Published online 30 October 2021 *Nucleic Acids Research, 2021, Vol. 49, No. 20* ***11459–11475***
https://doi.org/10.1093/nar/gkab967

DNA sequence and methylation prescribe the inside-out conformational dynamics and bending energetics of DNA minicircles

Jejoong Yoo [1,*], Sangwoo Park[3], Christopher Maffeo [2], Taekjip Ha[3,4,*] and Aleksei Aksimentiev [2,5,*]

[1]Department of Physics, Sungkyunkwan University, Suwon 16419, 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]Department of Biophysics and Biophysical Chemistry, Johns Hopkins University, Baltimore, MD 21205, USA, [4]Howard Hughes Medical Institute, Baltimore, MD 21218, USA and [5]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received November 23, 2020; Revised September 27, 2021; Editorial Decision October 04, 2021; Accepted October 11, 2021

ABSTRACT

Eukaryotic genome and methylome encode DNA fragments' propensity to form nucleosome particles. Although the mechanical properties of DNA possibly orchestrate such encoding, the definite link between 'omics' and DNA energetics has remained elusive. Here, we bridge the divide by examining the sequence-dependent energetics of highly bent DNA. Molecular dynamics simulations of 42 intact DNA minicircles reveal that each DNA minicircle undergoes inside-out conformational transitions with the most likely configuration uniquely prescribed by the nucleotide sequence and methylation of DNA. The minicircles' local geometry consists of straight segments connected by sharp bends compressing the DNA's inward-facing major groove. Such an uneven distribution of the bending stress favors minimum free energy configurations that avoid stiff base pair sequences at inward-facing major grooves. Analysis of the minicircles' inside-out free energy landscapes yields a discrete worm-like chain model of bent DNA energetics that accurately account for its nucleotide sequence and methylation. Experimentally measuring the dependence of the DNA looping time on the DNA sequence validates the model. When applied to a nucleosome-like DNA configuration, the model quantitatively reproduces yeast and human genomes' nucleosome occupancy. Further analyses of the genome-wide chromatin structure data suggest that DNA bending energetics is a fundamental determinant of genome architecture.

INTRODUCTION

Living cells have evolutionarily developed proteins that bend and twist DNA to perform biological functions such as gene regulation (1,2). One outstanding example is the eukaryotic nucleosome (3), where a 50 nm fragment of double-stranded DNA (dsDNA) is tightly wrapped around an 8 nm diameter protein core. When nucleosomes were first identified back in 1974 (3,4), Crick and Klug expected the double-helical structure of the nucleosomal DNA to contain local distortions—kinks (5)—because the nucleosome's radius was found to be an order of magnitude smaller than the DNA's persistence length (50 nm) (6). Defying the expectations, the crystal structure of a nucleosome revealed the nucleosomal DNA to locally maintain its double-helical structure, bending around the protein core, and no evidence of the kinks (7,8).

The discovery of the highly-bent structure of nucleosomal DNA raised the possibility that the sequence-dependent mechanical properties of DNA prescribe which regions of the eukaryotic genomes are more susceptible to nucleosome formation, i.e. nucleosome occupancy (9,10). Indeed, experiments have found the nucleosome occupancy to depend on the overall GC content (10,11) and the presence of AA, TT, or TA dinucleotides at the inward-facing minor grooves of the nucleosomal DNA (12–14). Further, it was found that methylation of DNA lowers the nucleosome occupancy experimentally (15,16) and computationally (17).

The relationship between the sequence of a dsDNA fragment and its propensity to forming highly-bent states was probed by DNA cyclization assays (18,19), which quantified the probability of spontaneous minicircle formation by a linear DNA construct containing short

*To whom correspondence should be addressed. Tel: +1 217 333 6495; Fax: +1 866 467 5398; Email: aksiment@illinois.edu
Correspondence may also be addressed to Jejoong Yoo. Email: jejoong@skku.edu
Correspondence may also be addressed to Taekjip Ha. Email: tjha@jhu.edu

complementary single-strand overhangs. The application of the single-molecule fluorescence resonance energy transfer (smFRET) technique to the DNA cyclization assay (15,20,21) revealed an orders-of-magnitude dependence of the cyclization (looping) time on the sequence (20) and epigenetic state (15) of short (100 bp or less) dsDNA fragments. The causal relationship between the DNA looping times and the nucleosome stability was established directly through measurements of the force required to unwrap individual nucleosomes (21).

While it is clear that both the nucleotide sequence and the epigenetic modifications influence the energetics of cyclized DNA, the kinetics of DNA looping, and nucleosome occupancy, the exact manner in which these effects come about remains undetermined. Several elastic models of DNA were developed with the sequence-dependent stiffness parameters derived from either molecular dynamics (MD) simulations of short DNA fragments (22–28), cyclization experiments of periodic DNA sequences (29), or a database of structures (30–32). Models that emphasize the role of kinks were developed as well (33–35).

Here, we quantitatively connect the nucleotide sequence and methylation to nucleosome occupancy data through a discrete worm-like chain (WLC) model of DNA derived from a massive all-atom MD simulation of DNA minicircles. In contrast to previous MD studies that characterized near-equilibrium elasticity of short (10–20 base pairs) DNA fragments, we characterized the sequence-dependent stiffness of highly-bent DNA, obtaining stiffness constants for DNA states realized in the nucleosomes. We find the nucleotide sequence of a DNA minicircle to precisely define which minor groove of the DNA is facing the center of the minicircle—i.e. potentially making contact with the protein core in an assembled nucleosome—with small changes to the sequence or methylation causing major conformational transitions. By decomposing the free energy landscapes of many minicircles into dinucleotide step energies, we develop a discrete WLC model that reproduces experimental DNA looping times and the nucleosome occupancy data. Finally, we applied our model to characterize DNA bending energies near the transcription start sites and across entire chromosomes in eukaryotes, which revealed a potential mechanism for controlling the gene activity by both sequence and methylation.

MATERIALS AND METHODS

Molecular dynamics simulations

General simulation protocol. All MD simulations were carried out in a constant-temperature/constant-pressure ensemble with a 2-fs time step using the Gromacs 4.5.5 package (36). The temperature was kept constant at 300 K using the Nosé-Hoover scheme (37,38). The pressure was coupled to 1 bar using the Parrinello-Rahman scheme (39). Van der Waals (vdW) forces were evaluated using a 10–12 Å switching scheme. The particle-Mesh Ewald (PME) summation scheme (40) was employed to compute long-range electrostatic forces using a 1.5-Å grid spacing and a 12-Å real-space cutoff. Covalent bonds to hydrogen in non-water and in water molecules were constrained using LINCS (41) and SETTLE (42) algorithms, respectively. AMBER ff99-based force fields (43) were used for the molecular mechanics models of DNA molecules in aqueous solutions: bsc0 parameters for unmodified DNA (44) and methylated cytosine, NaCl parameters of Joung and Cheatham (45), the TIP3P water model (46). Custom NBFIX corrections (CUFIX) were applied to the vdW interactions for cation–DNA phosphate and cation–anion pairs (47,48).

Construction of ideal DNA minicircles. For a given 90-bp sequence, a corresponding double-stranded DNA structure of strand I and J with nine turns was created using the nucleic acid builder (NAB) module of the AMBER package (49). To prepare a 90-bp DNA minicircle of linking number of 8 (Lk = 8), the 90 bp construct was first untwisted by one turn of the DNA through a homogeneous coordinate transformation. Following that, the linear dsDNA fragment was repositioned to have the axis of the dsDNA parallel to the z-axis and the DNA center of mass located at $(0,\ R,\ 0)$, where $R\,(=\ 0.34\frac{\text{nm}}{\text{bp}} \times 90\frac{\text{bp}}{2\pi} \approx 4.9\ \text{nm})$ is the radius of curvature of the target minicircle. The linear dsDNA was converted to a minicircle by applying the following transformations to each atom: translation by $-z$ along the z-axis and rotation by $\theta\ =\ z/R$ radian with respect to x-axis; z denotes the z-coordinate of the atom before the transformations. This procedure produced an ideal minicircle of a ∼4.9 nm radius of curvature and homogeneous dinucleotide bending angle of about 4°. Custom-made Perl scripts were used to implement the procedure mentioned above. Reproducing experimental conditions where DNA minicircles were synthesized without any nicks with the help of a ligase enzyme, we covalently linked the first and the last nucleotide of each strand. This all-atom model of an ideal minicircle was placed in a rectangular box of pre-equilibrated water molecules, arranging the minicircle to be parallel to the y–z plane. Sodium and chloride ions were randomly added to produce an electrically neutral system containing 1 M NaCl solution (15,20).

To efficiently utilize available computational resources, we used the following two methods to carry out free-equilibration simulations. In the first method, the simulation systems contained about 140 000 atoms and measured $\sim 8.5 \times 13.0 \times 13.0\ \text{nm}^3$; the DNA minicircles diffused freely about the simulation volume. In the second method, the simulation systems contained about 95 000 atoms and measured $6.0 \times 13.0 \times 13.0\ \text{nm}^3$; the DNA minicircle was initially placed parallel to the y–z plane. This orientation of the minicircle was maintained by subjecting all heavy atoms of DNA to two planar half-harmonic potentials (50), parallel to the y-z plane, that turn on at $x\ =\ \pm 3$ nm, $U(x)$:

$$U\,(x) = \begin{cases} 0.5k(x - 3\,\text{nm})^2, & x \geq 3\,\text{nm} \\ 0, & |x| < 3\,\text{nm} \\ 0.5k(x + 3\,\text{nm})^2, & x \leq -3\,\text{nm} \end{cases}$$

where the force constant $k\ =\ 1000\ \text{kJ mol}^{-1}\ \text{nm}^{-2}$. Method 1 was used for the microseconds-long simulations of NGO, NGOme, NGOat and NGOgc minicircles (Figure 2D and Supplementary Figure S3A,D,G). Method 2 was used for all other equilibration simulations. For the four NGO-variant minicircles, method 2 produced the same

equilibrium poloidal angles as method 1 (Figure 2E and Supplementary Figure S3B, E, H), which directly confirmed our intuitive expectation that the planar half-harmonic restraints do not affect the internal conformational energy and geometry of the minicircle.

Equilibration strategy for DNA minicircles. A standard approach to equilibrating the initial conformation of a DNA minicircle would have been to run an equilibrium MD simulation applying position restraints to atoms of DNA using the coordinates of the ideal minicircle as a reference. However, this approach is not the best way to relax the minicircle because the conformation of an ideal minicircle itself is far from the equilibrium one. Thus, we developed a custom equilibration strategy, where 90 hydrogen bonds (one per base pair) were reinforced, and no position restraints were used. Specifically, we added a half-harmonic restraint potential between the N1 and N3 atom pair of each base pair, a restraint that applies a non-zero force only when the N1–N3 distance becomes greater than 4 Å; the force constant for the half-harmonic restraint was 1000 kJ mol^{-1} nm^{-2}. Conveniently, this type of restraint potential has already been implemented in the Gromacs package as the 'restraint potential' of bondtype 10; see Gromacs manual for more details. Note that, for the majority of the time, these restraints apply zero force because the average N1–N3 distance is below 3 Å. This equilibration strategy has allowed us to fully relax the global conformation without putting the structural integrity of a minicircle at risk. These N1–N3 restraint potentials were applied during the first 200 ns of each equilibration simulation, regardless of the system's geometry or the presence of the planar restraining potentials.

Mathematical definition of poloidal angle ϕ. When a minicircle stably maintains a toroid-like geometry, the rotation with respect to the helical axis of dsDNA, i.e. the rotation in the poloidal angle space, occurs due to thermal fluctuations. Because each 90-bp DNA minicircle undergoes a highly correlated thermal motion in the poloidal angle space, an arbitrary position on the minicircle surface can be used as a reference point for monitoring the poloidal rotational motion. For convenience, we chose the phosphorus atom of the first nucleotide of the initially linear strand I's sequence (P1) as a reference point; see Supplementary Table S1 for the list of the DNA sequences. As a measure of the rotation, we defined the poloidal angle, ϕ, using P1 as follows. First, the center of mass of the 11-bp helical segment with the P1-containing base pair at the segment's center, O, was computed. Second, the center of mass of the entire minicircle, X, was computed. Third, two unit vectors from O to X, $\vec{u}$, and from O to P1, $\vec{v}$, were determined. Last, the poloidal angle was computed as the angle between $\vec{u}$ and $\vec{v}$: $\phi = A\cos^{-1}(\vec{u}\cdot\vec{v})$ where $A = +1$ when $\vec{v}\times\vec{u}$ is in the 5′ to 3′ direction of strand I and $A = -1$ otherwise.

Poloidal angle constraints for the measurements of the torque. By definition, $\phi = 0$ when the distance between point P1 and the plane of the minicircle (the gray plane in Figure 2C) is zero and P1 is pointing toward the center of the minicircle. In other words, one can maintain ϕ near zero by constraining P1 to the plane of the minicircle. Similarly, one can constrain P2 (the second phosphorus in the sequence of strand I) to the plane of the minicircle, resulting in $\phi = 32°$. By changing the constraining phosphorus atom from P1 to P11 (indicated by the colored spheres in Figure 2C), one can monotonically increase ϕ from 0° to 352° with an increment of 32°. For convenience and only for the simulations of the effective torque, we maintained the plane of the minicircle parallel to the x-y plane by harmonically restraining the z coordinate of the center of mass of the following three 11-bp segments equally spaced along the minicircle: segment 1, $s = (86, \cdots, 90, 1, \cdots, 6)$; segment 2, $s = (26, \cdots, 36)$; and segment 3, $s = (56, \cdots, 66)$. Finally, the z coordinate of one of P1 to P11 atoms was harmonically restrained to the minicircle plane, $z = 0$. For all restraints, the harmonic constant was 1000 kJ mol^{-1} nm^{-2}. Because the harmonic restraint on a phosphorus atom applies a tangential force in the poloidal direction, the applied torque τ equals the average restraining force multiplied by 9 Å, which is the distance between the local helical axis and the phosphorus atom. For each free energy landscape, we computed $\tau(\phi)$ for ϕ ranging from 0° to 352° with an increment of 32° (e.g. Figure 3A). The free energy landscape, ΔG, was computed by integrating $-\tau(\phi)$ with respect to ϕ using the trapezoidal scheme (e.g. Figure 3B):

$$\Delta G(\phi) = -\int_0^{\phi} \tau(\phi)\, d\phi. \quad (1)$$

Frenet-Serret analysis of the local bending at a dinucleotide step. To analyze the deformation of dsDNA by bending, we employed the Frenet-Serret formalism, which was developed using methods of differential geometry and generalized to describe the behaviors of polymers (51) and nanorods (52). Following the Frenet-Serret formalism, we represent the DNA minicircle using a contour connecting the center of masses of base pairs $s = 1, \cdots, 90$, and orthonormal vector triads, $\{\hat{u}_{i,s} \mid i = 1, 2, 3\}$, each of which is attached to the center of mass of a base pair s. For a base pair s, $\hat{u}_{3,s}$ is the normal unit vector of the least-squares plane approximating the plane formed by the heavy atoms of the base pair, N1, C2, N3, C4, C5, C6, N7, C8, and N9. The direction of $\hat{u}_{3,i}(s)$ is the same as the 5′ to 3′ direction of strand I. The unit vector $\hat{u}_{1,s}$ is from the center of mass of strand I's nucleotide to the center of mass of strand J's nucleotide of the base pair s, projected onto the least-squares plane of the base pair. Then, $\hat{u}_{2,s} = \hat{u}_{3,s} \times \hat{u}_{1,s}$. See Supplementary Figure S10A for the illustration. Having defined the orthonormal triad for each base pair, we used the generalized Frenet-Serret formula (51) to quantitatively characterize local structural deformations by measuring the generalized torsions between $\hat{u}_{i,s}$ and $\hat{u}_{i,s+1}$, $\omega_{i,s}$:

$$(\hat{u}_{i,s+1} - \hat{u}_{i,s})/\Delta s = \in_{ijk} \omega_{j,s} \hat{u}_{k,s} \quad (2)$$

where $\in_{ijk}$ is the Levi-Civita tensor and $\Delta s = 3.4$ Å. The $\omega_{1,s}$ and $\omega_{2,s}$ torsions characterize the two bending angles at the s, $s+1$ dinucleotide step whereas $\omega_{3,s}$ indicates the twist angle. Under a harmonic approximation, the bending

energy per step between s and $s + 1$

$$\bar{E}_s = \frac{k_B T}{2}\,\Delta s\left\{\alpha_1\left(n_{s,s+1}\right)\left(\frac{\omega_{1,s}}{\Delta s}\right)^2 + \alpha_2\left(n_{s,s+1}\right)\left(\frac{\omega_{2,s}}{\Delta s}\right)^2\right\}, \tag{3}$$

where α_1 and α_2 are the sequence-dependent bending moduli that depend on the type of the dinucleotide step, $n_{s,s+1}$, and $k_B T$ is the Boltzmann factor. Note that the persistence length, l_p, is related to the bending moduli as

$$\frac{2}{l_p} = \frac{1}{\alpha_1} + \frac{1}{\alpha_2} \tag{4}$$

for a homogeneous polymer with bending moduli α_1 and α_2 (53). When ω_1 and ω_2 are highly correlated, $\omega_2 \approx -0.3\omega_1$, $\bar{E}_s$ can be simplified to

$$\bar{E}_s = K\left(n_{s,s+1}\right)\omega_{1,s}{}^2 \tag{5}$$

where

$$K\left(n_{s,s+1}\right) = \frac{k_B T}{2\Delta s}\left\{\alpha_1\left(n_{s,s+1}\right) + 0.09\alpha_2\left(n_{s,s+1}\right)\right\} \tag{6}$$

is the sequence-dependent stiffness constant. The computer code for the Frenet-Serret analysis was written in C and implemented as a part of the Gromacs package; the code is available upon request.

Determination of the sequence-dependent stiffness constants. For a 90-bp minicircle, we found $\omega_{1,s}$ to show a periodic dependence along the DNA contour (Figure 3C, D and Supplementary Figure S10B, E) that can be approximated by a sine square function, regardless of the sequence and the poloidal angle value:

$$\omega_{1,s}^{\text{circle}}\left(\phi\right) = A\sin^2\left\{\frac{\pi}{P}s - \frac{\phi - 1.5}{2}\right\} \tag{7}$$

where $P = 11.25$ is the periodicity of a 90-bp minicircle with a linking number 8 and $A = 9^\circ$. Then, the bending energy of a step s, $\bar{E}_s(\phi)$, can be approximated for any given ϕ as

$$\bar{E}_s\left(\phi\right) = K\left(n_{s,s+1}\right)\left\{\omega_{1,s}^{\text{circle}}\left(\phi\right)\right\}^2 \tag{8}$$

where $K(n_{s,s+1})$ is the sequence-dependent stiffness constant from Eq. 6. The total bending energy of a minicircle, $E^{\text{circle}}(\phi)$, thus equals

$$E^{\text{circle}}\left(\phi\right) = \sum_{s=1}^{90}\bar{E}_s\left(\phi\right). \tag{9}$$

The expression for $E^{\text{circle}}(\phi)$ can be rewritten as a sum over all dinucleotide step types

$$E^{\text{circle}}\left(\phi\right) = \sum_{n=1}^{21} K_n W_{1,n}^{\text{circle}}\left(\phi\right) \tag{10}$$

where $W_{1,n}^{\text{circle}}(\phi)$ is the sum of $\omega_{1,s}^{\text{circle}}(\phi)^2$ among the dinucleotide steps of step type n and $\mathrm{K_n}$ the stiffness constant for the step of type n. By equating $E^{\text{circle}}(\phi)$ to the free energy landscapes, $\Delta G(\phi)$, determined from MD simulations via the torque integration method: $E^{\text{circle}}(\phi) - E^{\text{circle}}(0) = \Delta G(\phi) - \Delta G(0)$, we can obtain a system of linear equations with 21 variables, K_n. We determined the values of K_n by numerically solving this over-determined linear algebra equations using the NumPy package.

Coarse-grained simulations

Energy calculation using the DNABEND model. We obtained the six dinucleotide parameters (twist, roll, tilt, shift, slide, and rise) by analyzing the trajectories from our torque simulations using the 3DNA program (54). Then, we computed the energy at a given poloidal angle using the quadratic energy functions in Ref. (31).

General simulation protocols of 3SPN.2C for oxDNA models. Coarse-grained models of all DNA minicircles were constructed by leveraging a number of convenient functions in the MrDNA package (55). MrDNA simulations were performed, starting with an idealized initial configuration of each minicircle. For technical reasons, two equal-length double-helical regions were used to construct each circle, with the regions connected on 3′ and 5′ termini on both ends. The mathematical equation for a circle with a circumference equal to 3.4 **Å** × N_{bp} was evaluated at 100 points through which splines were fit. Similarly, a discrete sequence of rotations was applied to an orientation matrix representing the local basis of the helix so that the helical axis was always tangential to the circle and the remaining axes rotated about this exactly eight times over the contour length of the circle, yielding a linking number of eight. The resulting orientation matrices were converted into quaternions through which splines were fit. The appropriate sequence was assigned to the minicircle before configuration files, for either the 3SPN.2C (32,56) model or the oxDNA2 model (57), were generated by the package. The interaction sites and, for oxDNA, the orientation of each nucleotide were obtained by interpolating the splines and applying transforms to rotate and shift the idealized base pair coordinates. Support for the 3SPN.2C model was added to the MrDNA package to enable these simulations. Additional files describing harmonic restraints to hold base pairs intact were generated using the same script.

oxDNA model. First, simulations were performed using the sequence-specific oxDNA2 model (57) with 1000 steps of minimization followed by 10^7 steps of MD simulation using a timestep of 6.06 fs, an Anderson-like thermostat set to 291 K with velocities refreshed every 103 steps, the CUDA backend with mixed precision, and an ion concentration of 1 M. The simulation trajectory was written to a file every 1000 steps. Harmonic restraints were applied between nucleotides forming each basepair ($k = 57.09$ pN/nm; $r_0 = 1.28$ nm) to ensure that the bases remain intact.

3SPN.2C model. All simulations were performed using the 3SPN.2C model with a Langevin thermostat holding the temperature to 291 K, timestep of 20 fs, 18-Å short range interaction cutoff and 50-Å electrostatics cutoff. The source code of the USER-3SPN2 module for LAMMPS was modified to allow simulation of circular DNA strands,

although we were unable to avoid loss of cross-stacking interactions at the nucleotides forming the interface between nucleotides having the first and last index in each strand. A patch with the modifications to the USER-3SPN2 module with commit ID 2b87bf1 is provided as Supplementary Information. Two such sites occurred on opposite sides of each circle. Each circle was subject to 1000 steps of minimization followed by 10^6 steps of simulation. Weak harmonic restraints were placed between beads forming base pairs ($k = 10$ kJ/Å^2, $r_0 = 6.4$ Å) to ensure they remained intact. The 90-bp 3SPN.2C minicircles with a linking number of eight were generally observed to lose B-form helical structure in at least a portion of each minicircle, though this region was typically mobile. The loss of secondary structure was seen to migrate around the circle in most cases and was not, in general, colocalized with the ends of the strands, where cross-stacking interactions were neglected.

Single-molecule looping experiments

A single-molecule DNA cyclization assay was previously developed to measure the flexibility of dsDNA molecules (20). All DNA strands (90 bp) were prepared by ligating two synthesized DNA fragments (Integrated DNA Technologies) and purified in 14% urea–PAGE gel. After purification of ligated strands, the complementary strands were annealed to obtain dsDNA. The final DNA product was composed of an 80-bp double-strand core and 10-bp single-strand overhangs. The single-strand overhangs were labeled with Cy3 and Cy5 as FRET pair at each 5′ end to monitor the looping of individual molecules as base pairing of the overhangs. The prepared DNA was immobilized on a PEG-coated quartz slide through NeutrAvidin-biotin interaction. Single-molecule image acquisition was performed in the buffered solution (10 mM Tris–HCl pH 8.0, 10 mM or 1 M NaCl, 1% w/v D-glucose [Sigma], 150 U/ml glucose oxidase [Sigma], 2170 U/ml catalase [EMD Milipore] and 3 mM Trolox [Sigma]). Starting with low salt buffer (10 mM NaCl), in which most DNA did not make a loop, looping was initiated by an infusion of high salt buffer (1 M NaCl). Since looped and unlooped DNA molecules could be distinguishable from the corresponding high and low FRET signals, we were able to monitor an increasing number of molecules with a high FRET state over time. Through single exponential fittings of the fraction of looped molecules over time, the looping rate or looping time (the inverse of the looping rate) could be obtained. All measurements were performed at room temperature (∼23°C).

RESULTS

DNA minicircle as a model of nucleosomal DNA

X-ray crystallography (7) and statistical analysis of nucleosome occupancy in the yeast genome (12,13) have established that DNA in an assembled nucleosome has a preferential orientation characterized by a regular pattern of inward-facing nucleotides that make direct contacts with the protein (histone) core, Figure 1A and B. To test if the mechanical properties of DNA alone can produce such a specific pattern of contacts, we constructed two 85 base pair (bp) DNA minicircles representing the left (601LH) and the right (601RH) halves of the 601 sequence (58), Figure 1C,D. In doing so, we intentionally chose the initial configuration of the minicircles to have the phosphates, which are inward-facing in the crystal structure (yellow spheres), facing outward. Starting from these configurations, we performed an MD simulation of each minicircle submerged in 1 M NaCl solution, Figure 1E, using the parmbsc0 force field (43,44) and the CUFIX corrections (47) to ion-DNA interactions. See Methods for the simulation details. We chose to carry out our simulations at such high electrolyte concentration to reproduce the conditions of the DNA cyclization assays (20,21), where ions are thought to stabilize the cyclized DNA molecules in the absence of stabilizing proteins.

During the simulations, each DNA minicircle was observed to spontaneously rotate about its local helical axis, reaching an equilibrium orientation within 150 ns, Figure 1F and G and Supplementary Movies S1 and S2. Importantly, the DNA nucleotides that make contact with the histone core in the crystal structure (yellow spheres in Figure 1F, G) were facing, predominantly, inward also in the equilibrium configurations of the DNA minicircles despite the absence of histone proteins and the topological difference between the minicircle and the conformation of nucleosomal DNA (closed versus open loop). The agreement between the crystal structure and the equilibrium MD conformations was excellent for the internal stretches of the nucleosomal DNA and less so for the terminal domains, consistent with the experiment that found the internal domain of the 601 sequence to be the primary determinant of the 601 DNA bendability (21). Thus, we found the preferential orientation of the 601 DNA in the nucleosome to be determined predominantly by the DNA mechanics, which makes a DNA minicircle an excellent model of nucleosomal DNA.

Sequence and methylation uniquely define the inside-out (poloidal) orientation of a DNA minicircle

To further investigate the inside-out dynamics of DNA minicircles, we constructed an all-atom model of a 90-bp minicircle of the same sequence (referred to as NGO) as the DNA construct used in the previous smFRET cyclization study (15). The all-atom model was simulated unrestrained in 1 M NaCl solution for 11 microseconds, two orders of magnitude longer than the previous MD simulations of DNA minicircles (59–62). The minicircle maintained its overall toroid-like conformation for the whole duration of the simulation, Figure 2B and Supplementary Movie S3, developing no kinks or any other base pair-level distortions. Choosing the linking number of eight was critical for the stability of the 90-bp minicircle, as the same minicircle was observed to undergo supercoiling transition when built and simulated with the linking number of seven or nine, Supplementary Figure S1 and Movies S4 and S5.

To characterize the overall geometry of the minicircle, we converted the all-atom structure to a wireframe representation, a collection of lines that connect the centers of mass (CoM) of the neighboring base pairs (blue tubes in Figure 2A, B) (52). The wireframe representation makes it easy to see that the overall shape of the equilibrated DNA minicircle resembles more of an octagon than of a perfect circle. That is, the minicircle accommodates the bending

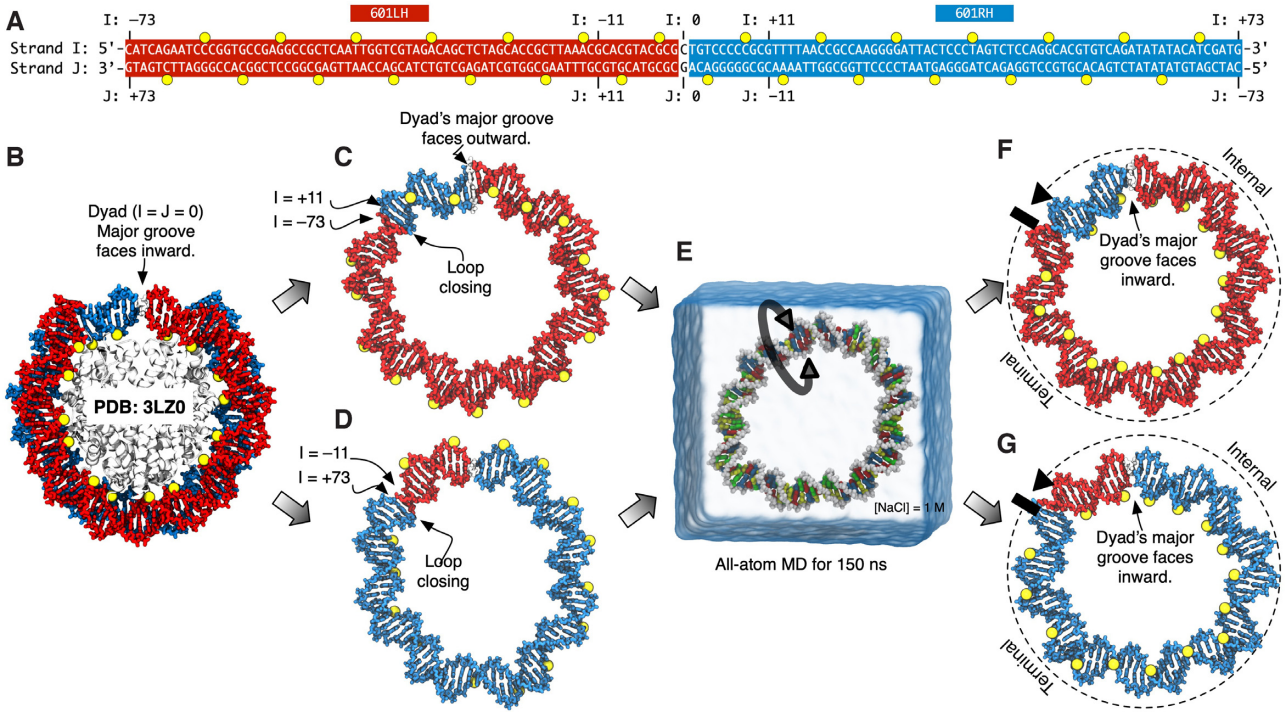


Figure 1. Spontaneous inside-out transition in DNA minicircles reproduces the configuration of nucleosomal DNA. (A, B) The sequence of the 601 nucleosome-forming DNA (**A**) (58) and the crystal structure of the assembled nucleosome (**B**), PDB ID 3LZ0 (8). Yellow spheres indicate the phosphate groups in contact with the histone core complex. The two halves of the 601 sequence, 601LH and 601RH, are highlighted in red and blue, respectively. (C, D) Two 85-base pair DNA minicircles containing the 601LH (**C**, $I = -73$ to $I = +11$) and 601RH (**D**, from $I = -11$ to $I = +73$) fragments of the 601 sequence. In these initial configurations, the phosphate groups (yellow spheres), which are in contact with the histone proteins in the crystal structure, were arranged to be at the outer face of the minicircles. (**E**) All-atom model of an 85-bp DNA minicircle submerged in a rectangular volume of 1 M NaCl solution (blue semi-transparent surface); ions are not shown for clarity. During equilibrium MD simulations (Supplementary Movies S1 and S2), the 601LH and 601RH minicircles maintain circular conformation but undergo spontaneous inside-out transition, locally twisting about the helical axis. (F, G) The final equilibrium configurations of the 601LH (**F**) and 601RH (**G**) minicircles. The yellow-colored phosphate groups are now located at the inner face of the minicircles, similar to the arrangement seen in the crystal structure.

stress unevenly throughout its structure, compressing the major groove at the inner surface of the minicircle at eight sites. Although the minicircle was found to undergo considerable conformational fluctuations throughout the simulation, the octagon-like geometry was omnipresent at any instance, Figure 2B and Supplementary Movie S3.

While maintaining its overall toroid-like conformation, the minicircle was observed to undergo spontaneous rotation with respect to its local helical axis, Figure 2B, similar to that observed for the 601LH and 601RH minicircles, Figure 1. To characterize this internal rotation quantitatively, we defined the poloidal angle, ϕ, of a DNA minicircle as the local rotation angle of a backbone phosphorus atom about the local helical axis with respect to a reference orientation, Figure 2C, see Methods mathematical definition and Supplementary Movie S6. Correlation analysis of the 90 poloidal angle trajectories (of each phosphorous atom of strand I) revealed a highly coordinated rotary motion, with all angle values changing in sync, Supplementary Figure S2A and B. Therefore, all subsequent analyses of the poloidal angle change were performed using coordinates of the phosphorous atom of the first nucleotide of strand I (red sphere in Figure 2A–C).

We hypothesized that the nucleotide sequence and the methylation state of a DNA minicircle encode its preferential value of the poloidal angle. Analysis of the 11 μs trajectory of the NGO minicircle, Figure 2D, indeed revealed one dominant population of the poloidal angle, although the angle was seen to sample the whole range (0–360°) of values. The same equilibrium poloidal angle was also observed at the end of six ∼300 ns simulations of the same minicircle, each beginning from a different initial value of ϕ, Figure 2E. Equilibrium simulations of 41 additional minicircles, Supplementary Figures S3–S9, revealed a complicated dependence of the equilibrium poloidal angle on the DNA sequence and methylation, with the equilibrium poloidal angle spanning the full range of values. Thus, our simulations of the $(XYtatatatat)_8$at minicircles, where XY nucleotides were AA, AT, TT, CC, CG, ^{m}CG, GC or GG, converged to equilibrium ϕ angles that varied dramatically depending on the choice of the XY nucleotides, Figure 3A–H and Supplementary Figure S6. Conversely, a minicircle of uniform sequence, $(ta)_{45}$, freely rotated without any equilibrium ϕ angle, Figure 3I and Supplementary Movie S7, confirming that the non-uniform distribution of the sequence indeed determines the equilibrium ϕ angle, Figure 3J. A similar variation of the equilibrium ϕ values was observed in the simulations of the seven $(XYcgcgcgcgc)_8$gc minicircles, Figure 3K and Supplementary Figure S7. Thus, our data show that the sequence and methylation of a DNA minicircle

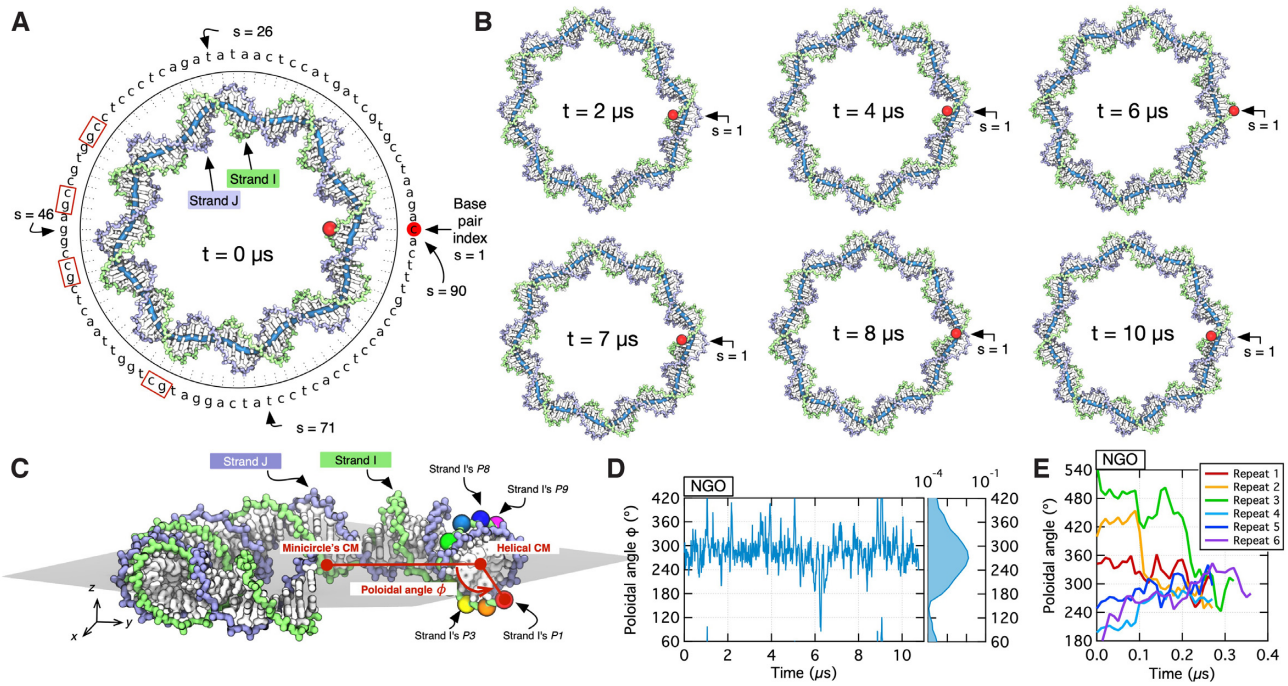


Figure 2. Sequence-dependent energetics of the inside-out transition in DNA minicircles. (**A**) Schematic of a DNA minicircle containing two complementary 90-nucleotide strands, I and J. The sequence of strand I is defined using a base pair index, *s*, that ranges from 1 to 90; the red dot indicates the nucleotide at s = 1. The specific sequence shown is the NGO construct (15). (**B**) Equilibrium, 11-μs MD simulation of the NGO minicircle. For visual comparison, molecular configurations were aligned with the initial configuration using coordinates of all non-hydrogen atoms. The blue wireframe connects the centers of mass of the neighboring base pairs. (**C**) Poloidal angle ϕ, defined as the angle between the line connecting the phosphorous atom of the first nucleotide of strand I, P1, to the center of the helix and the line connecting the center of the helix to the center of the minicircle. The gray surface indicates the plane of the minicircle. For clarity, a quarter of minicircle is not shown. (**D**) Poloidal angle as a function of simulation time for NGO minicircle. Each data point shows a 10-ns block average of 2-ps sampled instantaneous conformations. Histograms show normalized distributions of ϕ using a log scale, log $P(\phi)$. (**E**) Poloidal angle as a function of time in six independent simulations of the NGO minicircle started from different initial values.

programs which nucleotides are located at its inward and outward surfaces. Supplementary Table S1 provides a complete list of DNA sequences examined.

Dinucleotide stiffness analysis of the poloidal angle free energy landscapes

We determined the free energy of 26 DNA minicircles as a function of their poloidal angle by simulating each minicircle in eleven conformations corresponding to eleven poloidal angles equally spaced between 0 and 360°. The harmonic restraints that we used to maintain the prescribed poloidal angle reported on the effective torque required to maintain each orientation; see Methods for details. Figure 4A shows the resulting dependence of the mean torque on the poloidal angle for the NGO sequence. Integration of torque with respect to ϕ yielded the free energy profile of the DNA minicircle along the poloidal angle coordinate (black curve in Figure 4B). A quantitatively similar free energy profile was obtained using Boltzmann inversion (blue curve in Figure 4B), i.e. $-k_BT \log P(\phi)$, where $P(\phi)$ is the distribution of the ϕ values observed during the unrestrained equilibration, Figure 2D. Repeating the torque integration procedure for the 25 minicircles revealed a collection of rugged free energy landscapes, each characterized by a single minimum at the preferential angle, Figure 4C and Supplementary Figures S3-S9. The location of the minima of each free-energy landscape was found to be in excellent agreement (Pearson's correlation of 0.98) with the equilibrium poloidal angle observed in free-equilibration simulations, Figure 4D.

To check if existing models of DNA can reproduce the sequence dependence of the minicircle's free energy on the poloidal angle observed in our MD simulations, we performed coarse-grained simulations using either the oxDNA2 (35,57) or the 3SPN.2C model (32,56) and energy calculations using the DNABEND model (31); see Supplementary Methods for a detailed description. OxDNA minicircles composed of 85 bp were seen to be stable, whereas 90-bp oxDNA minicircles all developed kinks during the simulations (Supplementary Figure S10A, B) (63,64). The 3SPN.2C minicircles did not kink, although the B-DNA helical structure sometimes transiently unraveled (Supplementary Figure S10D). Despite those defects, the poloidal angle did not plateau during most simulations (Supplementary Figure S10E). The free energy profiles for select sequences obtained from oxDNA (Supplementary Figure S10C) and 3SPN (Supplementary Figure S10F) models differed significantly from the all-atom counterparts, resulting in a low correlation between the preferred poloidal angles in 3SPN.2C and all-atom MD simulations (Supplementary Figure S10G). The overall geometry of the minicircles was less circular in the 3SPN and oxDNA simulations than in all-atom MD. The eccentricity—the average ratio of the major and minor axes of an ellipse—for the all-atom MD,

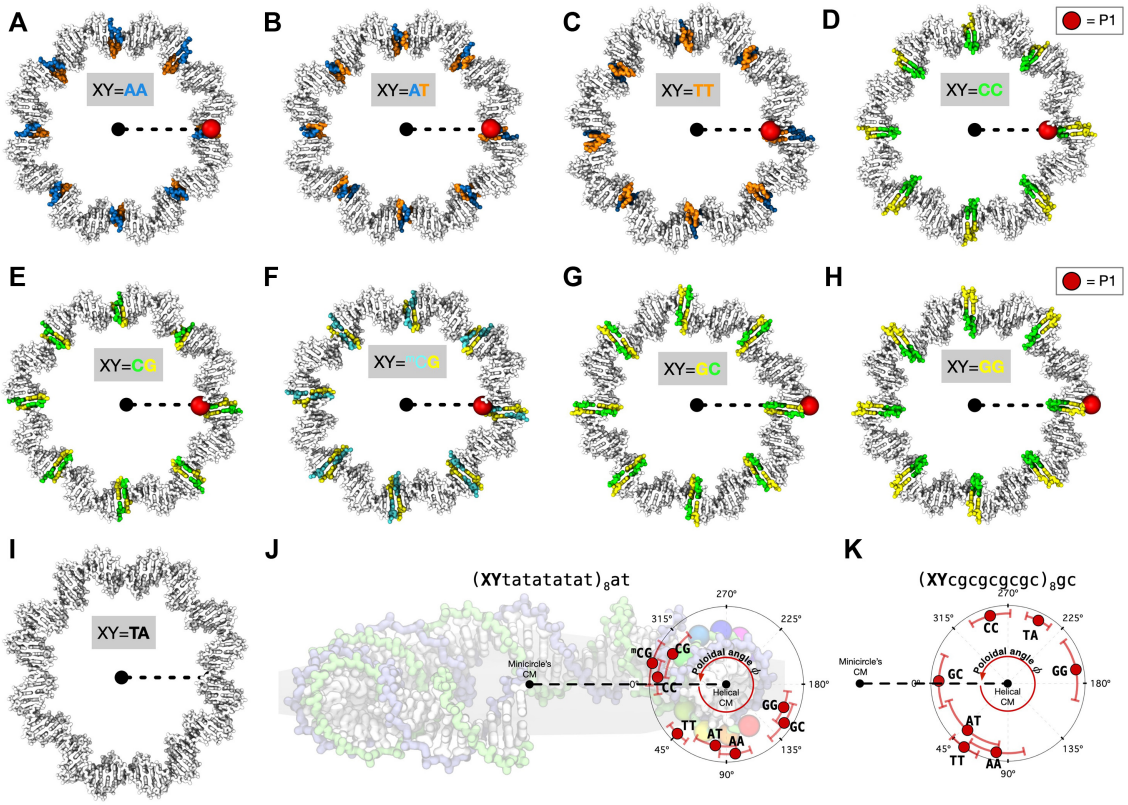
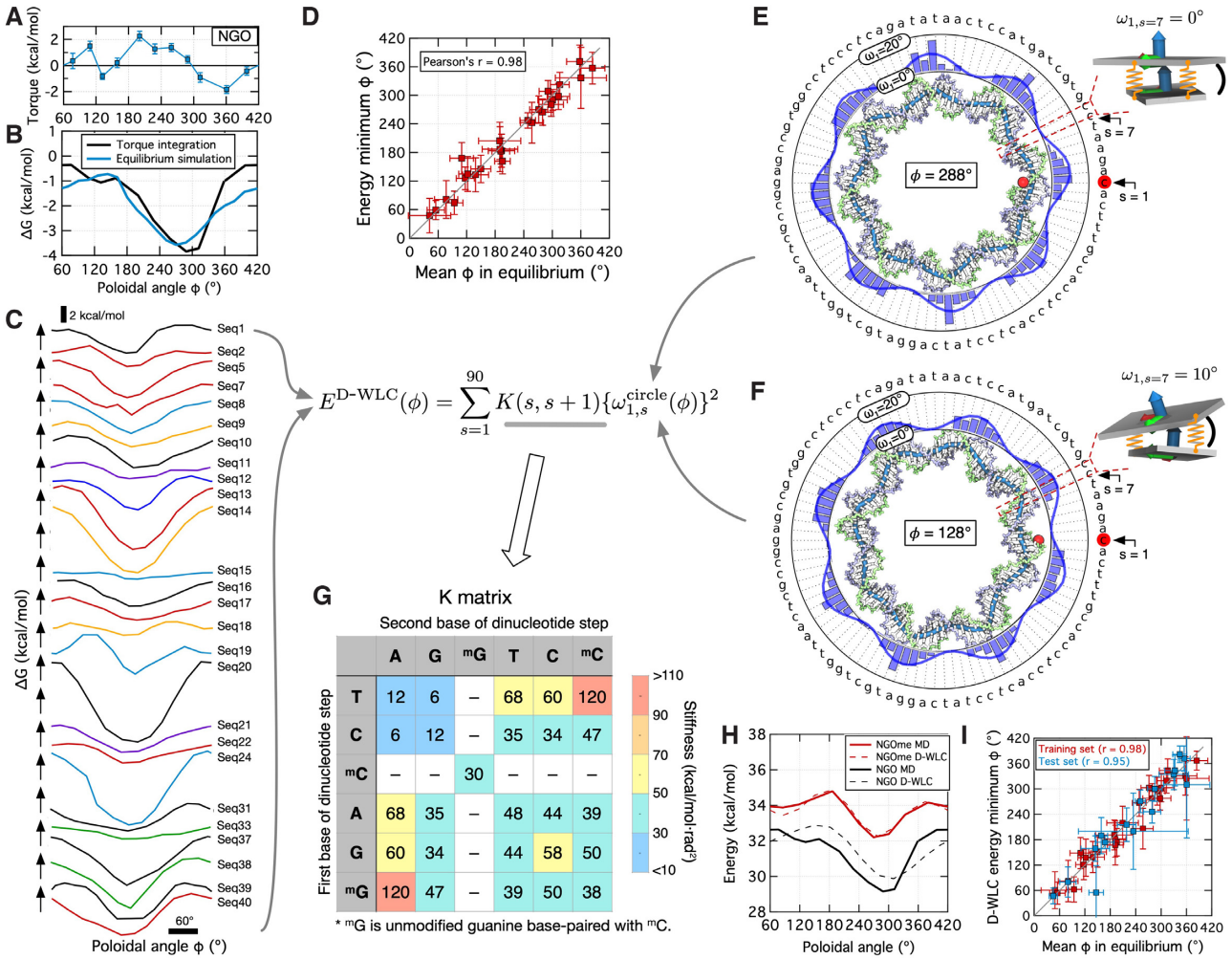


Figure 3. The preferential poloidal angle of $(XYtatatatat)_8at$ and $(XYcgcgcgcgc)_8at$ minicircles. (**A**–**I**) The equilibrium conformation of the $(XYtatatatat)_8at$ minicircles (ID 17–25). The location of P1 atom is indicated using a red sphere. See Supplementary Table S1 for nucleotide sequences. The type of X and Y nucleotides are color-coded as specified in the inset; all other nucleotides are shown in white. In panel I, the P1 atom is not shown because this particular sequence, $(ta)_{45}$, does not have any particular equilibrium poloidal angle. (**J**, **K**) Mean poloidal angle observed in unrestrained equilibration of $(XYtatatatat)_8at$ (panel J) and $(XYcgcgcgcgc)_8gc$ (panel K) minicircles, the type of nucleotides X and Y are specified at each data point. Error bars indicate the standard deviations of 10-ns block averages.

oxDNA, and 3SPN models was 1.19, 1.25, and 1.48, respectively, for the 85-bp 601RH minicircle and 1.15, 1.55, and 1.74, respectively, for the 90-bp 601RH minicircle. For the DNABEND model, the minimum-energy poloidal angles were also found to be in poor agreement with the preferred poloidal angle observed in all-atom MD simulations, Supplementary Figure S10H. Refitting the parameters of the oxDNA, 3SPN.2C or DNABEND models against our all-atom MD data may present an opportunity to refine further the description of the sequence-specific effects in those models.

Here, we account for the sequence and methylation dependence of the DNA minicircles' energy at the level of a discrete worm-like chain (WLC) model under zero intrinsic curvature approximation, following the works of Vologodskii and coworkers (29,65). Our discrete WLC model is an adaptation of the continuum WLC model (66,67), which has been widely used for the analysis of DNA conformational properties and elasticity (68–71). Our choice of the zero intrinsic curvature approximation was in part motivated by the outcome of the DNA cyclization assays that found the presence of a poly(dA:dT) tract to inhibit the cyclization of 100-bp (20) and 200-bp (72) dsDNA constructs, despite carrying a significant intrinsic curvature (73,74).

To fit the discrete WLC model to our free energy landscape data, we quantified the stepwise bending of DNA using the Frenet-Serret formalism, which was previously employed to describe the deformation of linear polymers (51), alpha-helices (53), and DNA (31,52). Briefly, we define an orthonormal triad of unit vectors $\hat{u}_{1,s}$, $\hat{u}_{2,s}$, and $\hat{u}_{3,s}$ attached to a base pair index s (Supplementary Figure S11A) and compute the generalized torsions for a step between base pair s and $s+1$, $\omega_{1,s}$, $\omega_{2,s}$ and $\omega_{3,s}$ (51,52), see Methods for detailed descriptions. For all DNA minicircles, two torsions that dictate the bending modes, $\omega_{1,s}$ and $\omega_{2,s}$, show a periodic oscillatory pattern regardless of ϕ. For example, the profiles of $\omega_{1,s}$ of the NGO minicircle (blue bars in Figure 4E) oscillate between 0° and 10° with eight peaks and a period of 11.25 base pairs at $\phi \sim 290°$. When ϕ changes by a half-turn ($\phi \sim 110°$), the $\omega_{1,s}$ profile shifts by half of the period (compare Figure 4E and F), reflecting the half-period shift in the location of the minor groove facing the interior of the minicircle. Further analysis established $\omega_{1,s}$ to exhibit more pronounced modulations than $\omega_{2,s}$ and $\omega_{1,s}$ and $\omega_{2,s}$ to be highly correlated: $\omega_{2,s} \approx -0.3\omega_{1,s}$ (Supplementary Figure S11D, G). Regardless of the sequence, $\omega_{1,s}$ of a minicircle can be approximated as an analytic function that solely depends on ϕ, $\omega_{1,s}^{\text{circle}}(\phi)$ (Supplementary Figure



	A	G	ᵐG	T	C	ᵐC
T	12	6	–	68	60	120
C	6	12	–	35	34	47
ᵐC	–	–	30	–	–	–
A	68	35	–	48	44	39
G	60	34	–	44	58	50
ᵐG	120	47	–	39	50	38

Figure 4. Dinucleotide stiffness determines the free energy of a DNA minicircle. (**A**) Average torque required to maintain a prescribed value of the poloidal angle for NGO minicircle. Each torque value was obtained from a 400 ns MD simulation in which coordinates of the first eleven phosphorus atoms, P1 to P11, were harmonically restrained. Error bars indicate standard errors of 10-ns block averages. (**B**) Free energy landscapes of the NGO minicircle as a function of poloidal angle obtained by integration of the torque with respect to the angle (black) and through Boltzmann inversion of the poloidal angle distribution sampled by the free-equilibration simulation (blue). (**C**) Free energy landscapes of 26 DNA minicircles computed using the torque integration method. The curves are shifted vertically and horizontally to aid visual comparison. (**D**) Mean poloidal angle observed in unrestrained equilibration versus poloidal angle at the free energy minimum for 26 minicircles (see Supplementary Table S1 for sequence information). For the equilibrium simulations, error bars indicate the standard deviations. For the free energy landscapes, error bars indicate the Gaussian standard deviation, σ; the energy minimum was fitted to a harmonic function, $E = 0.5k(\phi - \phi_0)^2 + c$, then, $\sigma = \sqrt{k_B T/k}$. (**E**, **F**) Local bending of a DNA minicircle at two values of the poloidal angle: 288° (E) and 128° (F). The local bending profiles described by the 90 values of the torsion parameter $\omega_{1,s}$ are shown in the bar graphs. Blue solid curves indicate the analytic fit $\omega_{1,s}^{\text{circle}} = (9°)\sin^2\{\frac{\pi}{P}s - \frac{(\phi-1.5)}{2}\}$, where periodicity $P = 11.25$ bp. Insets illustrate local DNA bending at a CC step ($s = 7$). (**G**) Sequence-dependent dinucleotide stiffness matrix, K. The 16 matrix values were obtained by fitting the 26 free energy landscapes. In the matrix table, ᵐC indicates a methylated cytosine, and ᵐG indicates guanine of a methylated CpG site. (**H**) Free energy landscapes of the NGO and NGOme minicircles computed using the discrete WLC model (dashed lines) and the torque integration method (solid lines). (**I**) Poloidal angles at the free energy minimum of the discrete WLC model versus mean poloidal angles in equilibrium simulations of 41 different sequence minicircles (Supplementary Table S1). Sixteen points shown in blue belong to the test sequence set that were not included in the training data set used to derive the K matrix values.

S11B, E). Although crystallographic studies suggest that the conformations of DNA in nucleosomes substantially depend on DNA sequence (75), we found the local bending profile to be mostly sequence-independent within the resolution of our discrete WLC model.

The discrete WLC model (29,65) describes the stepwise bending energy as a quadratic function, $\bar{E}_s(\phi) = K(s,s+1)\{\omega_{1,s}^{\text{circle}}(\phi)\}^2$, where $K(s,s+1)$ is the stiffness constant for the step s, which was assumed to depend solely on the identity of nucleotide s and $s+1$ (65). The total energy of a minicircle is then $E^{\text{D-WLC}}(\phi) = \sum_{s=1}^{90} \bar{E}_s(\phi)$. For a DNA molecule containing four canonical nucleotides and a methylated cytosine at CpG sites, $K(s,s+1)$ has only sixteen unique values, corresponding to the sixteen different dinucleotide steps. We determined the numerical values of K, Figure 4G, by equating the total bending energy $E^{\text{D-WLC}}(\phi)$ to the free energy profiles computed using the torque-integration method (Figure 4C and Supplementary Figures S3–S9). As each type of a dinucleotide step ap-

pears approximately 150 times among 26 sequences of the minicircles, finding the sixteen values of K reduced to solving an overdetermined linear algebra problem, see Methods for the detailed description of the procedure. Validating our approach, we found the discrete WLC energy profiles, $E^{\text{D-WLC}}(\phi)$ to be in excellent agreement with the free energy landscapes computed using the torque-integration method (Figure 4H and Supplementary Figures S3-S9). Furthermore, the predicted location of the $E^{\text{D-WLC}}(\phi)$ minima showed a quantitative agreement (Pearson's correlation > 0.95) with the equilibrium poloidal angle observed in our equilibrium simulations of the 26 minicircles whose free-energy profiles were used for the determination of the K matrix values (red symbols in Figure 4I) and also for the other sixteen minicircles that were not used to determine the K values (blue symbols in Figure 4I). To further test the transferability of our discrete WLC model, we used the model to compute the free energy landscape for 85-bp 601-derivative minicircles, finding the results in good agreement with our all-atom MD torque-integration data, Supplementary Figure S5A,B.

The K matrix values, Figure 4G, provide quantitative information about the energetics of highly bent DNA. The values of the matrix indicate that pyrimidine (Y)–purine (R) steps such as TpA or CpG are more flexible than RY, YY, and RR steps, consistent with the previous analysis of crystal structures that described YR steps to act as 'hinges' (30). Consistent with the previous report showing that poly(dA:dT) tracts induce the depletion of nucleosome (13) and inhibit DNA cyclization (20,72), we also found TpT/ApA steps to be among the stiffest ones.

Given the periodic oscillatory profile of $\omega_{1,s}$ and the varying stiffness of the dinucleotide steps, the minicircle appears to minimize its energy by avoiding placement of stiff dinucleotide steps at the highly bent regions of the minicircle, which are the compressed major grooves at the inner surface. For example, the accumulated penalty of $\bar{E}_s$ over the 90 dinucleotide steps of the NGO minicircle (Supplementary Figure S11C,F,H) are minimal at $\phi \sim 290^\circ$, resulting in $E^{\text{D-WLC}}(\phi)$ at $\phi \sim 290^\circ$ about 3 kcal/mol lower than that at $\phi \sim 110^\circ$ (Figure 4H). Another good example is the (cgc**CCGG**cgcg)$_8$cg minicircle with its minimum energy conformation at $\phi \sim 0^\circ$ (Figure 5B). At this conformation, the CCGG motifs are highly bent (Figure 5C and Supplementary Movie S8), suggesting that they are more flexible than the other parts of the minicircle. When we methylated the CpG sites of the CCGG motifs, the minicircle underwent a complete inside-out transition, a 180° change of its poloidal angle (Figure 5A), placing the C$^{\text{m}}$CGG motifs at the straight parts of the octagon (Figure 5D and Supplementary Movie S9), a consequence of the C$^{\text{m}}$CGG motifs being stiffer than the CCGG ones. It is noteworthy that a similar effect of methylation on the DNA conformation was seen in the computational analysis of nucleosomal DNA (17). Stepwise distributions of bending energy at $\phi \sim 0^\circ$ (Figure 5E,F,G) quantitatively show that the increase in the total minicircle energy by 8 kcal/mol upon methylation (Figure 5B) is due to the step energy penalties localized near CpG sites. On the effects of cytosine methylation on DNA flexibility, two independent single-molecule experiments drew contradictory conclusions (15,76). The K matrix shows that the stiffness of $^{\text{m}}$CpG is about 2.5 times larger than that of CpG, i.e. that methylation makes DNA stiffer. Methylation was also found to increase the energy of all tested DNA minicircles with CpG sites at equilibrium poloidal angle (Figure 5H).

Discrete WLC model accounts for the sequence-dependence of DNA looping time

Encouraged by the ability of our discrete WLC model to quantitatively describe sequence-dependent energetics of a DNA minicircle, we investigated if our model can predict experimental DNA looping times as determined by the sm-FRET cyclization assay (20,21). Presently, the path that a DNA molecule takes when moving from a fully extended state to a closed minicircle is not known. In our attempt to rationalize DNA cyclization experiments, we assumed DNA cyclization to occur via a pathway where a single intermediate state of a low end-to-end distance is the rate-limiting step of the DNA cyclization reaction (Figure 6A). Specifically, we considered two possible intermediate states: a hairpin and an open circle (Figure 6A). Under this assumption, the Arrhenius equation predicts the looping time as $Ae^{\Delta G^{\ddagger}/k_B T}$, where $\Delta G^{\ddagger} = \sum_s \bar{E}_s$ is the energy of an intermediate state and A is a pre-exponential factor. For the local bending profile $\omega_{1,s}$ of the open circle conformation, we adopted the analytic profile of the minicircle, $\omega_{1,s}^{\text{circle}}$, and computed the open circle bending energy using the discrete WLC model, $\Delta G^{\ddagger,\text{circle}} = \sum_s K(s, s+1)\{\omega_{1,s}^{\text{circle}}\}^2$. To obtain the local bending profile $\omega_{1,s}$ of the hairpin-like conformation, we arranged a 90-bp dsDNA fragment into a hairpin structure by breaking the covalent bonds between the first and last base pairs of the equilibrated NGO minicircle and re-equilibrating the resulting structure for 500 ns enforcing a 3 nm distance between the centers of mass of the two terminal 10 bp segments. Figure 6B shows a typical configuration of the equilibrated hairpin. The resulting $\omega_{1,s}$ profile features a decaying oscillatory behavior with the maximum amplitude located near the center of the hairpin, Figure 6C, suggesting accumulation of the bending energy at the hairpin center. Assuming that the analytic function $\omega_{1,s}^{\text{hairpin}}$ fitted to the $\omega_{1,s}$ profile (Figure 6C) represents the hairpin bending profile regardless of the DNA nucleotide sequence, we computed the hairpin bending energy using the discrete WLC model, $\Delta G^{\ddagger,\text{hairpin}} = \sum_s K(s, s+1)\{\omega_{1,s}^{\text{hairpin}}\}^2$. Despite the assumptions we made about the pathway of the DNA cyclization reaction, the discrete WLC energy values for both $\Delta G^{\ddagger,\text{hairpin}}$ and $\Delta G^{\ddagger,\text{circle}}$ values show reasonable correlations with the logarithm of the looping times experimentally measured for different sequences (Figure 6D,E). It is noteworthy that $\Delta G^{\ddagger,\text{hairpin}}$ values are significantly larger than $\Delta G^{\ddagger,\text{circle}}$ values, suggesting that the quantitative estimation of the looping time will require more detailed information about the reaction pathway.

To test if our model of DNA looping has predictive power, we turned the NGO sequence into an extremely rigid one by introducing nucleotide substitutions at the four CpG sites of the NGO sequence (red boxes in Figure 2A).

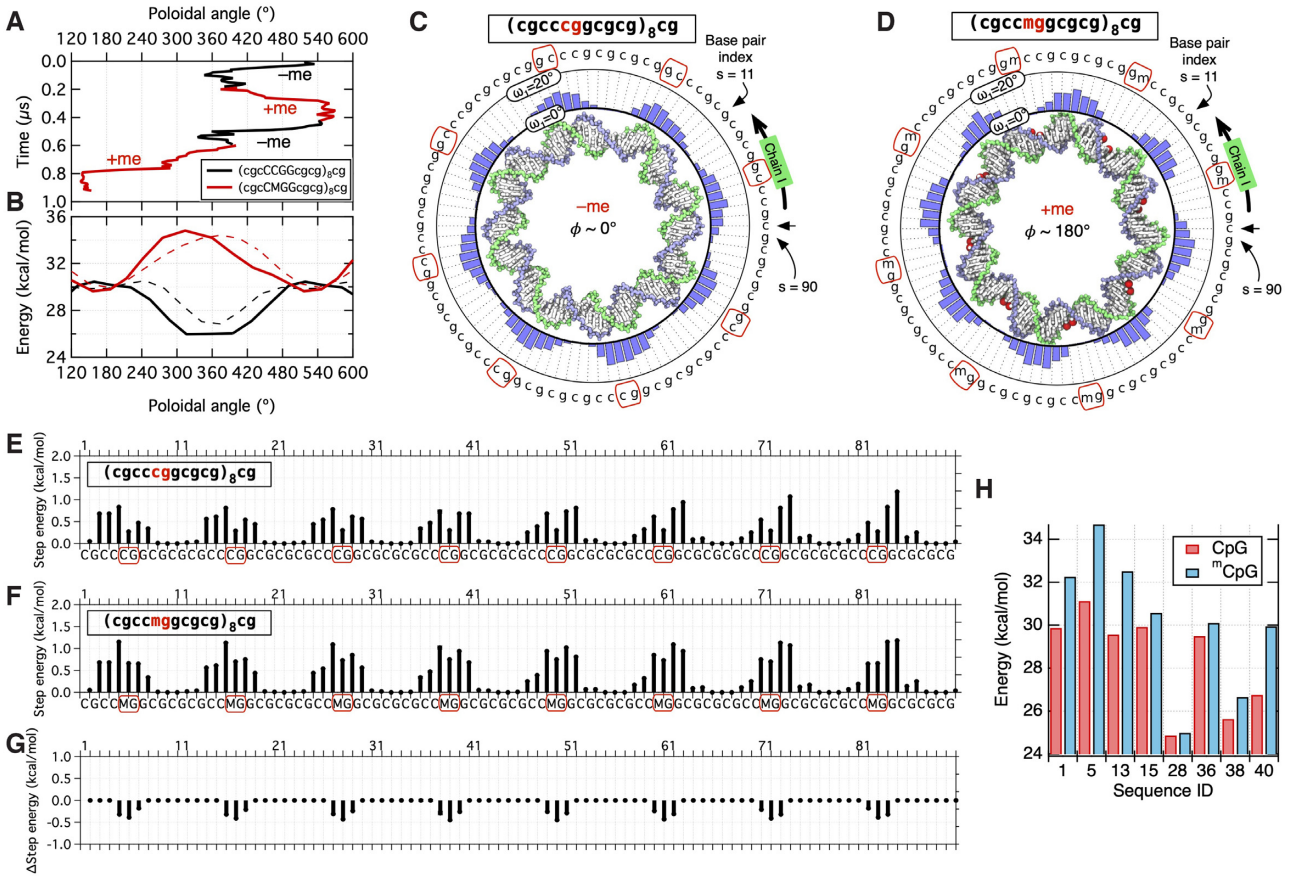


Figure 5. Methylation can cause both conformational and energetic perturbations. (**A**) Free-equilibration trajectories illustrating the change of the poloidal angle upon methylation of the (cgcccggcgcg)$_8$cg minicircle, sequence ID 41 and 42. (**B**) Methylation-dependent free energy landscapes of the (cgcccggcgcg)$_8$cg minicircle computed by the torque-integration method (solid) and using the discrete WLC model (dashed line). (**C, D**) Energy minimum conformations of the minicircle without (panel C) and with (panel D) CpG methylation. Red spheres in panel D indicate the methyl groups. Bar graphs in the outside show the distributions of ω_1. (**E, F**) Distribution of step energy values without (panel E) and with (panel F) CpG methylation at $\phi = 0°$. (**G**) The difference in step energy values for the data shown in panels E and F. (**H**) Changes in the total energy of seven DNA minicircles upon methylation.

Because we found the TpC/GpA and Tp^{m}C/^{m}GpA steps to be among the most rigid ones (Figure 4H), we designed the new NGOtcga sequence to have four TpCpGpA motifs by mutating only six nucleotides (Figure 6F). We then designed the NGOtmga sequence to have the four CpG sites of the NGOtcga sequence methylated (Figure 6F). In accordance with our predictions, the hairpin energy, $\Delta G^{\ddagger,\mathrm{hairpin}}$, increased upon the first mutation and even more after the methylation (Figure 6F). Furthermore, the cyclization assay confirmed that the fraction of looped molecules dramatically decreased following the prediction (Figure 6G). Note that we did not estimate the looping time of NGOtcga and NGOtmga because the fractions of looped molecules in the experiment were too low to be fitted by an exponential function (Figure 6G).

Sequence-dependent elasticity of extremely bent DNA describes genome-wide nucleosome occupancy

Using our simple yet accurate model of sequence-dependent energetics of highly-bent DNA, we were able to reproduce the experimental nucleosome occupancy data (13). To compute the nucleosomal DNA bending energy at a genomic locus i, we assumed that the locus i is at the dyad axis of a nucleosome and summed the step-wise bending energy $\bar{E}_s$ from −73 to +73 positions from the locus, $\sum_{s=-73}^{73} \bar{E}_{i+s}$. Considering that the minicircle represents well the bending of DNA within a nucleosome (Figure 1), we computed $\bar{E}_s$ using the analytical fit to the bending profile of a minicircle, $\omega_{1,s}^{\mathrm{circle}}$, with the periodicity of 10 bp. The specific poloidal angle of the $\omega_{1,s}^{\mathrm{circle}}$ function was set by requiring the major groove of the nucleosomal DNA to be at the dyad axis and to face inward (Figure 1B), which is a conserved feature of all nucleosome structures (77). The computed bending energy showed a significant variation depending on the locus, Figure 7A. Despite the simplicity of the model and the absence of the histone–DNA interactions, we found the bending energy to be strongly anticorrelated with the experimental nucleosome occupancy, Figure 7A. The anticorrelation between the computed bending energy and the experimental occupancy data (13) persisted when the data were averaged over the entire *Saccharomyces cerevisiae* and human genomes (Figure 7B). This analysis reaffirms the transferability of our discrete WLC model of highly bent DNA to biological systems, such as nucleosomes, across species. It is noteworthy that the GC content is also highly

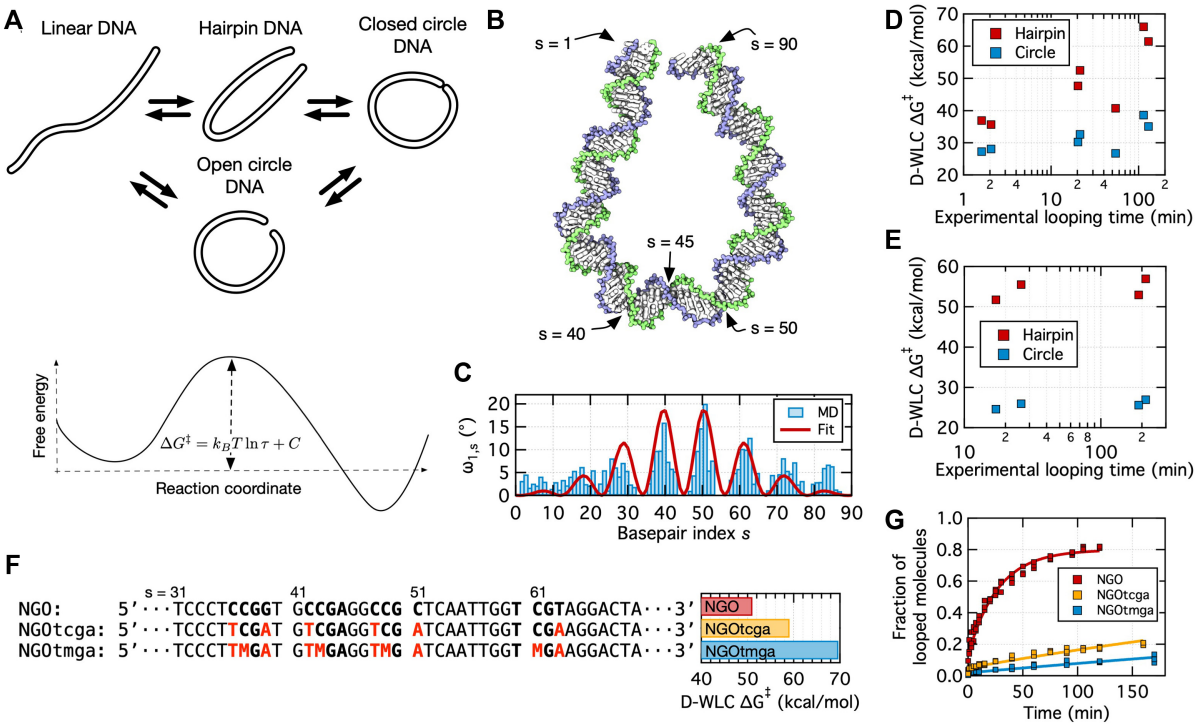


Figure 6. The dinucleotide stiffness model reproduces and predicts results of single-molecule DNA looping assay. (**A**) Schematics of the assumed mechanism of minicircle formation, where formation of an intermediate conformation (e.g. a hairpin or an open circle) is the rate-limiting step. (**B**, **C**) A hairpin structure of the NGO construct (panel B) and the local bending profile $\omega_{1,s}$ of the hairpin (bar graphs in panel C) at the end of a 1 μs simulation. In panel C, red line indicates a fitting function, $\omega_{1,s}^{\text{hairpin}} = \{20^{\circ}\}\, e^{-\left(\frac{s-L/2}{L/4}\right)^2} \sin^2\{\frac{\pi}{11}(s - L/2)\}$, where $L = 90$ is the length of dsDNA. The local bending profile of an open circle structure can be approximated using the $\omega_{1,s}^{\text{circle}}$ function derived from the closed circle simulations (e.g. Figure 4E, F). (**D**, **E**) Discrete WLC energy of the intermediate conformations (red and blue symbols for hairpin and open circle, respectively) versus experimentally measured looping times. Experimental data in panels D and E were taken from Ref. (20) (73-bp TA, E8, R73, E8A10, E8A17, E8A26 and E8A38 sequences) and (21) (83-bp 601LH, 601RH, 601MFLH, 601MFRH sequences), respectively. (**F**) Designs of new NGO variants, NGOtcga and NGOtmga. In the NGO sequence, bold font highlights the nucleotide substitution targets. Red font in the other two sequences indicates the mutated nucleotides. Discrete WLC energy of each hairpin variant is shown in the plot on the right. (**G**) Experimental single-molecule cyclization assay performed using NGO, NGOtcga, and NGOtmga constructs.

correlated with the nucleosome occupancy (Figure 7B) (78). However, considering that the GC content cannot explain the significant changes in looping time due to minor mutations and methylations (e.g. NGO, NGOtcga, and NGOtmga in Figure 6) and the presence of the free energy landscape as a function of poloidal angle, it seems that the discrete WLC model is a more general physical model than the GC content-based model (11,78).

To investigate the effects of local DNA stiffness on gene activities, we averaged the bending energy and the nucleosome occupancy over transcription start sites (TSS) (13,79), in a similar way to the previous MD-based approach (80). In *Saccharomyces cerevisiae*, the bending energy near TSS is higher than the genome-wide average energy level by 3 kcal/mol. This elevated energy level indicates that TSSs of *S. cerevisiae* evolved to be depleted of nucleosomes, with the nucleosome occupancy as low as 0.2 in the upstream region of TSS (Figure 7C) (13). Pearson's r value between the bending energy and the occupancy near TSS is –0.93 (Figure 7C).

Unlike TSSs of *S. cerevisiae*, human TSSs inside unmethylated CpG islands appear to be highly enriched with nucleosomes, with the occupancy as high as 0.8 and the bending energy lower than that away from TSSs by 4 kcal/mol (Figure 7D). Interestingly, we found that DNA methylations of CpG sites can significantly increase the bending energy at human TSSs inside the CpG islands (Figure 7D). Among the four cell types (B, Hela, H1, and HepG2), B cells showed the most substantial increase in the energy level, close to that of TSSs of *S. cerevisiae*. This finding suggests that the human genome evolved to have high intrinsic nucleosome occupancy that can be conditionally reduced by DNA methylations. The methylation-induced nucleosome depletion adds mechanics-based evidence to the genetics-based evidence (81) that challenge the conventional view that DNA methylation generally silences transcription. On the other hand, human TSSs residing outside the CpG islands have relatively a flat occupancy profile and, thus, a weaker susceptibility to DNA methylation (Figure 7E).

At a chromosome scale, the two high-throughput sequencing-based techniques, DamID (82) and Hi-C (83), dramatically expanded our understanding of genomic architecture. Thus, DamID determined lamina-associated domains of the human genome (mostly heterochromatin), whereas Hi-C revealed that the human genome is physically separated into A (mostly euchromatin) and B (mostly heterochromatin) compartments. Interestingly, we found the bending energy profile of nucleosomal DNA to be highly

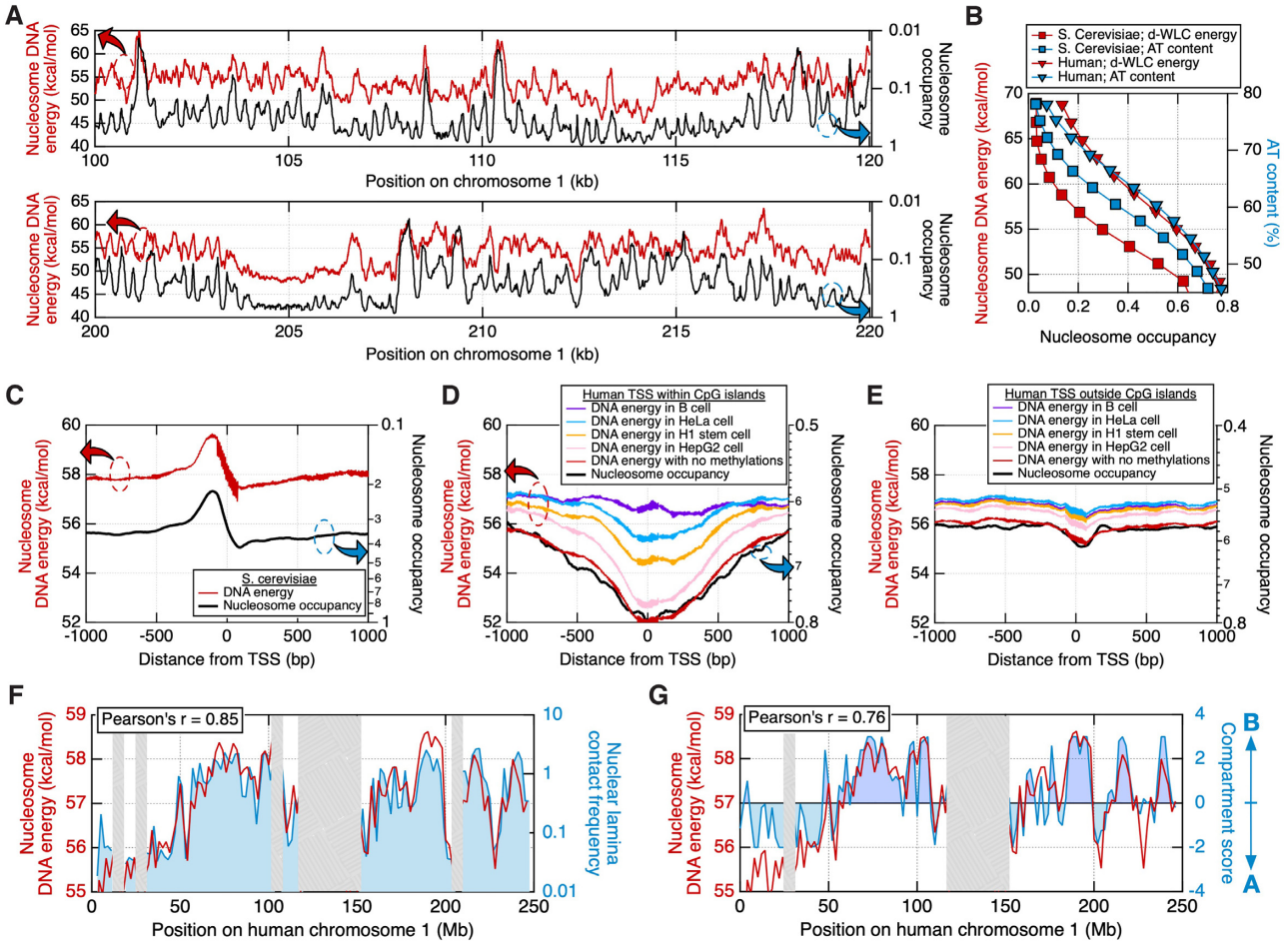


Figure 7. Deformation energy of DNA determines genome-wide nucleosome occupancy. (**A**) Correlation between the bending energy of nucleosomal DNA (red lines) and the experimental nucleosome occupancy data (13) (black lines) for two 20-kb segments of chromosome 1 of *Saccharomyces cerevisiae*. The nucleosomal bending energy at a given position was computed by summing dinucleotide step deformation energy over 147 base pairs. (**B**) Genome-wide correlation of the experimental nucleosome occupancy of *Saccharomyces cerevisiae* and human with the bending energy and the AT (or equivalently GC) content of nucleosomal DNA. (**C**) Nucleosome occupancy and bending energy averaged over all transcription start sites (TSS) in the genome of *Saccharomyces cerevisiae*. (**D, E**) Nucleosome occupancy and bending energy averaged over all transcription start sites (TSS) inside (panel D) and outside (panel E) CpG islands in human genomes. (**F**, **G**) Comparisons of the nucleosome bending energy (red lines) with a nuclear lamina contact frequency profile (82) (blue-filled line in panel F) and Hi-C compartment profile (83) (blue filled line in panel G) in the human chromosome 1; mean values in 2-Mb bin are shown. Gray boxes indicate domains with no experimental reads.

correlated with the nuclear lamina contact frequency (Figure 7F) and with the A/B compartment score (Figure 7G). These results suggest a yet undetermined role for DNA mechanics in the large-scale genome organization. Compared to euchromatin, heterochromatin domains that have high nuclear lamina contact frequency and a tendency to be localized within the B compartment (84) also have higher bending energy and, consequently, a lower nucleosome occupancy. Overall, our analysis raises the possibility that the DNA bending energetics plays a major role in the architecture of the chromosomes.

DISCUSSION

We have shown that the nucleotide sequence of highly bent DNA encodes a rugged free energy landscape along which the DNA searches for the minimum free energy orientation through spontaneous inside-out transition. A genomic locus will have a high likelihood of nucleosome formation when the well-defined configuration seen in the crystal structures (7,8) can minimize the DNA bending energy. Perturbing the DNA orientation out of this equilibrium conformation, for example, by the twist generated by a DNA polymerase or a chromatin remodeling factor, is expected to destabilize the nucleosome by increasing the bending energy of the nucleosomal DNA. The inside-out transitions of DNA might also be important for the diffusion of nucleosome along the genome (85). For sequences with relatively flat or rugged energy landscapes, one can expect diffusion to occur predominantly via sliding or hopping, respectively.

Compared to linear fragment systems previously used to parametrize coarse-grained models (22–28), our minicircle system provides a much better sampling of highly bent configurations and their correlated fluctuations. Although our discrete WLC model based on the minicircle simulations was successful in the reproduction as well as the prediction of the experimental measurements, its accuracy can be further improved by using a non-quadratic energy function

growing 'much more slowly than quadratic in the curvature', as suggested by Cloutier *et al.* (86) and by introducing non-zero equilibrium values for the dinucleotide step angle to account for the specific intrinsic curvature at the dinucleotide level (72–74).

In contrast to the previous MD studies that found DNA minicircles to form kinks within 100 ns (59–62), we found them to maintain their kink-less structures for tens of microseconds (Supplementary Figure S12). The key methodological distinction of our approach is the initial equilibration of a DNA minicircle with weak restraints applied to maintain Watson–Crick base pairing. This equilibration allows the DNA minicircle to find a relaxed, octagon-like orientation, which can persist for tens of microseconds in the absence of any restraints. To directly demonstrate the critical nature of this initial equilibration step, we applied our equilibration protocol to the two DNA minicircles (106 and 102 bp) that were previously reported to develop kinks (61), finding their structure to remain kink-free within 500 ns under the simulation conditions (0.1 M NaCl, 300 K, and the parmbsc0 force field) identical to those in Ref. (61) (Supplementary Figure S13). Note that we did not observe any structural deterioration in our MD simulations of 90-bp minicircles; such structural perturbations have been observed experimentally (87) and have been associated with the negative supercoiling of the 90-bp minicircles.

Presently, several force field choices are available for modeling DNA mechanics, each coming with its own pros and cons. Our study used an older AMBER force field variant, parmbsc0 (43,44), supplemented with our CUFIX corrections to non-bonded interactions, a combination that has been extensively validated and improved through our previous work (47,88,89). The newest version of the AMBER DNA force field, parmbsc1 (62), provides superior accuracy for MD simulation of diverse DNA systems. To test how the force field choice affects the behavior of minicircles, we performed two 800-ns simulations of the 85-bp minicircle (601LH) using the parmbsc1 model, see Supplementary Movies S10 and S11. In one of the two simulations (Supplementary Movie S11), we observed distortion of the double-helical structure, similar to the behavior reported in the study that introduced the parmbsc1 model (62), where one broken configuration was observed out of three simulations of a 106-bp minicircle. We believe, however, that the reduced minicircle stability in parmbsc1 simulations derives from the differences in the overall DNA stiffness rather than from sequence-specific effects. Thus, according to the previous study (90), the average persistence length of a DNA duplex modeled using parmbc1 and parmbc0 is 59 and 52 nm, respectively, with the experimental value being close to 50 nm (6). Given that the diameter of a typical minicircle considered in our work is only 10 nm, stiffer DNA is more likely to lose its double-helical structure.

To place the parameters of our discrete WLC model in the context of other work, we compare the rank-order of our dinucleotide stiffness values to those independently extracted from the analysis of experimental genome-wide loopability data (91,92). Overall, the relative stiffness data extracted from the results of loop-seq experiments are in reasonable qualitative agreement with our computationally derived values (Supplementary Figure S14A). For example, the loop-seq data find the TCGA motifs to be the stiffest among all tetranucleotides and methylation of CpG sites to generally increase the DNA stiffness. Comparison to the loop-seq data also shows some deviation from our results. Most significantly, the loop-seq analysis categorized TG/CA and CG steps as relatively stiff ones (92) (Supplementary Figure S14A), which goes against the generally accepted idea that a pyrimidine–purine step acts as a hinge (30). Another comparison of our *K*-matrix values is shown in Supplementary Figure S14B, where our dinucleotide stiffness values, K, are plotted against the effective persistence length, α, extracted from DNA looping experiments by Geggier and Vologodskii (29). Although Geggier and Vologodskii derived the stiffness parameters using a WLC model, their stiffness parameters differ from ours (Supplementary Figure S14B). It is highly likely that the discord derives from our use of the non-uniform oscillatory pattern to describe dinucleotide bending (e.g. Figure 4E, F)—the key new feature of our discrete WLC model introduced to account for the MD results. Compared to the stiffness parameters derived in the absence of the oscillatory bending pattern, our stiffness parameters show a broader distribution because the highly bent and almost straight regions can afford lower and higher stiffness values, respectively. As a result, the ratio between the maximum and minimum stiffness values in our stiffness parameter set is higher, $K_{max}/K_{min} \approx 6.4$, than that in Ref. (29), $\alpha_{max}/\alpha_{min} \approx 1.3$ (Supplementary Figure S14B).

Finally, we point out that our results raise a question about the fundamental relationship between the chromatin states (euchromatin or heterochromatin) and the nucleosome occupancy. Contrary to the conventional wisdom that the gene activity in heterochromatin is low due to high nucleosome occupancies, our analysis and the comparison with the DamID and Hi-C results suggest that heterochromatin may have a lower nucleosome occupancy than euchromatin, which is consistent with the predictions of the model proposed by Widom and Segal (13). This conclusion may be consistent with the recent counterintuitive finding that the phase separation of heterochromatin is accompanied by the destabilization of nucleosomes (93). If it is true that heterochromatin is more nucleosome-depleted than euchromatin, there should be an alternative mechanism that can account for the compaction of heterochromatin. Possibly, histone-free DNA can promote the heterochromatin formation by providing binding sites for proteins that mediate DNA–nuclear lamina interactions (84) or by polyamine-mediated DNA condensation (94).

DATA AVAILABILITY

Raw simulation and experimental data that support the findings of this study are available from the corresponding authors, JY, TH, and AA, upon request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

J.Y., C.M. and A.A. gladly acknowledge supercomputer time provided through the Blue Waters Sustained Petascale Computer System (UIUC), the Illinois Campus Cluster, XSEDE Allocation Grant MCA05S028, and the National Supercomputing Center (Grant KSC-2020-CRE-0080). We thank Dr. Noam Kaplan for kindly sharing the nucleosome occupancy data.

FUNDING

National Science Foundation (USA) through the Center for the Physics of Living Cells [PHY-1430124] and [EFMA-1933303 to T.H.]; National Institutes of Health [R01-GM137015 to A.A.; R35-GM122569 to T.H.]; National Research Foundation of Korea grant funded by the Korea government [2020R1A2C1101424 to J.Y.]. Funding for open access charge: National Research Foundation of Korea [2020R1A2C1101424].

Conflict of interest statement. None declared.

REFERENCES

1. Blackwood,E.M. and Kadonaga,J.T. (1998) Going the distance: a current view of enhancer action. *Science*, **281**, 60–63.
2. Johnson,S., Linden,M. and Phillips,R. (2012) Sequence dependence of transcription factor-mediated DNA looping. *Nucleic Acids Res.*, **40**, 7728–7738.
3. Kornberg,R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**, 868–871.
4. Olins,A.L. and Olins,D.E. (1974) Spheroid chromatin units (v bodies). *Science*, **183**, 330–332.
5. Crick,F.H. and Klug,A. (1975) Kinky helix. *Nature*, **255**, 530–533.
6. Baumann,C.G., Smith,S.B., Bloomfield,V.A. and Bustamante,C. (1997) Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 6185–6190.
7. Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251–260.
8. Vasudevan,D., Chua,E.Y.D. and Davey,C.A. (2010) Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol.*, **403**, 1–10.
9. Satchwell,S.C., Drew,H.R. and Travers,A.A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.*, **191**, 659–675.
10. Thastrom,A., Lowary,P.T., Widlund,H.R., Cao,H., Kubista,M. and Widom,J. (1999) Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.*, **288**, 213–229.
11. Tillo,D., Kaplan,N., Moore,I.K., Fondufe-Mittendorf,Y., Gossett,A.J., Field,Y., Lieb,J.D., Widom,J., Segal,E. and Hughes,T.R. (2010) High nucleosome occupancy is encoded at human regulatory sequences. *PLoS One*, **5**, e9129.
12. Segal,E., Fondufe-Mittendorf,Y., Chen,L., Thastrom,A., Field,Y., Moore,I.K., Wang,J.P. and Widom,J. (2006) A genomic code for nucleosome positioning. *Nature*, **442**, 772–778.
13. Kaplan,N., Moore,I.K., Fondufe-Mittendorf,Y., Gossett,A.J., Tillo,D., Field,Y., LeProust,E.M., Hughes,T.R., Lieb,J.D., Widom,J. *et al.* (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, **458**, 362–366.
14. Brogaard,K., Xi,L., Wang,J.P. and Widom,J. (2012) A map of nucleosome positions in yeast at base-pair resolution. *Nature*, **486**, 496–501.
15. 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.
16. Luo,G.Z., Hao,Z., Luo,L., Shen,M., Sparvoli,D., Zheng,Y., Zhang,Z., Weng,X., Chen,K., Cui,Q. *et al.* (2018) N(6)-methyldeoxyadenosine directs nucleosome positioning in Tetrahymena DNA. *Genome Biol.*, **19**, 200.
17. Portella,G., Battistini,F. and Orozco,M. (2013) Understanding the connection between epigenetic DNA methylation and nucleosome positioning from computer simulations. *PLoS Comput. Biol.*, **9**, e1003354.
18. Shore,D., Langowski,J. and Baldwin,R.L. (1981) DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 4833–4837.
19. Cloutier,T.E. and Widom,J. (2004) Spontaneous sharp bending of double-stranded DNA. *Mol. Cell*, **14**, 355–362.
20. 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.
21. Ngo,T.T., Zhang,Q., Zhou,R., Yodh,J.G. and Ha,T. (2015) Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. *Cell*, **160**, 1135–1144.
22. Mitchell,J.S., Glowacki,J., Grandchamp,A.E., Manning,R.S. and Maddocks,J.H. (2017) Sequence-dependent persistence lengths of DNA. *J. Chem. Theory Comput.*, **13**, 1539–1555.
23. Marin-Gonzalez,A., Vilhena,J.G., Moreno-Herrero,F. and Perez,R. (2019) DNA crookedness regulates DNA mechanical properties at short length scales. *Phys. Rev. Lett.*, **122**, 048102.
24. Eslami-Mossallam,B., Schram,R.D., Tompitak,M., van Noort,J. and Schiessel,H. (2016) Multiplexing genetic and nucleosome positioning codes: a computational approach. *PLoS One*, **11**, e0156905.
25. Walther,J., Dans,P.D., Balaceanu,A., Hospital,A., Bayarri,G. and Orozco,M. (2020) A multi-modal coarse grained model of DNA flexibility mappable to the atomistic level. *Nucleic Acids Res.*, **48**, e29.
26. Drsata,T., Spackova,N., Jurecka,P., Zgarbova,M., Sponer,J. and Lankas,F. (2014) Mechanical properties of symmetric and asymmetric DNA A-tracts: implications for looping and nucleosome positioning. *Nucleic Acids Res.*, **42**, 7383–7394.
27. Pasi,M., Maddocks,J.H., Beveridge,D., Bishop,T.C., Case,D.A., Cheatham,T. 3rd, Dans,P.D., Jayaram,B., Lankas,F., Laughton,C. *et al.* (2014) muABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.*, **42**, 12272–12283.
28. Ma,N. and van der Vaart,A. (2016) Anisotropy of B-DNA groove bending. *J. Am. Chem. Soc.*, **138**, 9951–9958.
29. Geggier,S. and Vologodskii,A. (2010) Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 15421–15426.
30. Olson,W.K., Gorin,A.A., Lu,X.J., Hock,L.M. and Zhurkin,V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 11163–11168.
31. Morozov,A.V., Fortney,K., Gaykalova,D.A., Studitsky,V.M., Widom,J. and Siggia,E.D. (2009) Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Res.*, **37**, 4707–4722.
32. Freeman,G.S., Hinckley,D.M., Lequieu,J.P., Whitmer,J.K. and de Pablo,J.J. (2014) Coarse-grained modeling of DNA curvature. *J. Chem. Phys.*, **141**, 165103.
33. Du,Q., Smith,C., Shiffeldrim,N., Vologodskaia,M. and Vologodskii,A. (2005) Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 5397–5402.
34. Le,T.T. and Kim,H.D. (2014) Probing the elastic limit of DNA bending. *Nucleic Acids Res.*, **42**, 10786–10794.
35. Harrison,R.M., Romano,F., Ouldridge,T.E., Louis,A.A. and Doye,J.P.K. (2019) Identifying physical causes of apparent enhanced cyclization of short DNA molecules with a coarse-grained model. *J. Chem. Theory Comput.*, **15**, 4660–4672.
36. Abraham,M.J., Murtola,T., Schulz,R., Páll,S., Smith,J.C., Hess,B. and Lindahl,E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1-2**, 19–25.
37. Nosé,S. and Klein,M.L. (1983) Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, **50**, 1055–1076.
38. Hoover,W.G. (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys Rev A Gen Phys*, **31**, 1695–1697.
39. Parrinello,M. and Rahman,A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182–7190.

40. 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.
41. Hess,B., Bekker,H., Berendsen,H.J.C. and Fraaije,J.G.E.M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
42. 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.
43. 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.
44. Perez,A., Marchan,I., Svozil,D., Sponer,J., Cheatham,T.E. 3rd, Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.
45. 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.
46. 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.
47. Yoo,J. and Aksimentiev,A. (2018) New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *PCCP*, **20**, 8432–8449.
48. 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.
49. Case,D.A., Cheatham,T.E. III, Darden,T., Gohlke,H., Luo,R., Merz,K.M. Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
50. Yoo,J. and Aksimentiev,A. (2012) Improved parametrization of Li(+), Na(+), K(+), and Mg(2+) ions for all-atom molecular dynamics simulations of nucleic acid systems. *J. Phys. Chem. Lett.*, **3**, 45–50.
51. Panyukov,S. and Rabin,Y. (2000) Fluctuating filaments: statistical mechanics of helices. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics*, **62**, 7135–7146.
52. Yoo,J. and Aksimentiev,A. (2013) In situ structure and dynamics of DNA origami determined through molecular dynamics simulations. *PNAS*, **110**, 20099–20104.
53. Choe,S. and Sun,S.X. (2005) The elasticity of alpha-helices. *J. Chem. Phys.*, **122**, 244912.
54. 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.
55. Maffeo,C. and Aksimentiev,A. (2020) MrDNA: a multi-resolution model for predicting the structure and dynamics of DNA systems. *Nucleic Acids Res.*, **48**, 5135–5146.
56. Hinckley,D.M., Freeman,G.S., Whitmer,J.K. and de Pablo,J.J. (2013) An experimentally-informed coarse-grained 3-Site-Per-Nucleotide model of DNA: structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.*, **139**, 144903.
57. Snodin,B.E.K., Randisi,F., Mosayebi,M., Šulc,P., Schreck,J.S., Romano,F., Ouldridge,T.E., Tsukanov,R., Nir,E., Louis,A.A. *et al.* (2015) Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *J. Chem. Phys.*, **142**, 234901.
58. Lowary,P.T. and Widom,J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.*, **276**, 19–42.
59. Lankas,F., Lavery,R. and Maddocks,J.H. (2006) Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure*, **14**, 1527–1534.
60. Mitchell,J.S., Laughton,C.A. and Harris,S.A. (2011) Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA. *Nucleic Acids Res.*, **39**, 3928–3938.
61. Sutthibutpong,T., Matek,C., Benham,C., Slade,G.G., Noy,A., Laughton,C., JP,K.D., Louis,A.A. and Harris,S.A. (2016) Long-range correlations in the mechanics of small DNA circles under topological stress revealed by multi-scale simulation. *Nucleic Acids Res.*, **44**, 9121–9130.
62. Ivani,I., Dans,P.D., Noy,A., Perez,A., Faustino,I., Hospital,A., Walther,J., Andrio,P., Goni,R., Balaceanu,A. *et al.* (2016) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55–58.
63. Skoruppa,E., Nomidis,S.K., Marko,J.F. and Carlon,E. (2018) Bend-Induced twist waves and the structure of nucleosomal DNA. *Phys. Rev. Lett.*, **121**, 088101.
64. Caraglio,M., Skoruppa,E. and Carlon,E. (2019) Overtwisting induces polygonal shapes in bent DNA. *J. Chem. Phys.*, **150**, 135101.
65. Vologodskii,A. and Frank-Kamenetskii,M.D. (2013) Strong bending of the DNA double helix. *Nucleic Acids Res.*, **41**, 6785–6792.
66. Schellman,J.A. (1974) Flexibility of DNA. *Biopolymers*, **13**, 217–226.
67. Shimada,J. and Yamakawa,H. (1984) Ring-closure probabilities for twisted wormlike chains. application to DNA. *Macromolecules*, **17**, 689–698.
68. Bustamante,C., Marko,J.F., Siggia,E.D. and Smith,S. (1994) Entropic elasticity of lambda-phage DNA. *Science*, **265**, 1599–1600.
69. Wiggins,P.A., van der Heijden,T., Moreno-Herrero,F., Spakowitz,A., Phillips,R., Widom,J., Dekker,C. and Nelson,P.C. (2006) High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotechnol.*, **1**, 137–141.
70. Shin,J., Lee,O.-C. and Sung,W. (2015) How a short double-stranded DNA bends. *J. Chem. Phys.*, **142**, 155101.
71. Salari,H., Eslami-Mossallam,B., Naderi,S. and Ejtehadi,M.R. (2015) Extreme bendability of DNA double helix due to bending asymmetry. *J. Chem. Phys.*, **143**, 104904.
72. Vologodskaia,M. and Vologodskii,A. (2002) Contribution of the intrinsic curvature to measured DNA persistence length. *J. Mol. Biol.*, **317**, 205–213.
73. Crothers,D.M., Haran,T.E. and Nadeau,J.G. (1990) Intrinsically bent DNA. *J. Biol. Chem.*, **265**, 7093–7096.
74. Koo,H.S., Wu,H.M. and Crothers,D.M. (1986) DNA bending at adenine. thymine tracts. *Nature*, **320**, 501–506.
75. Olson,W.K. and Zhurkin,V.B. (2011) Working the kinks out of nucleosomal DNA. *Curr. Opin. Struct. Biol.*, **21**, 348–357.
76. Choy,J.S., Wei,S., Lee,J.Y., Tan,S., Chu,S. and Lee,T.H. (2010) DNA methylation increases nucleosome compaction and rigidity. *J. Am. Chem. Soc.*, **132**, 1782–1783.
77. Andrews,A.J. and Luger,K. (2011) Nucleosome structure(s) and stability: variations on a theme. *Annu. Rev. Biophys.*, **40**, 99–117.
78. Tillo,D. and Hughes,T.R. (2009) G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics*, **10**, 442.
79. Schones,D.E., Cui,K., Cuddapah,S., Roh,T.Y., Barski,A., Wang,Z., Wei,G. and Zhao,K. (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell*, **132**, 887–898.
80. Deniz,O., Flores,O., Battistini,F., Perez,A., Soler-Lopez,M. and Orozco,M. (2011) Physical properties of naked DNA influence nucleosome positioning and correlate with transcription start and termination sites in yeast. *BMC Genomics*, **12**, 489.
81. Suzuki,M.M. and Bird,A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
82. Kind,J., Pagie,L., de Vries,S.S., Nahidiazar,L., Dey,S.S., Bienko,M., Zhan,Y., Lajoie,B., de Graaf,C.A., Amendola,M. *et al.* (2015) Genome-wide maps of nuclear lamina interactions in single human cells. *Cell*, **163**, 134–147.
83. Rao,S.S., Huntley,M.H., Durand,N.C., Stamenova,E.K., Bochkov,I.D., Robinson,J.T., Sanborn,A.L., Machol,I., Omer,A.D., Lander,E.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.
84. van Steensel,B. and Belmont,A.S. (2017) Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell*, **169**, 780–791.
85. Rudnizky,S., Khamis,H., Malik,O., Melamed,P. and Kaplan,A. (2019) The base pair-scale diffusion of nucleosomes modulates binding of transcription factors. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 12161–12166.
86. Cloutier,T.E. and Widom,J. (2005) DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 3645–3650.
87. Du,Q., Kotlyar,A. and Vologodskii,A. (2008) Kinking the double helix by bending deformation. *Nucleic Acids Res.*, **36**, 1120–1128.

88. You,S., Lee,H.G., Kim,K. and Yoo,J. (2020) Improved parameterization of protein-DNA interactions for molecular dynamics simulations of PCNA diffusion on DNA. *J. Chem. Theory Comput.*, **16**, 4006–4013.
89. Yoo,J., Winogradoff,D. and Aksimentiev,A. (2020) Molecular dynamics simulations of DNA-DNA and DNA-protein interactions. *Curr. Opin. Struct. Biol.*, **64**, 88–96.
90. Dans,P.D., Balaceanu,A., Pasi,M., Patelli,A.S., Petkeviciute,D., Walther,J., Hospital,A., Bayarri,G., Lavery,R., Maddocks,J.H. *et al.* (2019) The static and dynamic structural heterogeneities of B-DNA: extending Calladine-Dickerson rules. *Nucleic Acids Res.*, **47**, 11090–11102.
91. Basu,A., Bobrovnikov,D.G., Qureshi,Z., Kayikcioglu,T., Ngo,T.T.M., Ranjan,A., Eustermann,S., Cieza,B., Morgan,M.T., Hejna,M. *et al.* (2021) Measuring DNA mechanics on the genome scale. *Nature*, **589**, 462–467.
92. Basu,A., Bobrovnikov,D.G., Cieza,B., Qureshi,Z. and Ha,T. (2020) Deciphering the mechanical code of genome and epigenome. bioRxiv doi: https://doi.org/10.1101/2020.08.22.262352, 24 August 2020, preprint: not peer reviewed.
93. Sanulli,S., Trnka,M.J., Dharmarajan,V., Tibble,R.W., Pascal,B.D., Burlingame,A.L., Griffin,P.R., Gross,J.D. and Narlikar,G.J. (2019) HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature*, **575**, 390–394.
94. 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.