNA molecules by means of the periodic boundary condition. A recent study found the propensity of short dsDNAs to bend and form loops to be strongly correlated with the presence of the C5 methyl group; i.e. higher methylation content makes DNA less flexible and causes the nucleosomal DNA to unwrap easier (5). To determine if DNA–DNA association, which is also strongly affected by the C5 methyl content, is linked to DNA looping, we designed a multi-color single molecule FRET (smFRET) measurement that could simultaneously characterize DNA looping and DNA–DNA association. We prepared a pair of TA-rich DNA constructs, one with Cy3 at one end and the other of the same sequence with Cy5 and Cy7 at two ends, Figure 6A. By co-encapsulating them in lipid vesicles immobilized on surface, we could observe weak and transient interaction between two DNA molecules, which could not be achieved through our ensemble FRET measurement because of un-

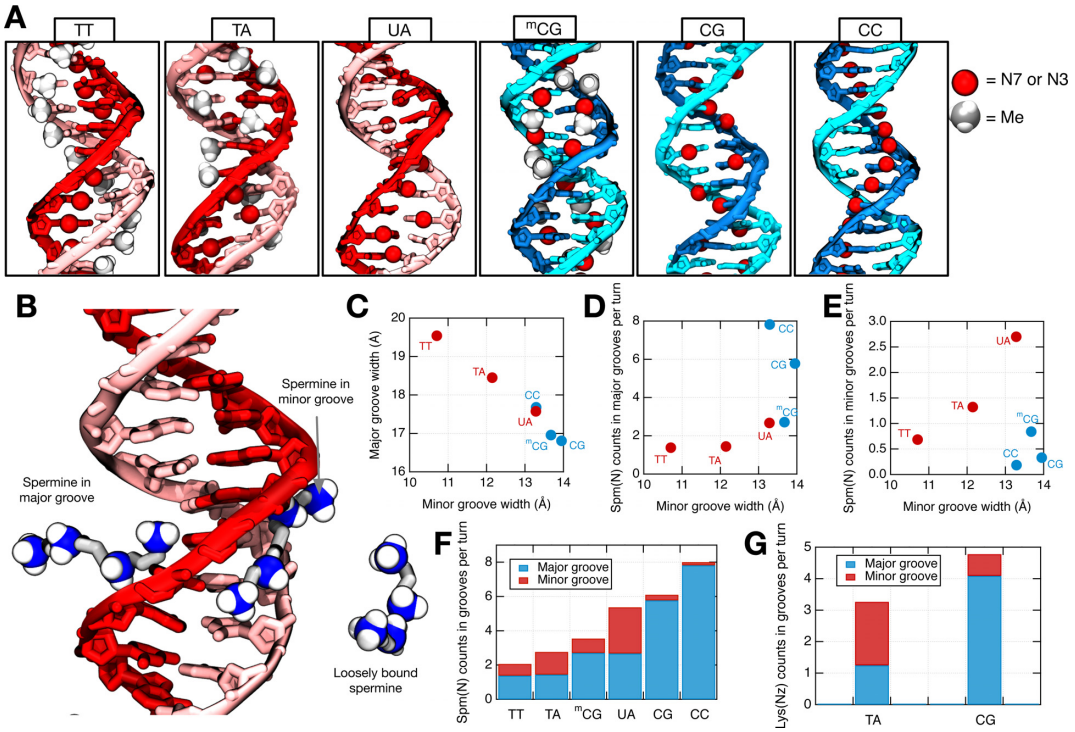

Figure 5. Binding of polycations to major and minor grooves of dsDNA. (**A**) Location of the cation binding sites (N7 and N3 atoms in the major and minor grooves, respectively) and the C5 methyl groups in dsDNA molecules of difference nucleotide sequence. (**B**) Three representative configurations of spermine near dsDNA. The snapshot was taken from the simulation of UA-repeat DNA. (**C**) The width of the major and minor grooves as a function of DNA sequence. The widths were measured using the 3DNA package (75). (D, E) The number of spermine nitrogen atoms in the major (**D**) and minor (**E**) grooves of a pair of 20-bp DNA molecules as a function of the minor groove width. Each data point indicates the average values of the respective umbrella sampling simulations; the standard error is smaller than the symbol size. Atoms located within 9Å of the DNA central axis were considered to be groove-bound. (F,G) The total number of groove-bound spermine nitrogen atoms (**F**) and lysine amine nitrogen atoms (**G**) as a function of the DNA sequence.

controlled DNA aggregation. We performed three-color smFRET measurements with alternating excitation of Cy3 and Cy5 dyes, which allowed us to simultaneously characterize binding of the two DNA molecules (Cy3–Cy5 or Cy3–Cy7 FRET) and looping of one of the DNA constructs (Cy5–Cy7 FRET).

Analysis of the smFRET measurements identified events that correspond to the parallel binding of two dsDNA constructs (Cy3–Cy5 FRET), anti-parallel binding (Cy3–Cy7 FRET), and the looping motion of the dsDNA (Cy5–Cy7 FRET). Supplementary Figure S9 shows representative examples of FRET traces indicative of DNA binding, looping, and their combination. As we analyzed the relative fraction of the traces that show each kind of motion, parallel binding events turned out to be significantly more frequent than anti-parallel binding events (Supplementary Figure S10A, B), indicating the preference for binding geometry with matching sequence or methylation patterns. The observed three-color traces exhibited various FRET levels and broad range of dwell times, implying the various configurations between the pair of dsDNA molecules (Supplementary Figure S10C).

Interestingly, in the traces showing concurrent events of parallel binding and looping, these events are strongly anti-correlated with each other in time, in both Lys_6 and spermine solutions (Figure 6B and C). That is, when DNA looping occurred, DNA binding was significantly less likely to occur and vice versa. Taking into account that a pair of dsDNA molecules showed either parallel binding, anti-parallel binding or no binding over long periods of time without transitioning between different binding behaviors (Supplementary Figure S9), the inter-dsDNA FRET dynamics is more likely to reflect on partial dissociation of the two molecules rather than on complete unbinding. In that case, the negative temporal correlation would indicate that dissociated parts of dsDNA form loops easier than bound pairs of dsDNA molecules. Such negative correlation is further supported by the fact that the number of traces showing both binding and looping dynamics is much smaller than the number of traces showing individual dynamics (Supplementary Figure S10A).

We calculated cross-correlation between Cy3–Cy5 and Cy5–Cy7 FRET signals from three-color traces that lasted over 30 second. When averaged over 100 such traces, strong negative correlation emerged with a decay time of 6.3 and 9.0 sec for Lys_6 and spermine, respectively (Figure 6D, E). The scatter plot of the average Cy3–Cy5 and Cy5–Cy7 FRET levels from these traces recapitulated the negative correlation between the two types of events (Figure 6F). While we observed many traces exhibiting strongly bound or strongly looped configuration, represented by high average of Cy3–Cy5 or Cy5–Cy7 FRET efficiency, hardly any trace exhibited high average of both kinds. If the binding and looping dynamics were independent from each other,

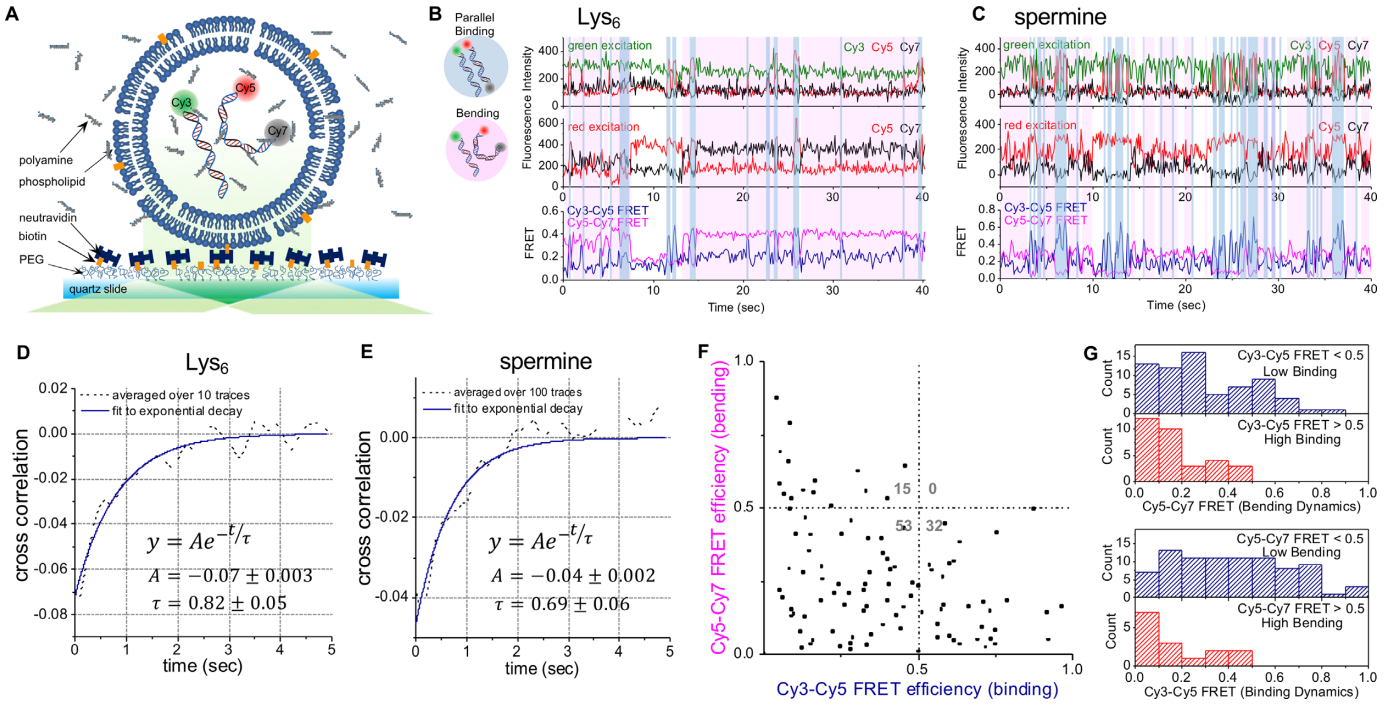

Figure 6. Three-color single molecule FRET measurement revealed the anti-correlated dynamics of binding and bending. (**A**) 120 bp TA-rich dsDNAs labeled with Cy3 and Cy5/Cy7 were co-encapsulated in a 200 nm diameter lipid vesicle. The vesicle was immobilized on a quartz slide through biotin-neutravidin binding. 532 nm (green) and 647 nm (red) lasers were alternated to excite Cy3 and Cy5, respectively. All measurements were done in the presence of 25 mM NaCl and at 25°C. (**B**, **C**) Representative traces for 1 mM Lys_6 (B) and 1 mM spermine (C). Blue and magenta shades denote the parallel binding and bending events, respectively. (**D**, **E**) Cross correlation curves between Cy3–Cy5 and Cy5–Cy7 FRET signals averaged from 10 traces for Lys_6 (D) and 100 traces for spermine (E), clearly showing negative correlation between two FRET signals. (**F**) Scatter plot of the average FRET levels of each trace. (**G**) Population histograms from the scatter plot demonstrating the negative correlation; i.e. the histogram of Cy3–Cy5 FRET level shifts lower when the traces with high Cy5–Cy7 FRET level were selected, and vice versa.

we would expect the FRET level distribution of one kind to be the same regardless of the other. In our measurements, however, the distribution of each FRET level shifted to lower values when the other FRET level was higher (Figure 6G). These results imply that the flexibility of bound dsDNA pair is different from that of isolated dsDNA and the affinity between the molecules may depend on the curvature of the molecules. Thus, attractive interaction and flexural motions should be simultaneously considered when addressing DNA condensation in the context of constrained geometry of the chromatin fiber.

Sequence-dependent condensation can produce phase separation

Our results thus far have shown that the strengths of polycation-mediated DNA condensation depend on the nucleotide sequence and methylation of DNA. It is thus possible that a mixture of dsDNA molecules or a long dsDNA molecule of a heterogeneous sequence can phase-separate spontaneously into domains enriched of particular nucleotide makeup.

To demonstrate this possibility, we have carried out coarse-grained simulations of a mixture of dsDNA molecules containing equal amounts of TA- and CG-rich DNA (Figure 7A). Starting from an initially random configuration (Figure 7B), TA-rich molecules spontaneously formed stable bundles whereas CG-rich DNA remained dispersed or weakly bound to the periphery of the TA-rich clusters (Figure 7C and Supplementary Movie 1). Considering that the condensation strength of ^{m}CG-rich DNA is comparable to that of TA-rich DNA in our MD simulations (Figure 4E), heavy methylation of DNA may also trigger phase separation. This result also suggests that double-stranded RNA without chemical

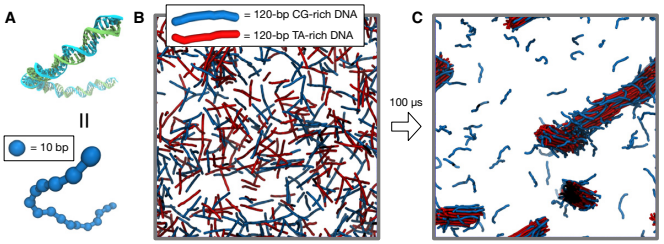

Figure 7. Phase separation of DNA in a mixture of dsDNA molecules driven by sequence-specific forces. (**A**) Coarse-graining of a 120-bp dsDNA molecule using the beads-on-a-string model, each bead representing 10 bp. The inter-molecular bead–bead interactions were parameterized to match the interaction free energy determined by the all-atom MD simulations (Supplementary Figure S4A). (**B**) Initial configuration of 250 poly$(CpG)_{60}$ (blue) and 250 poly$(TpA)_{60}$ (red) dsDNA molecules randomly placed in a cubic volume 200 nm on edge, corresponding to a DNA concentration of ~100 μM. (**C**) Configuration of the simulated system after 100 microseconds of a coarse-grained simulation.

modifications might be separated out from a mixture of dsDNA molecules as the former resists condensation (39).

DISCUSSION

Our work has characterized the effect of nucleotide content and sequence on the strength of DNA condensation mediated by biological polycations. Previous experimental studies have suggested a possibility for a dsDNA molecule to recognize the sequence of a nearby dsDNA in protein-independent manner both *in vitro* and *in vivo* (40–43). Such protein-independent homology recognition between dsDNA has been suggested to increase the efficiency of homologous recombination (42–44). However, the molecular mechanism of such direct homology sensing has remained elusive. The modulation of inter-dsDNA attraction by methylation pattern revealed here provides a plausible molecular mechanism for the direct sequence recognition between dsDNAs. Even for dsDNA having an overall similar nucleotide composition, the local variation of the nucleotide sequence builds a unique pattern of methylation density profiles. We can conjecture that matching up the methylation-rich domains between two dsDNA molecules would maximize their attractive interaction, thus leading to rough sequence alignment between the DNA pair. Further experiments and simulations using DNA of heterogeneous sequence composition may reveal the process of such sequence alignment.

The effect of the local DNA sequence on DNA condensation is intriguing. Comparing TA- and TT-rich DNA, the spacing between the C5 methyl groups in TA-rich DNA might already be small enough to efficiently block the binding of polycations to the major groove, thus not making much difference from TT-rich DNA regarding condensation strength in experiments. Considering that the N7 nitrogen of adenine is a weaker cation-binding site than the N7 nitrogen of guanine (13), the presence of non-consecutive methyl groups in TA-rich DNA might be sufficient to push the polycations out of the major groove. CC-rich DNA, on the other hand, allows the chain-like polycations to bind along the grooves more strongly than CG-rich DNA does because consecutive guanines on the complementary strand of the CC-rich strand arrange N7 nitrogen atoms into clusters (32).

Recent studies have suggested that the phase separation of chromatin leads to the formation of heterochromatic domains, consequently controlling the accessibility of transcription machineries and the organization of chromosomes (45–47). Differential affinity between DNA regions of distinct sequences revealed in our work may provide a molecular mechanism for such phase separation in chromosomes. The coarse-grained simulations in Figure 7 provide the preliminary evidence that polycation-mediated phase separation of DNA mixtures according to the local nucleotide content potentially may play a role in structural organization of the eukaryotic chromosomes (48).

Our three-color smFRET measurement directly revealed the temporal correlation between two important physical processes: DNA–DNA association and DNA looping. Establishing such anti-correlation was enabled by the unique capability of the technique to simultaneously record the binding and looping motions of a single pair of dsDNA molecules isolated from other DNA by vesicle encapsulation (49). As polycations make dsDNAs more attractive and bendable at the same time, a positive correlation between the two dynamics might be expected. However, we observed strong negative temporal correlation between the two events; the binding motion almost precluded the bending motion and vice versa. One possible explanation is that the attractive interaction between two dsDNA molecules is maximized when both are straight as such geometry would maximize the contact area and the possible number of bridging polycations between them. Alternatively, we may argue that there might exist certain configurations of polycations on dsDNAs that prefer binding or looping in mutually exclusive manner. For example, bridge-bound polycations promote DNA–DNA binding as shown from our computational data, while groove-bound polycations may facilitate DNA deformation or looping as they interrupt the base stacking interaction.

It should be noted that the smFRET signals do not directly represent the binding/dissociation events between dsDNAs or their looping/unlooping dynamics, because smFRET can detect up to ∼10 nm distance, not covering the full size of the DNAs used. For example, Cy3–Cy5 FRET transitions may sometimes represent scissor-like motions of two dsDNAs that are already bound, rather than the full binding-dissociation events. Nevertheless, considering that the bound state of dsDNAs would be maximally stable when the largest contact forms between them, Cy3–Cy5 and Cy3–Cy7 FRET can be regarded as dependable indicators of the binding event. Regarding the looping dynamics, as the J-factor of dsDNA increases with increasing length up to several hundred bps (50–53), the labeled ends of our 120 bp dsDNA are the most likely sites to come together when looping occurs. Thus Cy5–Cy7 FRET signal should work as a good indicator of the DNA looping motion.

From a broader view, our results may bear remote yet profound implications for the spatial organization of the chromosomes. Hi-C measurements have shown that the chromosomes are spatially arranged into topologically associated domains (TADs) which are grouped into two compartments. Interestingly, the distinction between the compartments strongly correlates with the local variation of TA richness (48,54). This was recently confirmed in real space by super-resolved imaging of the spatial arrangement of the TADs (55). Gene-poor domains that have higher TA content on average assembled near the nuclear membrane while gene-rich domains with lower TA content were grouped near the inner space of nuclei. Differential attractive interaction upon varying sequence composition revealed here may contribute to the observed spatial segregation of chromosomal domains. For better assessment of such relation, experiments on the role of polycations in the context of chromatin fibers rather than bare DNAs are required. Histone tails contain high density of positively charged residues and have been known to control DNA compaction through epigenetic modifications (56). To properly understand the control of DNA–histone interaction and nucleosome compaction under physiological setting, it is imperative to uncover the interplay between solution-based polycations and

histone proteins in controlling DNA looping and condensation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

J.Y. and A.A. gladly acknowledge supercomputer time provided through XSEDE Allocation Grant [MCA05S028] and the Blue Waters Sustained Petascale Computer System (UIUC). We thank Dr Min Ju Shon for helpful discussions.

FUNDING

National Research Foundation of Korea [2017R1D1A1B0 3036239, 2017M3A9E2062181, 2015R1C1A1A02036405]; Institute for Basic Science [IBS-R022-D1, IBS-R007-Y1]; National Science Foundation (USA) [PHY-1430124 to J.Y. and A.A.]; National Institutes of Health [P41 GM104601]. Funding for open access charge: National Research Foundation of Korea.
Conflict of interest statement. None declared.

REFERENCES

1. Misteli,T. (2007) Beyond the Sequence: Cellular organization of genome function. *Cell*, **128**, 787–800.
2. Dion,V. and Gasser,S.M. (2013) Chromatin movement in the maintenance of genome stability. *Cell*, **152**, 1355–1364.
3. Spector,D.L. (2003) The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.*, **72**, 573–608.
4. Grewal,S.I.S. and Moazed,D. (2003) Heterochromatin and epigenetic control of gene expression. *Science*, **301**, 798–802.
5. Ngo,T.T.M., Yoo,J., Dai,Q., Zhang,Q., He,C., Aksimentiev,A. and Ha,T. (2016) Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.*, **7**, 10813.
6. Yoo,J., Kim,H., Aksimentiev,A. and Ha,T. (2016) Direct evidence for sequence-dependent attraction between double-stranded DNA controlled by methylation. *Nat. Commun.*, **7**, 11045.
7. Leng,M. and Felsenfeld,G. (1966) The preferential interactions of polylysine and polyarginine with specific base sequences in DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **56**, 1325–1332.
8. Hoopes,B.C. and McClure,W.R. (1981) Studies on the selectivity of DNA precipitation by spermine. *Nucleic Acids Res.*, **9**, 5493–5504.
9. Raspaud,E., De La Cruz,M.O., Sikorav,J.-L. and Livolant,F. (1998) Precipitation of DNA by polyamines: a polyelectrolyte behavior. *Biophys. J.*, **74**, 381–393.
10. DeRouchey,J., Hoover,B. and Rau,D.C. (2013) A comparison of DNA compaction by arginine and lysine peptides: a physical basis for arginine rich protamines. *Biochemistry*, **52**, 3000–3009.
11. Besteman,K., Van Eijk,K. and Lemay,S. (2007) Charge inversion accompanies DNA condensation by multivalent ions. *Nat. Phys.*, **3**, 641–644.
12. Lipfert,J., Doniach,S., Das,R. and Herschlag,D. (2014) Understanding nucleic Acid–Ion interactions. *Annu. Rev. Biochem.*, **83**, 813–841.
13. Deng,H., Bloomfield,V.A., Benevides,J.M. and Thomas,G.J. Jr (2000) Structural basis of polyamine–DNA recognition: spermidine and spermine interactions with genomic B-DNAs of different GC content probed by Raman spectroscopy. *Nucleic Acids Res.*, **28**, 3379–3385.
14. Childs,A.C., Mehta,D.J. and Gerner,E.W. (2003) Polyamine-dependent gene expression. *Cell. Mol. Life Sci. CMLS*, **60**, 1394–1406.
15. DeRouchey,J.E. and Rau,D.C. (2011) Role of amino acid insertions on intermolecular forces between arginine peptide condensed DNA helices implications for Protamine-DNA packaging in sperm. *J. Biol. Chem.*, **286**, 41985–41992.
16. Todd,B.A., Adrian Parsegian,V., Shirahata,A., Thomas,T.J. and Rau,D.C. (2008) Attractive forces between cation condensed DNA double helices. *Biophys. J.*, **94**, 4775–4782.
17. Rau,D.C., Lee,B. and Parsegian,V.A. (1984) Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel DNA double helices. *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 2621–2625.
18. Rau,D.C. and Parsegian,V.A. (1992) Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. Evidence for long range attractive hydration forces. *Biophys. J.*, **61**, 246–259.
19. Das,R., Mills,T.T., Kwok,L.W., Maskel,G.S., Millett,I.S., Doniach,S., Finkelstein,K.D., Herschlag,D. and Pollack,L. (2003) Counterion distribution around DNA probed by solution X-Ray scattering. *Phys. Rev. Lett.*, **90**, 188103.
20. Qiu,X., Andresen,K., Kwok,L.W., Lamb,J.S., Park,H.Y. and Pollack,L. (2007) Inter-DNA attraction mediated by divalent counterions. *Phys. Rev. Lett.*, **99**, 038104.
21. Korolev,N., Lyubartsev,A.P., Laaksonen,A. and Nordenskiöld,L. (2003) A molecular dynamics simulation study of oriented DNA with polyamine and sodium counterions: diffusion and averaged binding of water and cations. *Nucleic Acids Res.*, **31**, 5971–5981.
22. Raspaud,E., Olvera de la Cruz,M., Sikorav,J.L. and Livolant,F. (1998) Precipitation of DNA by polyamines: a polyelectrolyte behavior. *Biophys. J.*, **74**, 381–393.
23. Cruz,M.O.D.L., Belloni,L., Delsanti,M., Dalbiez,J.P., Spalla,O. and Drifford,M. (1995) Precipitation of highly charged polyelectrolyte solutions in the presence of multivalent salts. *J. Chem. Phys.*, **103**, 5781–5791.
24. Tolokh,I.S., Pabit,S.A., Katz,A.M., Chen,Y., Drozdetski,A., Baker,N., Pollack,L. and Onufriev,A.V. (2014) Why double-stranded RNA resists condensation. *Nucleic Acids Res.*, **42**, 10823–10831.
25. Katz,A.M., Tolokh,I.S., Pabit,S.A., Baker,N., Onufriev,A.V. and Pollack,L. (2017) Spermine condenses DNA, but not RNA duplexes. *Biophys. J.*, **112**, 22–30.
26. Nathan,D. and Crothers,D.M. (2002) Bending and flexibility of methylated and unmethylated EcoRI DNA1. *J. Mol. Biol.*, **316**, 7–17.
27. Pérez,A., Castellazzi,Chiara L., Battistini,F., Collinet,K., Flores,O., Deniz,O., Ruiz,Maria L., Torrents,D., Eritja,R., Soler-López,M. *et al.* (2012) Impact of methylation on the physical properties of DNA. *Biophys. J.*, **102**, 2140–2148.
28. Conwell,C.C., Vilfan,I.D. and Hud,N.V. (2003) Controlling the size of nanoscale toroidal DNA condensates with static curvature and ionic strength. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 9296–9301.
29. Luckel,F., Kubo,K., Tsumoto,K. and Yoshikawa,K. (2005) Enhancement and inhibition of DNA transcriptional activity by spermine: A marked difference between linear and circular templates. *FEBS Lett.*, **579**, 5119–5122.
30. Sarhan,S. and Seiler,N. (1989) On the subcellular localization of the polyamines. *Biol. Chem. Hoppe-Seyler*, **370**, 1279–1284.
31. Kumar,S., Rosenberg,J.M., Bouzida,D., Swendsen,R.H. and Kollman,P.A. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, **13**, 1011–1021.
32. Souaille,M. and Roux,B.T. (2001) Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.*, **135**, 40–57.
33. Yoo,J. and Aksimentiev,A. (2016) The structure and intermolecular forces of DNA condensates. *Nucleic Acids Res.*, **44**, 2036–2046.
34. Schellman,J.A. and Parthasarathy,N. (1984) X-ray diffraction studies on cation-collapsed DNA. *J. Mol. Biol.*, **175**, 313–329.
35. Hud,N.V. and Polak,M. (2001) DNA–cation interactions: the major and minor grooves are flexible ionophores. *Curr. Opin. Struct. Biol.*, **11**, 293–301.
36. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251.
37. Drew,H.R. and Dickerson,R.E. (1981) Structure of a B-DNA dodecamer: III. Geometry of hydration. *J. Mol. Biol.*, **151**, 535–556.
38. Young,M.A., Jayaram,B. and Beveridge,D.L. (1997) Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: Fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.*, **119**, 59–69.

39. Li,L., Pabit,S.A., Meisburger,S.P. and Pollack,L. (2011) Double-stranded RNA resists condensation. *Phys. Rev. Lett.*, **106**, 108101.
40. Danilowicz,C., Lee,C.H., Kim,K., Hatch,K., Coljee,V.W., Kleckner,N. and Prentiss,M. (2009) Single molecule detection of direct, homologous, DNA/DNA pairing. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 19824–19829.
41. Kornyshev,A.A. and Leikin,S. (2001) Sequence recognition in the pairing of DNA duplexes. *Phys. Rev. Lett.*, **86**, 3666–3669.
42. Baldwin,G.S., Brooks,N.J., Robson,R.E., Wynveen,A., Goldar,A., Leikin,S., Seddon,J.M. and Kornyshev,A.A. (2008) DNA double helices recognize mutual sequence homology in a protein free environment. *J. Phys. Chem. B*, **112**, 1060–1064.
43. Gladyshev,E. and Kleckner,N. (2014) Direct recognition of homology between double helices of DNA in Neurospora crassa. *Nat. Commun.*, **5**, 3509.
44. Burgess,S.M., Kleckner,N. and Weiner,B.M. (1999) Somatic pairing of homologs in budding yeast: existence and modulation. *Genes Dev.*, **13**, 1627–1641.
45. Strom,A.R., Emelyanov,A.V., Mir,M., Fyodorov,D.V., Darzacq,X. and Karpen,G.H. (2017) Phase separation drives heterochromatin domain formation. *Nature*, **547**, 241.
46. Larson,A.G., Elnatan,D., Keenen,M.M., Trnka,M.J., Johnston,J.B., Burlingame,A.L., Agard,D.A., Redding,S. and Narlikar,G.J. (2017) Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*, **547**, 236.
47. Tang,S.-J. (2017) Potential role of phase separation of repetitive DNA in chromosomal organization. *Genes*, **8**, 279.
48. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
49. Cisse,I.I., Kim,H. and Ha,T. (2012) A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.*, **19**, 623–627.
50. Vafabakhsh,R. and Ha,T. (2012) Extreme bendability of DNA less than 100 base pairs long revealed by Single-Molecule cyclization. *Science*, **337**, 1097–1101.
51. Zoli,M. (2016) J-factors of short DNA molecules. *J. Chem. Phys.*, **144**, 214104.
52. Douarche,N. and Cocco,S. (2005) Protein-mediated DNA loops: Effects of protein bridge size and kinks. *Phys. Rev. E*, **72**, 061902.
53. Le,T.T. and Kim,H.D. (2014) Probing the elastic limit of DNA bending. *Nucleic Acids Res.*, **42**, 10786–10794.
54. Imakaev,M., Fudenberg,G., McCord,R.P., Naumova,N., Goloborodko,A., Lajoie,B.R., Dekker,J. and Mirny,L.A. (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods*, **9**, 999–1003.
55. Wang,S., Su,J.-H., Beliveau,B.J., Bintu,B., Moffitt,J.R., Wu,C.-T. and Zhuang,X. (2016) Spatial organization of chromatin domains and compartments in single chromosomes. *Science*, **353**, 598–602.
56. Korolev,N., Yu,H., Lyubartsev,A.P. and Nordenskiöld,L. (2014) Molecular dynamics simulations demonstrate the regulation of DNA–DNA attraction by H4 histone tail acetylations and mutations. *Biopolymers*, **101**, 1051–1064.
57. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
58. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E., Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817–3829.
59. Lindorff-Larsen,K., Piana,S., Palmo,K., Maragakis,P., Klepeis,J.L., Dror,R.O. and Shaw,D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct. Funct. Bioinformatics*, **78**, 1950–1958.
60. Hornak,V., Abel,R., Okur,A., Strockbine,B., Roitberg,A. and Simmerling,C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct. Funct. Bioinformatics*, **65**, 712–725.
61. Nerenberg,P.S. and Head-Gordon,T. (2011) Optimizing protein−solvent force fields to reproduce intrinsic conformational preferences of model peptides. *J. Chem. Theory Comput.*, **7**, 1220–1230.
62. Joung,I.S. and Cheatham,T.E. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, **112**, 9020–9041.
63. Dai,L., Mu,Y., Nordenskiöld,L. and van der Maarel,J.R.C. (2008) Molecular dynamics simulation of Multivalent-Ion mediated attraction between DNA molecules. *Phys. Rev. Lett.*, **100**, 118301.
64. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
65. Yoo,J. and Aksimentiev,A. (2012) Improved parametrization of Li+, Na+, K+, and Mg2+ ions for all-atom molecular dynamics simulations of nucleic acid systems. *J. Phys. Chem. Lett.*, **3**, 45–50.
66. Yoo,J. and Aksimentiev,A. (2016) Improved parameterization of amine–carboxylate and amine–phosphate interactions for molecular dynamics simulations using the CHARMM and AMBER force fields. *J. Chem. Theory Comput.*, **12**, 430–443.
67. Hess,B., Kutzner,C., Van Der Spoel,D. and Lindahl,E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.
68. Nosé,S. and Klein,M. (1983) Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, **50**, 1055–1076.
69. Hoover,W.G. (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*, **31**, 1695.
70. Parrinello,M. and Rahman,A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182–7190.
71. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
72. Miyamoto,S. and Kollman,P.A. (1992) SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.*, **13**, 952–962.
73. Hess,B., Bekker,H., Berendsen,H.J. and Fraaije,J.G. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
74. Case,D.A., Cerutti,D.S., Cheatham,T.E. III, Darden,T.A., Duke,R.E., Giese,T.J., Gohlke,H., Goetz,A.W., Greene,D., Homeyer,N. *et al.* (2017) *AMBER 2017*. University of California, San Francisco.
75. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.