SUPPLEMENTARY MATERIALS

Nucleosome-free DNA regions differentially affect distant communication in chromatin

Ekaterina V. Nizovtseva, Nicolas Clauvelin, Stefjord Todolli, Yury S. Polikanov, Olga I. Kulaeva, Scott Wengrzynek, Wilma K. Olson and Vasily M. Studitsky

Supplementary Methods – Computer Simulations

DNA model. Our mesoscale treatment of chromatin represents DNA in terms of a set of rigid-body parameters that specify the relative orientation and displacement of successive base pairs (1). These details, excluded in earlier models of fluctuating chromatin fibers (2,3), are needed to keep precise account of the spatial disposition of interacting proteins. The protein-free linker DNA and the enhancer and promoter sequences at the ends of the modeled constructs are governed by an ideal elastic potential that allows for deformations of double-helical structure consistent with the solution properties of DNA (4). The equilibrium rest state of the free DNA is that of an ideal, perfectly straight molecule with 10.5 base pairs per turn and 3.6 nm pitch, equivalent to ~34.3° helical twist and 0.34-nm vertical displacement at each base-pair step. The force constants impeding deformations of structure are chosen such that a bend of 4.84° or a change in twist of 4.09° raises the energy by 0.5 k_BT, the base-pair level equivalents of a bending persistence length of 47.7 nm and a twisting persistence length of 66.5 nm. The latter values are based on the measured persistence length of intrinsically straight, mixed-sequence DNA (5), the equilibrium topoisomer distributions of DNA minicircles (6), and the fluorescence depolarization anisotropy of ethidium bromide intercalated in DNA minicircles (7). The chain is treated as inextensible, with the spacing of successive base pairs held fixed by use of very large force constants.

Nucleosomes and regulatory proteins. The nucleosomes are modeled as rigid bodies with the associated DNA assigned the spatial pathway found in the currently best resolved nucleosome core-particle structure (8). Each nucleosome carries three reference frames, one on each H2A·H2B·H3·H4 tetramer and one on the system as a whole. The first two frames are used in the detection of excluded volume effects and the third, a frame midway between those of the tetramers, to keep track of charged residues for the computation of electrostatic energy. The enhancer-binding nitrogen regulatory protein C (NtrC) is represented by a pair of stacked cylinders (the DNA-binding domain 7 nm diameter $\times$ 3 nm high, the ATPase domain 15 nm diameter $\times$ 4.3 nm high) and the RNA polymerase-σ^{54} assembly, on the promoter, by a single cylinder (11 nm diameter $\times$ 13.5 nm high) consistent with the observed dimensions of the protein assemblies (9-11). The pathways imposed on the DNA by the regulatory proteins are based on the known high-resolution structures with 31 bp at the 5′-end of the modeled system bound to the polymerase and 27 bp at the 3′-end wrapped around the DNA binding domain of NtrC.

Excluded volume effects. The simulations include a collision-detection step to reject configurations in which proteins and/or DNA overlap. This step, implemented with the help of the ODE software library (12), accounts for excluded-volume effects within the chromatin construct. The tetramer frames define a pair of cylinders that encompass the folded histone core of each nucleosome, i.e., the protein residues other than the disordered tails. The software uses these objects along with 1-nm spheres centered on the DNA base pairs and the cylindrical representations of the regulatory proteins to check for all possible overlaps. Interactions within DNA separated by less than 30 bp are not considered.

Histone core charges. The charged atoms within the folded protein core — i.e., the 140 cationic N_η and N_γ atoms of arginine and lysine with an assumed charge of +1 esu and the 68 anionic O_δ and O_ε atoms of aspartic acid and glutamic acid with an assumed charge of –1 esu — are clustered on the basis of their interatomic distances into a set of point charges for coarse-grained assessment of electrostatic interactions. The number and size of clusters are determined by an optimization procedure that minimizes the difference at representative points between the electrostatic potential of the simplified model and the corresponding potential determined for all charged atoms in the octamer core. There are 26 such clusters (some involving atoms from different histones) with net charges ranging between –4 esu and +9 esu in the case of the octamer core from the best resolved core-particle structure. The charged atoms on the regulatory proteins are similarly reduced to set of point charges, 12 points on NtrC ranging in magnitude from –10 esu to +1 esu and 39 points on the RNA polymerase-σ^{54} complex with values between –17 esu and +17 esu.

Histone tail charges. The charged residues on the histone tails that lie outside the folded protein core — a total of 80 cationic and 4 anionic residues — are modeled as mobile point charges. Each of the tails is

represented by both a fixed point charge and a set of cationic charges that move at random within a spherical cap centered at the point where the tail exits the protein core. The fixed charge is located on the cationic nitrogen of the residue closest to the center of the globular protein assembly. The number of movable charges is taken as half the sum of the charged atoms in the tail, and the total charge is maintained by doubling the charge per point. The dimensions of the spherical caps (9-20 Å radii) are based on estimates of the radii of gyration of the histone tails obtained from atomic-level simulations of the individual proteins (13). The moves of the tails are rejected if they enter the geometrical objects representing the nucleosome cores, regulatory proteins, and DNA base pairs. Thus, the tails cannot enter the folded protein core but can approach the surface.

DNA polyanion. The anionic charges along the DNA sugar-phosphate backbone are also coarse-grained. The DNA is partitioned into groups of three base-paired nucleotides, and for each group a point charge is placed at the geometric center of the three base-pair origins. Thus the first point charge on DNA lies at or near the center of base-pair 2 and successive point charges lie at or near base-pair 5 up to base-pair $N - 1$ if the chain length N is a multiple of three, or 1-2 bp from the end if the chain length is not a multiple of three. The magnitude of the charge at the selected point is accordingly set to three times that of a single base pair. In most cases we coarse-grain the system with charges of –1.5 esu placed near the center of every third base pair. The cited number reflects the charge on each nucleotide upon neutralization by counterions, i.e., –0.25 esu, times the net charge per base pair and the spacing of base pairs. The percent neutralization is based on Manning's predictions of counterion uptake around B DNA in dilute solution (14). Whether the level of neutralization remains the same in a nucleosome-decorated chain is an open question. Available high-resolution structural data, however, show that the build-up of ions on the surface of DNA (15) resembles the pattern as the charged amino acid atoms in bound proteins (16).

Electrostatic treatment. Electrostatic interactions between charges on DNA and protein are determined using a Debye-Hückel potential with a Debye length computed on the basis of the salt composition/ionic strength of the experimental system, 8.33 Å in the current study. Energy contributions are determined for all point charges separated by a distance less than 10 times the Debye length, i.e., ~80 Å. The interactions include points within and between DNA linkers, between DNA and proteins, between proteins, but not within a given protein.

Configurational sampling. DNA/histone tail moves are based on Monte Carlo Markov Chain (MCMC) sampling and Metropolis selection. The step parameters of proposed new configurations are slight modifications of those in the last accepted configuration. In particular, one linker step is chosen at random and the rigid-body parameters associated with that step are changed, each independently of the other. The proposed increments in step parameters are given by the product of an assumed maximum move, or amplitude, and a random number in the range [–1, +1]. The magnitude of the amplitude is based on the elastic properties of DNA, *i.e*., the deviation of a step parameter that raises the energy by $k_BT/2$ (values given above). The global coordinate frame is located on the first DNA base pair and successive base pairs are related to the global frame by a product of transformation matrices (4). These matrices specify the rigid-body parameters at individual base-pair steps along the chain, including those on the nucleosome and regulatory proteins. Once the coordinates of the base-pair on a nucleosomal dyad are known, the positions of point charges on the histone cores, the axes and origins of tetramer frames, the origins and polar axes of tail half-spheres, etc. are determined with built-in transformations specific for the assumed core-particle structure. The positions of charges on the regulatory proteins are similarly expressed in terms of a coordinate frame at the 5´-terminal base pair of the bound DNA fragment.

The DNA is always sampled in attempting to generate a new configuration but the histone tails are moved only if a random number in the range [0,1] is less than a specified probability $p_{protein}$. The value of $p_{protein}$ is chosen so that configurations are generated with a desired acceptance ratio, here taken to be 0.27 following standard guidelines (17). The choice is based on a series of 'burn-in' runs of ~500,000 trials, where the acceptance ratio is computed for increasing values of $p_{protein}$. This procedure is repeated for each new system, e.g., choice of linker length between nucleosomes. The histone moves include all point

charges on all tails. The moves of point charges are expressed in terms of spherical coordinates within the half-spherical volumes described above. New positions are given by the product of the assumed maximum radial and angular moves times a random number, e.g., [0,1](*R*/10) along the radial coordinate where *R* is the assumed radius of the half-sphere of a given tail, [0,1](π/10) in the azimuthal direction, and [–1,1](2π/10) with respect to the polar axis. Moves are rejected if collisions of five different types are detected: sphere-sphere, cylinder-sphere, cylinder-cylinder, point-sphere, point-cylinder, where sphere refers to a DNA base pair, cylinder to a histone tetramer or regulatory protein, and point to part of a histone tail. Moves free of collisions are accepted according to the Metropolis criterion.

Enhancer-promoter interactions. Given the relative stiffness of the nucleosome-decorated DNA chains, we use a biased Monte Carlo procedure (18) to estimate the likelihood that the proteins on the enhancer and promoter sites come into direct contact. For example, only 469 of the 2422 base-pair steps in a saturated 13-nucleosome construct with 177-bp spacing undergo local fluctuations. Moreover, many of these deformations introduce collisions of protein and/or DNA that further limit the overall motions of the modeled system. We address this problem by determining a series of conditional probabilities $p(r_i|r_{i+1})$ that the distance r between the centers of the regulatory proteins is less than or equal to a value of r_i in a subset of configurations where the interprotein distance is no larger than r_{i+1}.

We find these subsets by rejecting any trial configurations with $r > r_{i+1}$ and counting the number of accepted configurations that meet the smaller distance criterion. The probability $p(r_0)$ that the distance between the proteins comes within a desired value r_0 over all possible configurations is then obtained from the product of the $p(r_i|r_{i+1})$ over a set of distance constraints $r_0 < r_1 < \cdots < r_n$, where r_n is larger than or equal to the maximum inter-protein separation, i.e., $p(r_0) = \prod p(r_i|r_{i+1})$, where $i = 0,n–1$. Here we set r_0 to 200 Å and evaluate the probability that the distance between regulatory proteins matches this limit using a sequential set of 6-10 conditional probabilities. The starting state for the determination of each $p(r_i|r_{i+1})$ value is the last structure generated in the computation of $p(r_{i-1}|r_i)$, and that for the lowest distance constraint a compact configuration with near-r_0 interprotein separation taken from an initial unconstrained Monte Carlo sample of ~10^4 chains (19), collected after every 200 accepted moves in a simulation of over a million total steps. Conditional probabilities are based on ~10^3 configurations matching the desired upper limit on r collected from several million simulation steps. The present calculations do not include constraints on the relative orientation of appropriately separated regulatory proteins.

We repeat the procedure for DNA chains of the same length in the absence of nucleosomes and estimate the enhancement of enhancer-promoter communication in the presence of nucleosomes from the quotient of the computed probabilities that the distance between regulatory proteins comes within the desired range in the two systems, i.e., $p_{\mathrm{nsm+DNA}}(r_0) \div p_{\mathrm{DNA}}(r_0)$. The statistical error is estimated by repeating the calculations for different values of r_0, here 175 and 225 Å.

References

1. 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.
2. Arya, G., Zhang, Q. and Schlick, T. (2006) Flexible histone tails in a new mesoscopic oligonucleosome model. *Biophys. J.*, **91**, 133-150.
3. Langowski, J. (2006) Polymer chain models of DNA and chromatin. *Eur. Phys. J. E Soft Matter*, **19**, 241-249.
4. Czapla, L., Swigon, D. and Olson, W.K. (2006) Sequence-dependent effects in the cyclization of short DNA. *J. Chem. Theor. Comp.*, **2**, 685-695.
5. 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., USA*, **102**, 5397-5402.
6. Horowitz, D.S. and Wang, J.C. (1984) Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *J. Mol. Biol.*, **173**, 75-91.

7. Heath, P.J., Clendenning, J.B., Fujimoto, B.S. and Schurr, J.M. (1996) Effect of bending strain on the torsion elastic constant of DNA. *J. Mol. Biol.*, **260**, 718-730.
8. Davey, C.A., Sargent, D.F., Luger, K., Mäder, A.W. and Richmond, T.J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.*, **319**, 1087-1113.
9. Murakami, K.S., Masuda, S. and Darst, S.A. (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science*, **296**, 1280-1284.
10. Lee, S.-Y., De La Torre, A., Yan, D., Kustu, S., Nixon, B.T. and Wemmer, D.E. (2003) Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes Dev.*, **17**, 2552-2563.
11. De Carlo, S., Chen, B., Hoover, T.R., Kondrashkina, E., Nogales, E. and Nixon, T. (2006) The structural basis for regulated assembly and function of the transcriptional activator NtrC. *Genes Dev.*, **20**, 1485-1495.
12. Smith, R. (2001–2014) http://www.ode.org.
13. Potoyan, D.A. and Papoian, G.A. (2001) Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J. Am. Chem. Soc.*, **133**, 7405-7415.
14. Manning, G.S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.*, **11**, 179-246.
15. Auffinger, P., D'Ascenzo, L. and Ennifar, E. (2016) Sodium and potassium interactions with nucleic acids. *Met Ions Life Sci.*, **16**, 167-201.
16. Olson, W.K., Colasanti, A.V., Li, Y., Ge, W., Zheng, G. and Zhurkin, V.B. (2006) In Sponer, J. and Lankas, F. (eds.), *Computational Studies of RNA and DNA*. Springer, Dordrecht, The Netherlands, pp. 235-257.
17. Gelman, A., Gilks, W.R. and Roberts, G.O. (1997) Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, **7**, 110-120.
18. Podtelezhnikov, A.A. and Vologodskii, A.V. (2000) Dynamics of small loops in DNA molecules. *Macromol.*, **33**, 2767-2771.
19. Metropolis, N.A., Rosenbluth, A.W., Rosenbluth, M.N., Teller, H. and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087-1092.

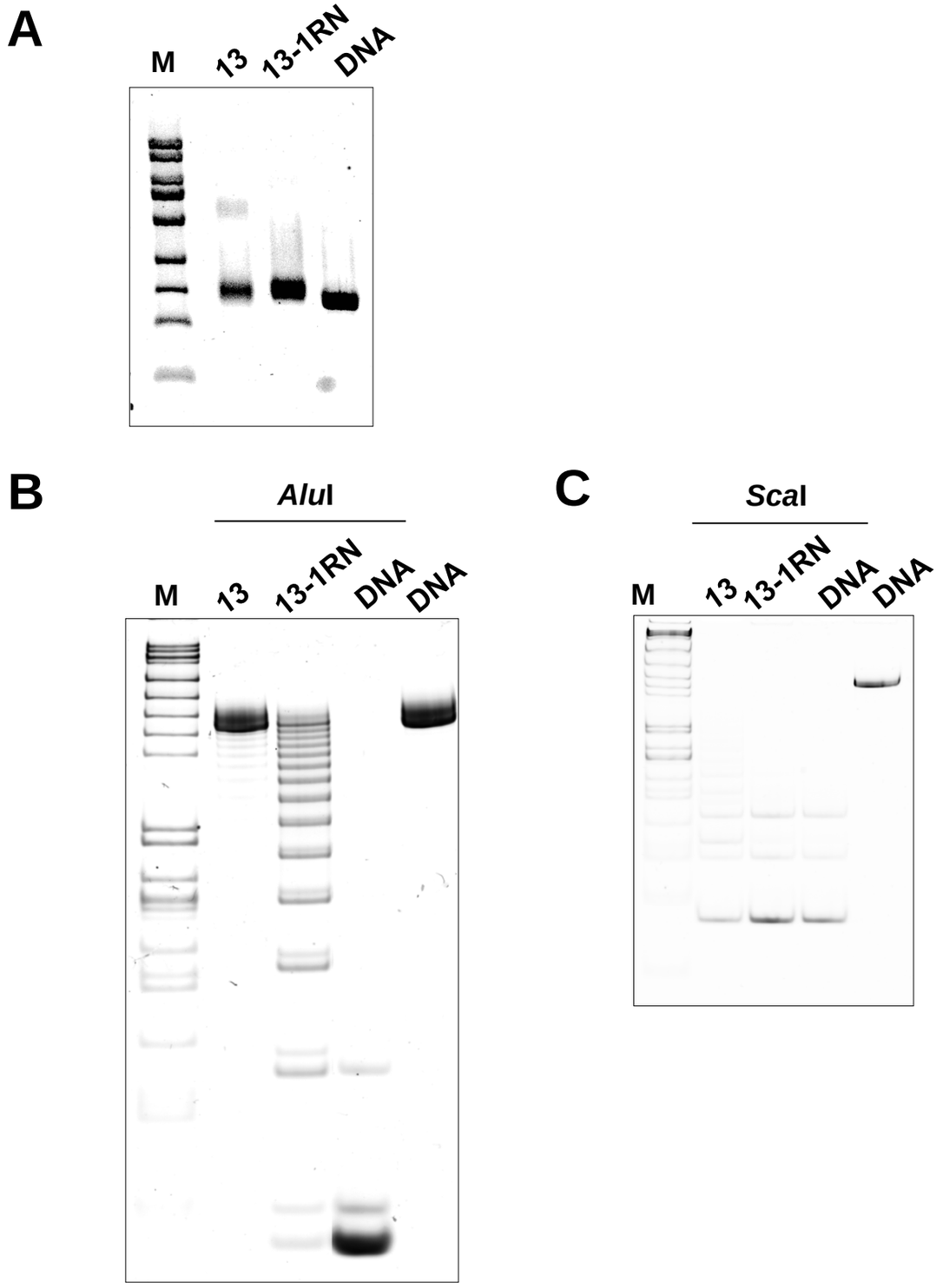

Figure S1. Analysis of nucleosome occupancy of 601_{177} arrays. (A) Saturated nucleosomal arrays (13), arrays missing a single nucleosome in random positions (13-1RN) or DNA templates were characterized by electrophoresis in 0.8% agarose gel. (B, C) Analysis using restriction enzyme sensitivity assay. The arrays or DNA templates were incubated in the presence of (B) *Alu*I cutting in the middle of the 601 sequences, or (C) *Sca*I cutting between the 601 sequences. Analysis of purified DNA by non-denaturing PAGE. Staining with ethidium bromide. M: 1 kb Plus DNA ladder.

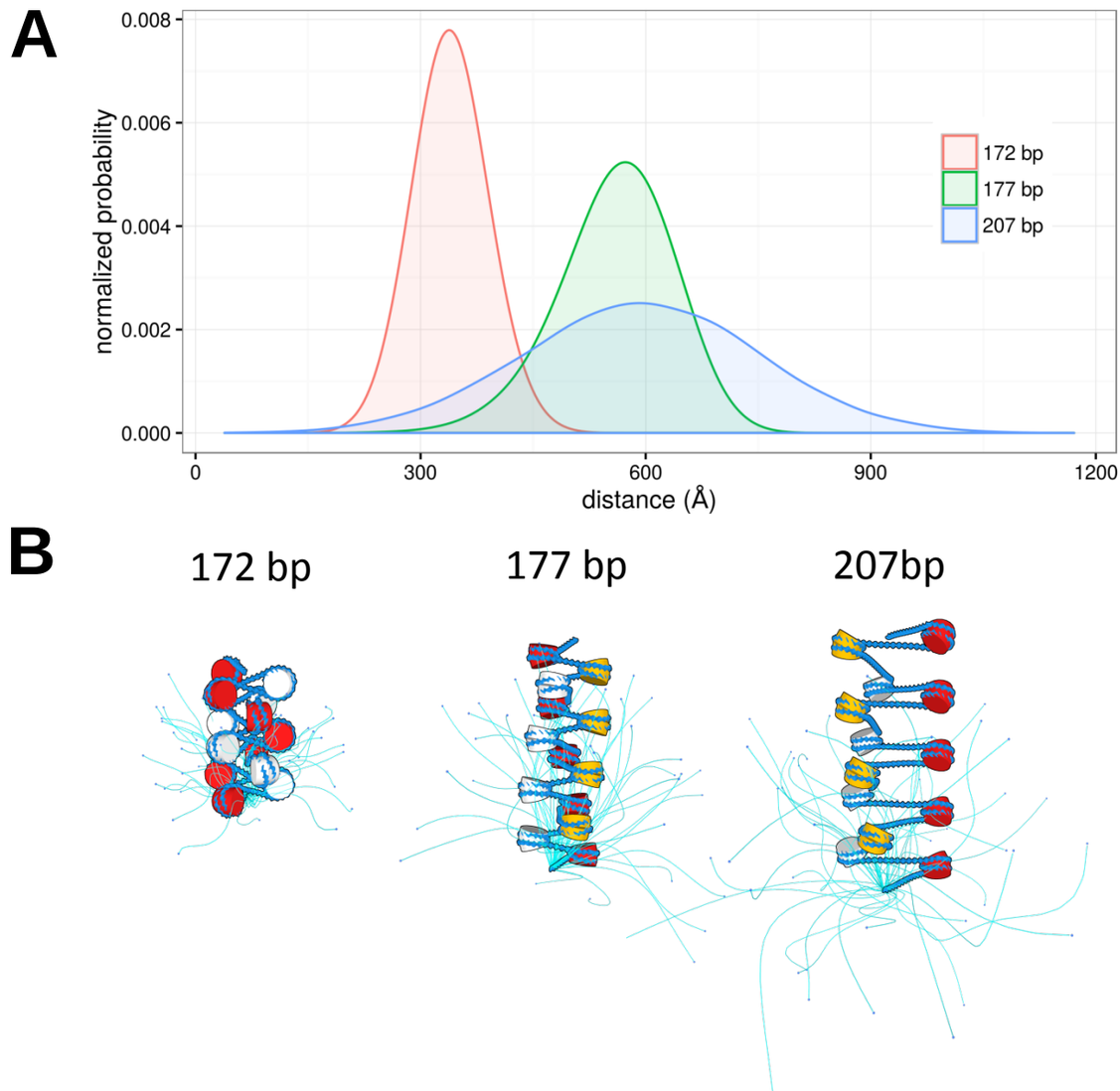

Figure S2. Changes in internucleosomal spacing lead to very different chromatin structures. (**A**) Distribution of the distances between the ends of simulated 13-nucleosome constructs with 172, 177, and 207-bp nucleosome repeat lengths. Note the decrease in chain extension and deformability (downward shift and narrowing of the distribution) that arises from the 5-bp decrease in linker length in the 172-bp construct and the limited increase in extension and wide broadening of the distribution associated with the 30-bp increase in linker length in the 207-bp construct. (**B**) 'Average' and 'ensemble of trajectories' of representative chain configurations of the modeled constructs. The twisted blue ribbons trace the average pathways taken by successive DNA base pairs in the collections of modeled structures. The ensemble of trajectories (thin lines emanating from the 5′-ends of the average structures and terminating at the small dots on their 3′-ends) illustrate the influence of the nucleosome repeat on overall chain deformability. The color-coding of successive nucleosomes (wedge-shaped objects that enclose the folded histone proteins along the average structures) emphasizes the different (two- and three-start) modes of superhelical nucleosome association in the three systems.

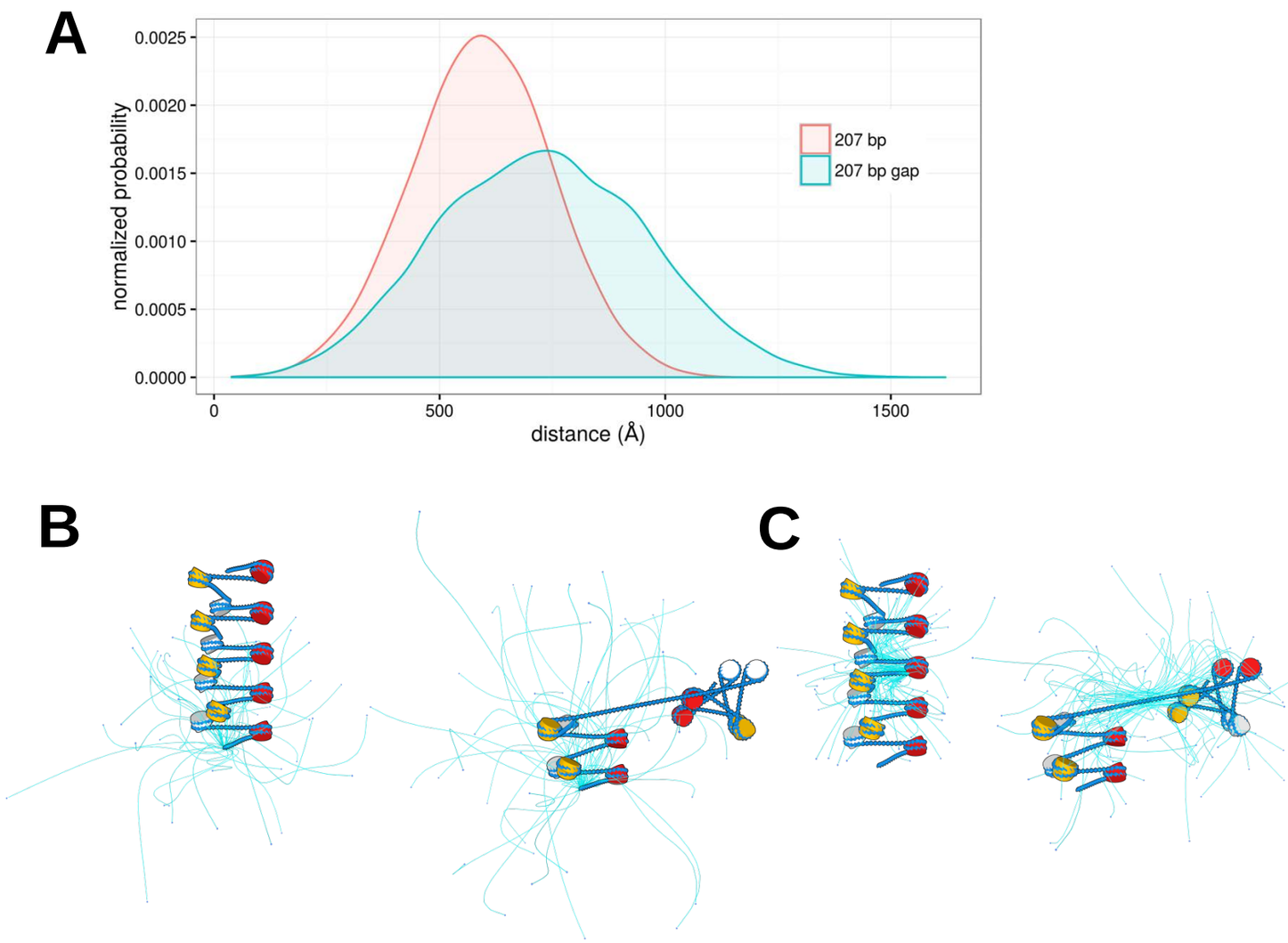

Figure S3. The presence of a single nucleosome-free gap enhances the flexibility of modeled chromatin structures. (**A**) Distribution of the distances between the ends of a saturated 13-nucleosome array with 207-bp inter-nucleosomal spacing and the same array missing the central nucleosome. Note the increase in chain extension and deformability (upward shift and broadening of the distribution) found in the absence of the nucleosome. (**B, C**) 'Average' and 'ensemble of trajectories' of representative chain configurations of the modeled constructs. The fully saturated structure on the left in B is depicted as in Figure S1. The 5′-nucleosome on the gapped structure to its right is oriented in the same way as the corresponding nucleosome on the saturated structure to highlight the changes in average chain direction and global flexibility that arise upon removal of the central nucleosome. The ensemble of trajectories in C are superimposed on the central base of the average structures to highlight the 'flexible joint' that arises at the gap. Note the different direction of overall bending in the gapped vs. intact constructs relative to the start of the fiber.

Table S1. Cloning approaches.

Obtained plasmid	Amplified/Cut from	Cloned into
pSW01	Double-stranded oligonucleotide 5'-ATTACAGGATCCGTCTAGATAGATCTTACGTA-3' 3'-TAATGTCCTAGGCAGATCTATCTAGAATGCAT-5' The oligonucleotide was cleaved by *Bgl*II/*Bam*HI	pYP04/ *Bgl*II
pSW08	177bp 601 fragment amplified from pSG183 plasmid using the following oligonucleotides: 5'-ATCCTTAGATCTCACATGCACAGGAT-3' 5'-ATCCTTAGATCTCCCTATACGCGGCC-3' The amplified product was cleaved by *Bgl*II	pYP04/ *Bgl*II
pSW12	pSG187/*Spe*I/*Xba*I	pSW01/ *Xba*I
pEN12	12x172SG/*Spe*I/*Xba*I	pSW01/ *Xba*I
pSW6	pSG217/*Spe*I/*Xba*I	pSW01/ *Xba*I
pSW6N	Double-stranded oligonucleotide 5'-CTAGAATGCCCATGGACGAA-3' 3'- TTACGGGTACCTGCTTGATC-5'	pSW6/ *Xba*I
pSW61	Lambda DNA amplified by oligonucleotides: 5'-GCACCTCCATGGGCTGTACTACAGCGTCTAAG-3' 5'-GCACCTCCATGGGATGGGCTACAGATACCTGAC-3' Amplified product was cleaved by *Nco*I	pSW6N/ *Nco*I
pSW616	pSG217/*Spe*I/*Xba*I	pSW61/ *Xba*I

Table S2. P-values for the indicated pairs of arrays determined by t-test.

	Array	**Array**	***P-value***
Figure 3	207 NRL	177NRL	0.03
	207 NRL	172 NRL	0.001
	177 NRL	172 NRL	0.005
Figure 4	4	4-1RN	0.0001
	7	7-1RN	0.005
	8	8-1RN	0.01
	10	10-1RN	0.06
Figure 6	13-1CN	13	0.01
	13-1CN	13-1RN	0.45
	13	13-1RN	0.02
Figure 7	13-1RN	13-1RN-1CN	0.001