Supplementary Information: Molecular Mechanism of Spontaneous Nucleosome Unraveling

David Winogradoff[a,b], Aleksei Aksimentiev[a,b,c]

[a] *Center for the Physics of Living Cells,*
[b] *Department of Physics,*
[c] *Beckman Institute for Advanced Science and Technology,University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

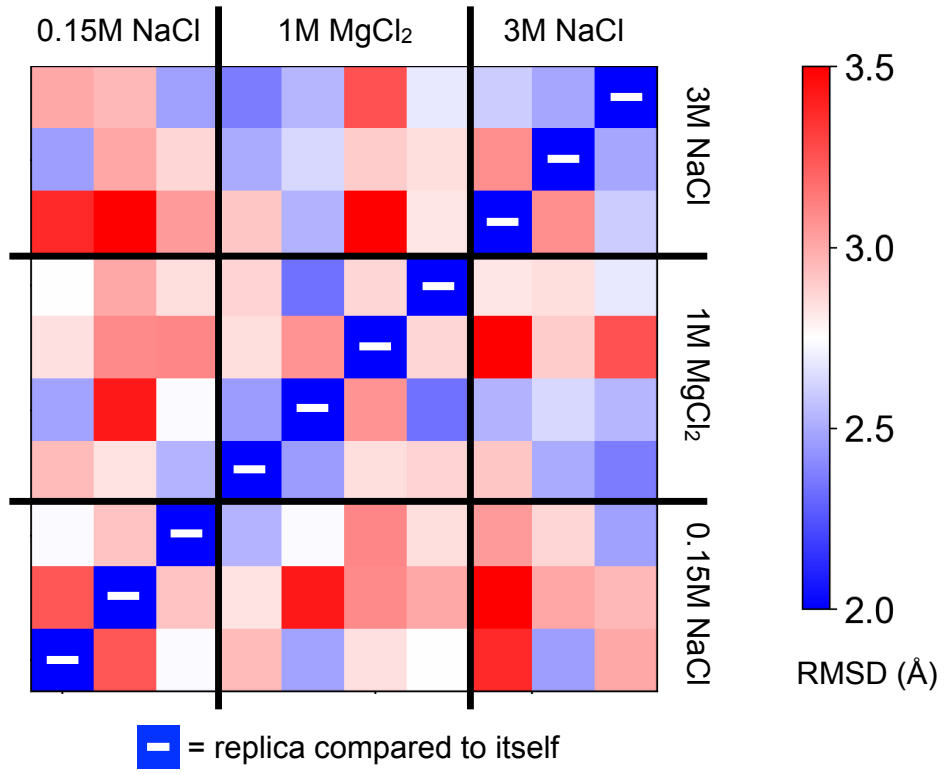


Figure S1 : Pairwise RMSD comparison of replicas of 601L nucleosomes. Each row and column represents one of ten independent replicas of 601L. This 10×10 matrix displays all their pairwise RMSD values—blue to red corresponds to low to high RMSD. The conformation at $t = 0$ (the first frame of production, after structural relaxation) was taken for each. The RMSD for each pair of unique replicas was calculated for non-hydrogen nucleic atoms, after alignment based on non-hydrogen protein atoms. The dark black lines separate the matrix into 3×3 blocks by ionic conditions, identified by labels along the top and right-hand side. Blue squares with a white minus sign along the diagonal are where replicas would be compared to themselves.

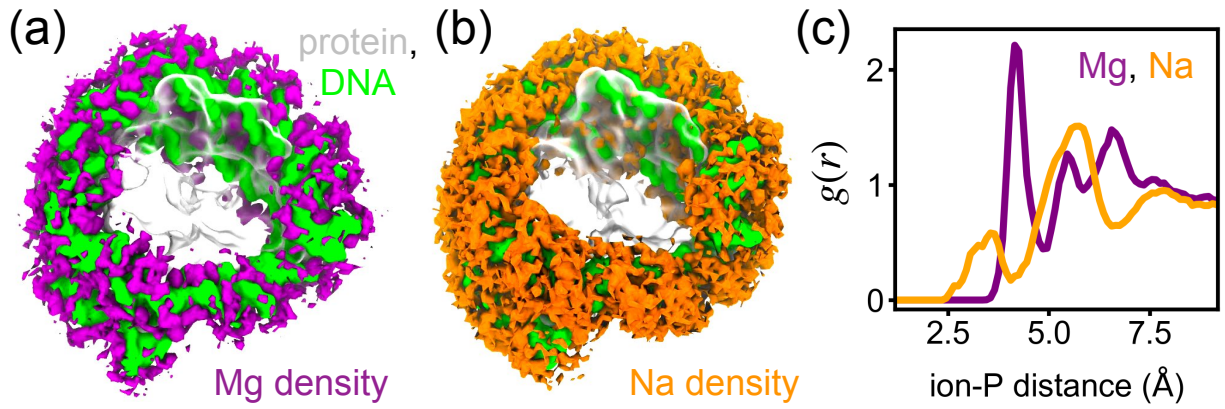


Figure S2 : (a, b) Distribution of cations around DNA in high ionic strength electrolytes. The distribution is visualized by drawing 2 ions/nm^3 isosurfaces for Mg^{2+} (purple, 1 M $MgCl_2$) and Na^+ (orange, 3 M NaCl) within 10 Å of nucleosomal DNA. Only frames in which the DNA is fully wrapped were considered for this analysis. The histone core and the surrounding DNA are shown as semi-transparent white and solid green isosurfaces, respectively. (c) Radial distribution function $g(r)$ between phosphorus atoms of DNA and Mg^{2+} or Na^+ ions at 1 M $MgCl_2$ or 3 M NaCl simulation, respectively.

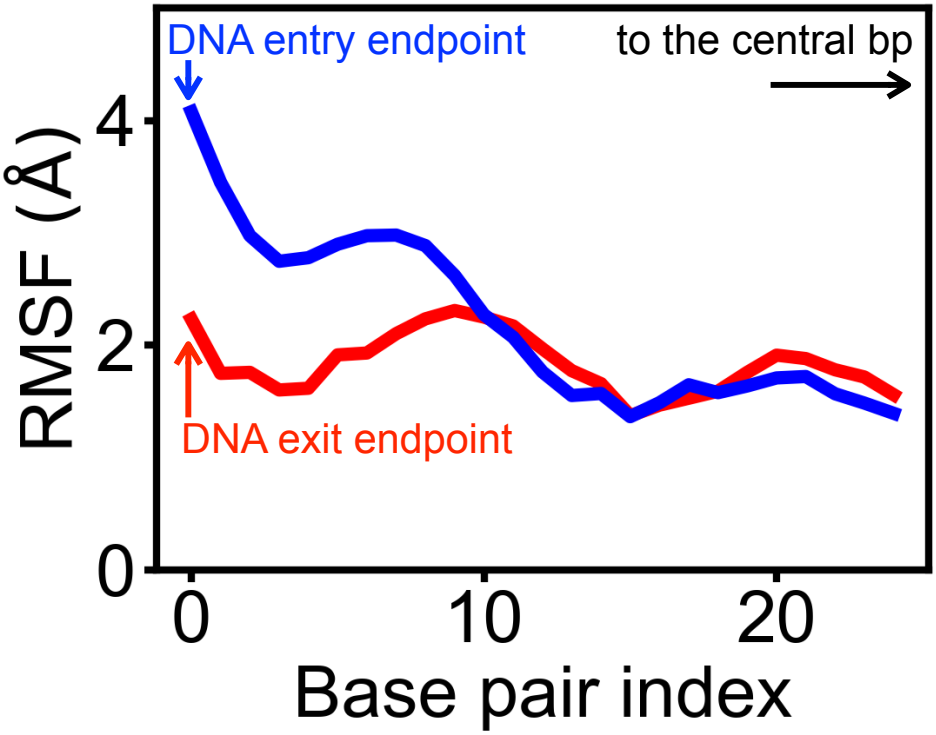


Figure S3 : Root mean squared fluctuations (RMSF) of the first (blue) and last (red) 25 bp of DNA as a function of base pair index for a 601L nucleosome simulated in 1 M $MgCl_2$. Matching the time frame depicted in main text Fig. 1e, only simulation times from 0.7 to 0.85 μs of the trajectory of the 601L nucleosome in 1 M $MgCl_2$ were considered.

System	DNA sequence
Widom 601L	−70 −60 −50 −40 −30 −20 −10 0 5′–ATCACAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGGAA TCCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTGTGAT–3′ +10 +20 +30 +40 +50 +60 +70
poly-AT	−70 −60 −50 −40 −30 −20 −10 0 5′–ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA–3′ +10 +20 +30 +40 +50 +60 +70
5S RNA gene	−70 −60 −50 −40 −30 −20 −10 0 5′–ACGAATAACTTCCAGGGATTTATAAGCCGATGACGTCATAACATCCCTGACCCTTTAAATAGCTTAACTTTCA TCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCACCGAAGTCAAGCAGCAT–3′ +10 +20 +30 +40 +50 +60 +70
poly-GC	−70 −60 −50 −40 −30 −20 −10 0 5′–GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG–3′ +10 +20 +30 +40 +50 +60 +70
Central 85bp of 601	−40 −30 −20 −10 0 5′–TGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGC TGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTC–3′ +10 +20 +30 +40

Table S1 : List of the DNA sequences used in the MD simulation of nucleosomes. The table lists the nucleotide sequence of one of the two DNA strands (strand I), from 5′ to 3′ ends.

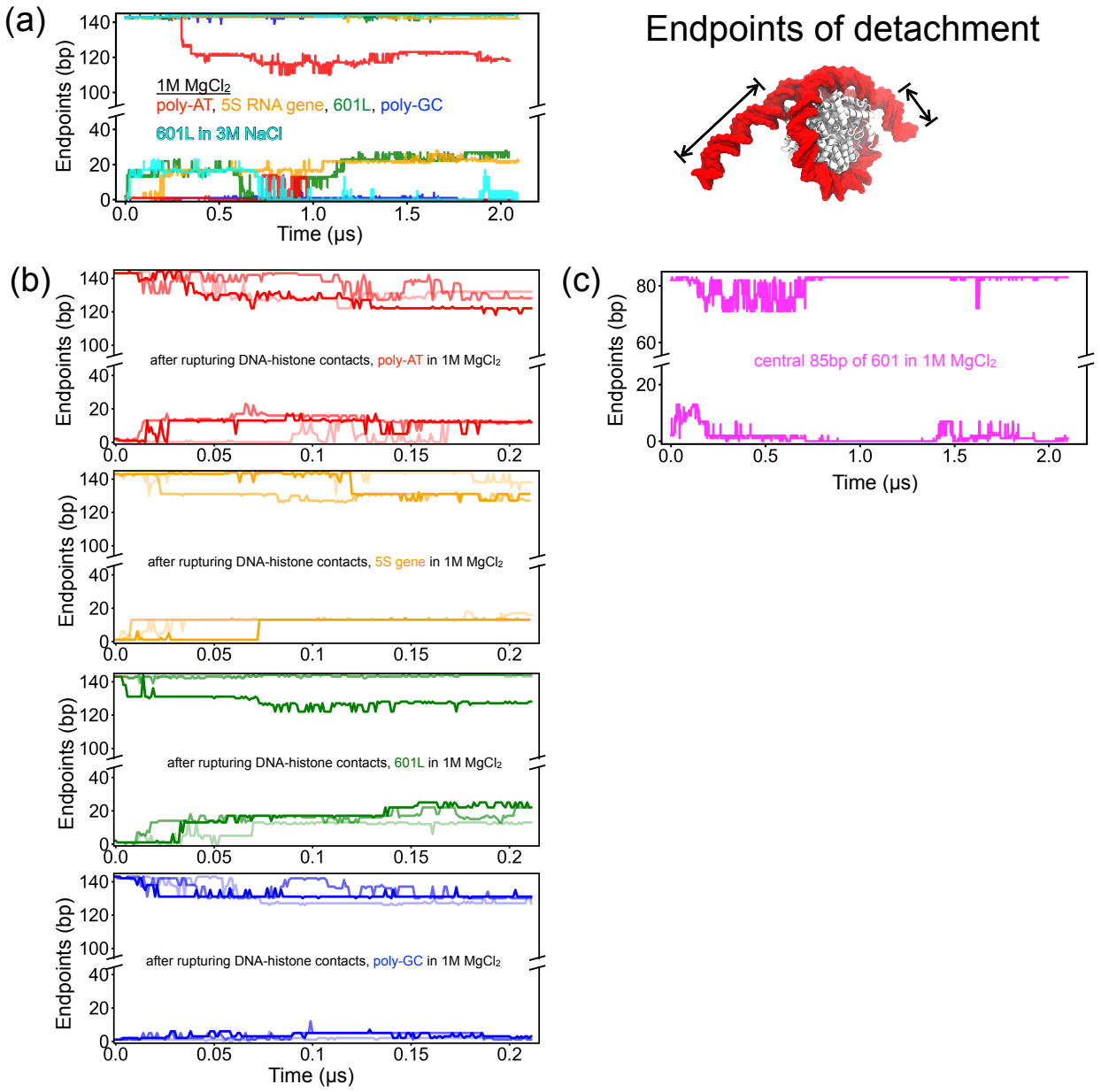


Figure S4 : Nucleosome unwrapping from both endpoints. (a) Terminal endpoints of nucleosomal DNA in contact with protein (in bp) as a function of simulation time. Zero and 144 correspond to the entry and exit sides, respectively. Right, an example of the two endpoints of poly-AT in 1M $MgCl_2$ in contact with the histone core at 0.8 μs. (b) Endpoints of DNA in contact with the histone core after rupturing key histone-DNA contacts. Replicas shown in different opacities. (c) Endpoints of DNA contact for a nucleosome system containing only the central 85 bp of 601 DNA.

Supplementary Note 1: Persistence length of DNA at physiological and high salt conditions

We investigated the effect of ionic conditions on DNA rigidity by performing additional all-atom MD simulations of five different 45-bp fragments of dsDNA in physiological and high salt. Defined in Fig. S5a, DNA sequences were chosen as the terminal segments of polyAT, 601L, polyGC, and the two ends of 5Sgene (since it alone is non-palindromic). A length of 45-bp was chosen because it extends beyond the longest continuous stretch of nucleosomal DNA to spontaneously detach from the histone core. Initial, idealized conformations of the segments of dsDNA were generated using the *Nucleic Acid Builder* tool from Amber12 [1]. Then, the segments were immersed in neutral aqueous solutions of either 0.150 M NaCl or 1.0 M $MgCl_2$. Matching the parameters defined in the "General simulation protocols" of Materials and Methods, minimization, structural relaxation, and initial unrestrained equilibration simulations of 10 ns were performed using the gromacs 5.0.4 MD package [2], and 0.74-μs production runs were performed in the NPT ensemble on the D. E. Shaw Research Anton2 supercomputer [3].

An example illustrated in Fig. S5b, instantaneous directional vectors were determined for increments of 10bp for the central 40bp — t_0, t_1, t_2, t_3 — using single-value decomposition for each simulation snapshot of the trajectory. From atomistic simulations, we estimated the *dynamic* persistence length, L_d, following the formalism of Ref. 4, defined as

$$e^{-s/L_d} = \frac{\langle t_0 \cdot t_i \rangle}{\hat{t_0} \cdot \hat{t_i}}, \tag{S1}$$

where s denotes contour length, t_i an instantaneous directional vector for the i^{th} 10bp fragment, $\langle \cdot \rangle$ indicating an ensemble average over all pairs of such fragments, and $\hat{t_0} \cdot \hat{t_i}$ is a normalization factor, accounting for the average intrinsic curvature of the segment of DNA. Contour length was calculated as a linear function of the number of bp, 3.4-nm per 10bp of separation. The raw data from the MD simulations is plotted in Fig. S5c. Shown in

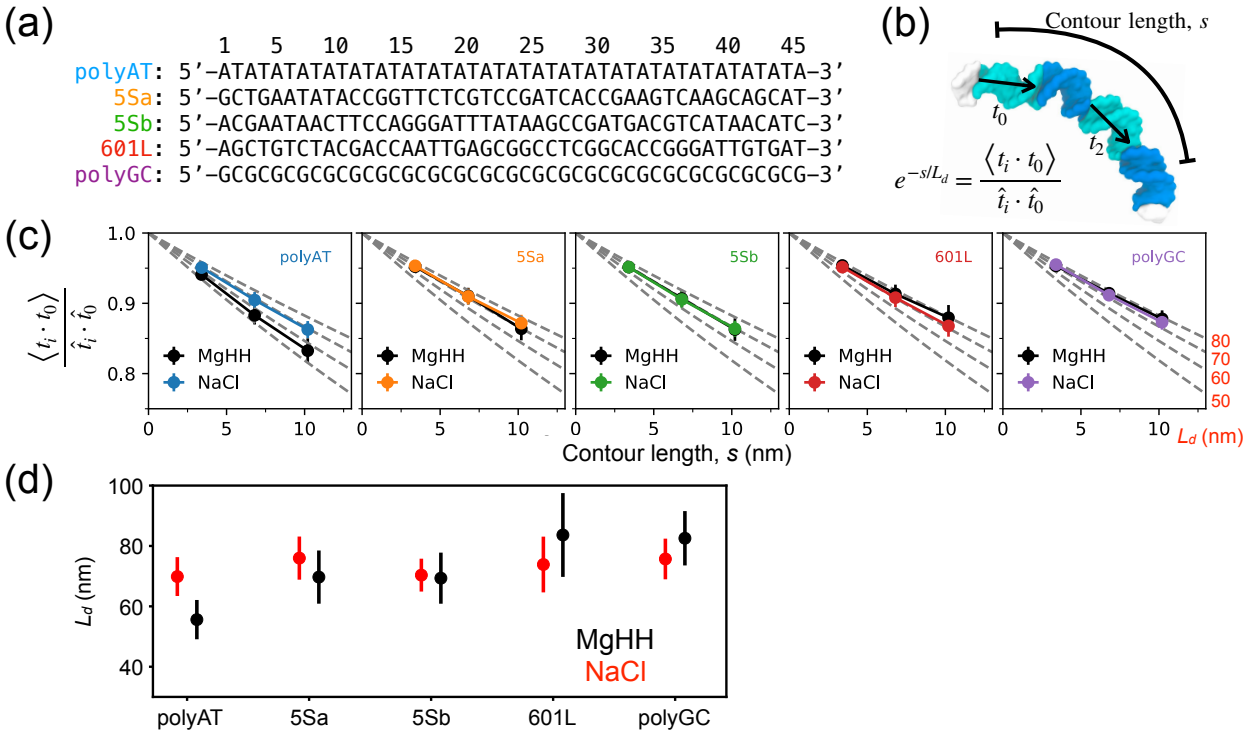


Figure S5 : DNA rigidity in physiological and high salt. (A) DNA sequences considered, 45bp each. (B) Schematic illustration of the persistence length calculations. The first three and final two bp (in white) were excluded for the purpose of analysis. For the remaining 40bp (in light and dark blue), directional vectors were determined using for the initial 10bp t_0, and for other 10bp segments t_i. Following Ref. 4, these instantaneous vectors were normalized by $\hat{t_0} \cdot \hat{t_i}$, an average accounting for the structure's intrinsic curvature. $\langle \cdot \rangle$ denotes an ensemble average. (C) Average normalized dot product of t_0 and t_i as a function of contour length. 'MgHH' (black curves) and 'NaCl' (colored curves) correspond to conditions of 1.0 M $MgCl_2$ and 0.15 M NaCl, respectively. Error bars represent standard deviation. When more than one pair of 10bp are compared, *i.e.* for the first two points of each panel, the geometric mean of the standard deviations was calculated. Dashed gray lines are theoretical curves for L_d values ranging from 50-80 nm. (D) Dynamic persistence length, L_d, for the DNA segments considered. For each system, colors highlight L_d in 1.0 M $MgCl_2$ (black) and 0.15 M NaCl (red). L_d values obtained by exponential fits to the data in panel C. System names, defined in panel A, provided at the bottom.

Fig. S5d, the average L_d values for each system indicate that ionic conditions do not have a major influence on the segments of DNA considered here: the largest effect of increasing salt from 0.15 M NaCl to 1.0 M $MgCl_2$ on DNA's rigidity can be seen for polyAT, for which the DNA is slightly more flexible in high salt, a difference of about two standard deviations; for the rest of the DNA sequences considered, the average values are within one standard deviation. Overall, these results confirm ionic conditions do not have a significant effect on DNA's rigidity.

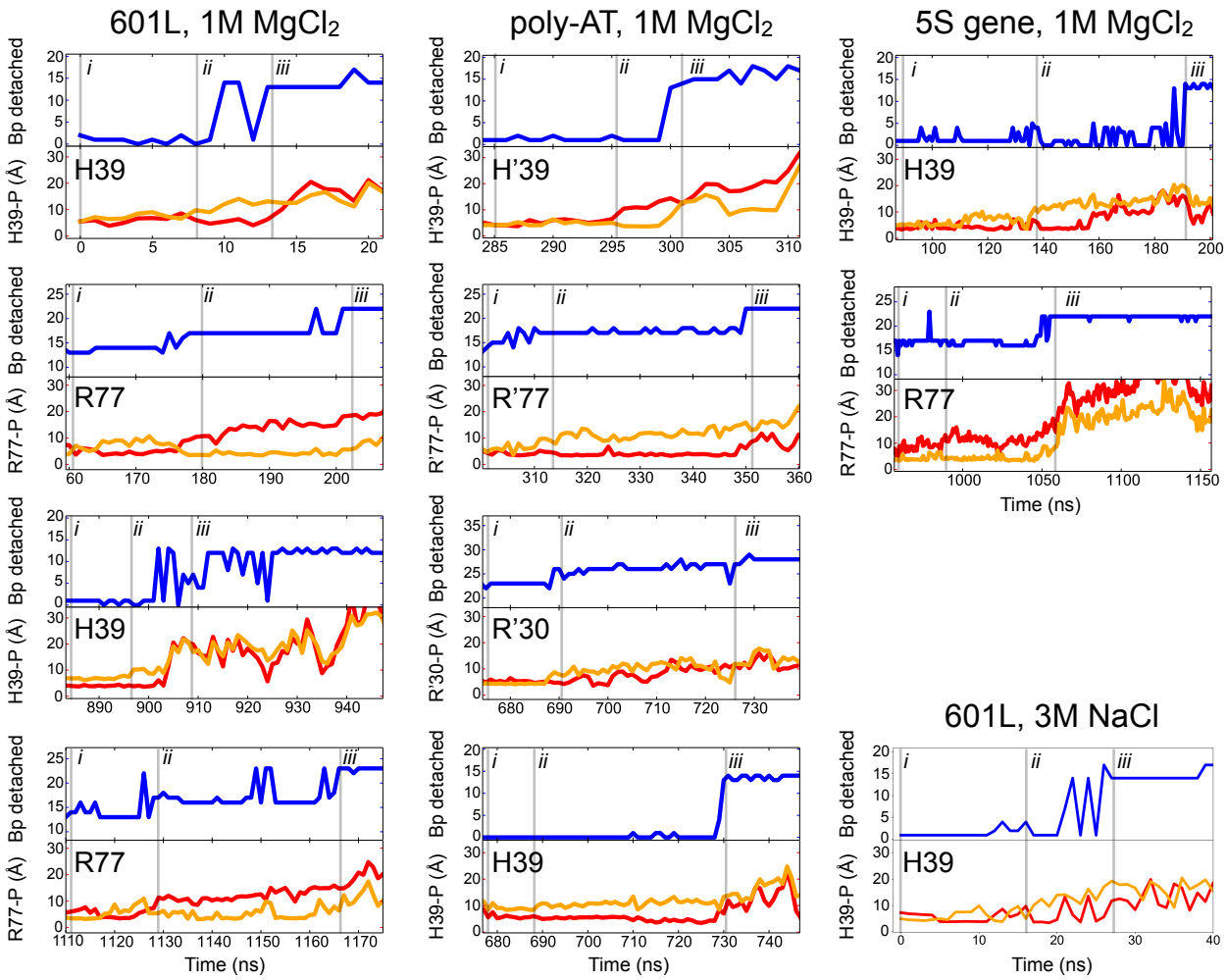


Figure S6 : Stepwise unraveling of nucleosomal DNA. In each panel, the number of endpoint bp detached (top) and the distance from specific histone residues to the nearest phosphorus atoms of the I (red) and J (orange) strands of nucleosomal DNA (bottom) are plotted *versus* simulation time. Note, base pairs detached are plotted only from one end of the nucleosomal DNA, matching the side on which the detachment is occurring. The time intervals featured in the individual panels correspond to the major DNA unbinding events observed in the 2 μs trajectories of the 601L, poly-AT and 5S systems, see main text Fig. 2a. Panel labels specify abbreviated names of the histone residues that control the DNA unbinding process: H3 His39 (H39), H3 Arg77 (R77), and H2A Arg30 (R30). The distance from these residues to the DNA strands was measured using center of mass coordinates of the guanidinium or α-amino groups of the arginine or histidine residues, respectively. A prime superscript or absence thereof denotes the histone copy, *e.g.* R77 stands for H3 Arg77 and R′77 for H3′ Arg77. Vertical lines indicate transitions in DNA conformation from being tightly bound to the histone core by a charged residue (*i*) to only being partially bound to the core (*ii*) and to being detached completely from the core (*iii*).

System	$i \rightarrow ii$		$ii \rightarrow iii$		Δt
	time	Δbp	time	Δbp	
601L, 1M $MgCl_2$	7	0	14	11	7
	180	4	205	6	25
	895	0	910	12	15
	1130	4	1165	6	35
601L, 3M NaCl	17	4	27	11	10
5Sgene, 1M $MgCl_2$	135	4	190	8	55
	990	3	1060	6	70
poly-AT, 1M $MgCl_2$	295	0	302	12	7
	312	6	350	4	38
	690	3	725	5	35
	690	0	730	13	40

Table S2 : Timeline of nucleosomal DNA detachment from the histone core. For each nucleosomal system considered, the table specifies the time of occurrence (in ns) and number of bp detached associated with the steps of each specific unbinding transition from the beginning of the unrestrained equilibration. Each row corresponds to one detachment event. The $i \rightarrow ii$ columns specify the time of occurrence and number of bp detached in the transition from DNA being fully-bound to a charged histone residue (i) to being partially bound (ii). The $ii \rightarrow iii$ columns specify the time of occurrence and number of bp detached in the transition from DNA being partially bound to the histone core (ii) to being detached completely (iii). The full course of each detachment event is illustrated in Supplementary Fig. S6 and main text Fig. 3a. In those figures, the "ii" and "iii" vertical lines correspond to the times specified in the second and fourth columns, respectively. The Δt column specifies the time the DNA spent in the intermediate partially-bound state. Because of the way the nucleosomal DNA is twisted, the two phosphates that form the minor groove pocket are about 4 ($\pm$1) bp apart along the DNA helical axis. Which of the two phosphate-histone contacts break first determines whether we see 0 or $\sim$4 bp detached in the intermediate step ($i \rightarrow ii$).

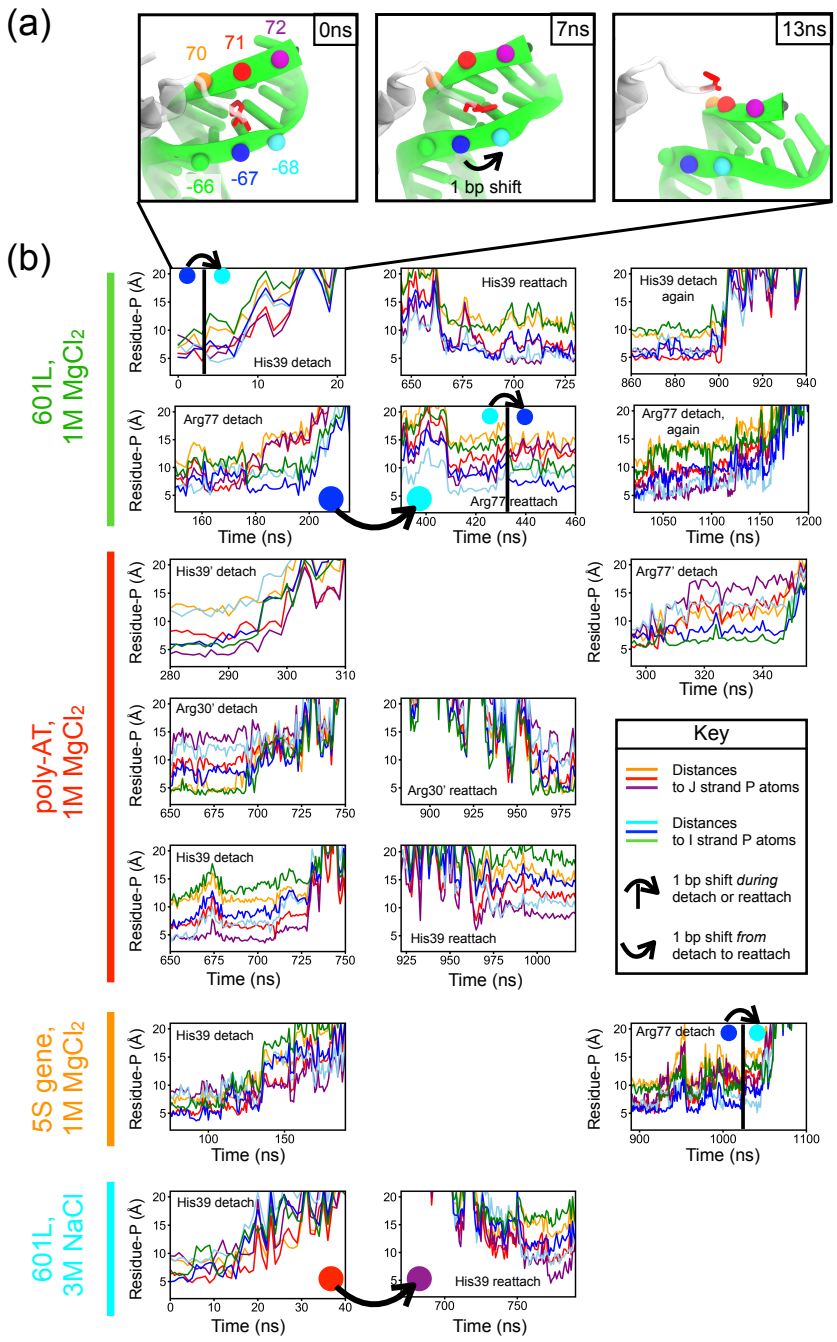


Figure S7 : Shifting accompanies nucleosome unwrapping and rewrapping. (a) Example of 601L DNA (green) spontaneously unwrapping from a histone core (semi-transparent). Spheres highlight the three nearest phosphorus atoms on one strand (orange, red, purple) and from the other strand (green, blue, cyan). At 7ns, there was a 1-bp shift in the nearest phosphorus to H3 histidine 39 (red tubes), from blue to cyan, before detachment occurred. (b) Distances between key histone residues—H3 His39, H2A Arg77, or H2A Arg30—and the six nearest phosphorus atoms, three from each strand of DNA, plotted as a function of time during spontaneous detachment or attachment. Individual bp shifts are highlighted during either detachment or reattachment (curve above the subplot), or a 1-bp shift in position from detachment to reattachment (curve beneath). Systems specified in labels to the left.

Supplementary Methods 1: Forced rupture of histone-DNA contacts

Bonds were ruptured between specific histone residues and the DNA without affecting the overall structure of the nucleosome in independent simulations of the poly-AT, 5S gene, 601L, and poly-GC nucleosomes carried out in 1.0 M $MgCl_2$. After performing minimization and structural relaxation (steps listed in Supplementary Table S4), where velocities were initiated with a different random seed, external forces were applied to specific histone residues to move them away from the DNA minor groove (over 10 ns), and those forces were maintained for the entire production MD simulation.

Supplementary Fig. S8 schematically illustrates the simulation protocol used. The ten target residues were subject to pairwise distance restraints to nine unique anchor residues of the histone core; the force constant of each restraint was 418 kJ mol^{-1} nm^{-2}. Each restraint was applied between either the guanidinium carbon or α-amino nitrogen of the target arginine or histidine residue, respectively, and the α carbon atom of the anchor residue. The target residues were chosen because of their proximity to the minor groove of the nucleosomal DNA in the MD trajectory. The anchor residues were chosen to guide the motion of the target residues' side chains away from the DNA minor grove. Table S3 specifies the residue pairs subject to the distance restraints along with the initial and target distance of each such restraint.

By measuring the protein root-mean-square deviations (RMSD) not including the ten shifted residues, we confirmed that the conformation of the rest of the histone core remained relatively unchanged, in comparison to 601L simulated without restraints (Supplementary Fig. S8, c). For both, the primary peak in the RMSD distribution is centered between 3.0 and 3.5 Å, indicating that the protein core remained stable.

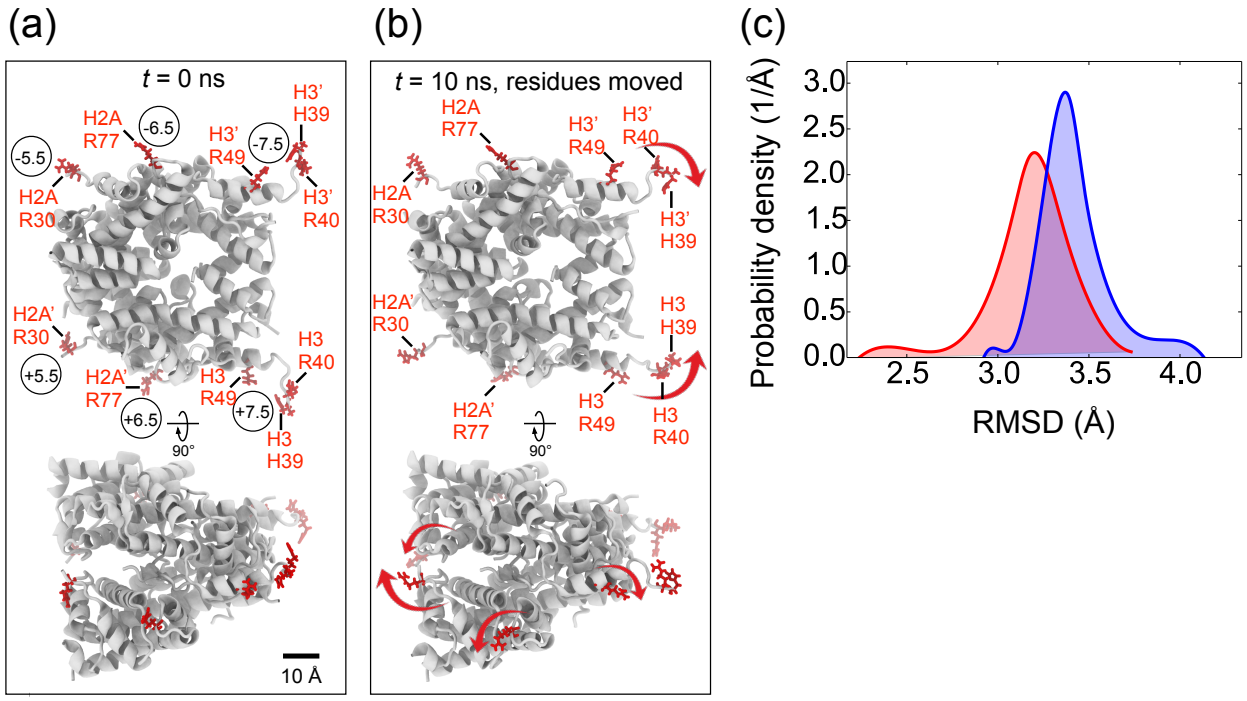


Figure S8 : Illustration of the simulation protocol used to break the bonds between select histone residues and DNA. Microscopic conformation of the specified histone residues before (A) and after (B) a 10 ns MD simulation in which the residues' side chains were moved away from the DNA. The histone proteins are shown in white, the target histone residues in red, and the DNA is not shown for clarity. Curved arrows indicate the general directions in which the side chains were moved. Circled values denote the number of turns from the central bp of the nucleosomal DNA. The ten histone residues that were moved away from the DNA are labeled in panels A and B. A prime superscript, or absence thereof, denotes the histone copy. For example, "H2A′ R77" denotes the second copy of histone H2A, residue Arg77. (C) Distribution of the RMSD values of the histone backbone atoms from the crystallographic coordinates for the 601L nucleosome systems simulated without any restraints (red) and the same nucleosome system simulated in the presence of the pairwise restraints that moved ten residues away from DNA (blue). The residues that were subjected to restraints were not included in the RMSD calculation. The RMSD values were collected over the entire 2 μs length of the simulation. Both systems were simulated at 1 M $MgCl_2$.

Target residue	Anchor residue	Initial distance	Target distance
H3 His39	H3 Pro43	14.3	5.0
H3 Arg40	H3 Pro43	8.6	5.0
H3 Arg49	H2A′ Gln112	18.8	5.0
H2A′ Arg77	H2A′ Asp48	13.6	5.0
H2A′ Arg30	H2B′ Arg29	14.6	5.0
H3′ His39	H3′ Pro43	14.4	5.0
H3′ Arg40	H3′ Pro43	8.3	5.0
H3′ Arg49	H2A Gln112	18.4	5.0
H2A Arg77	H2A Asp48	12.5	5.0
H2A Arg30	H2B Arg29	9.7	5.0

Table S3: Distance restraints used to break contacts between the target histone residues and DNA. The force constant of each restraint was 418 kJ mol^{-1} nm^{-2}. The initial and target distances are specified in Ångstroms.

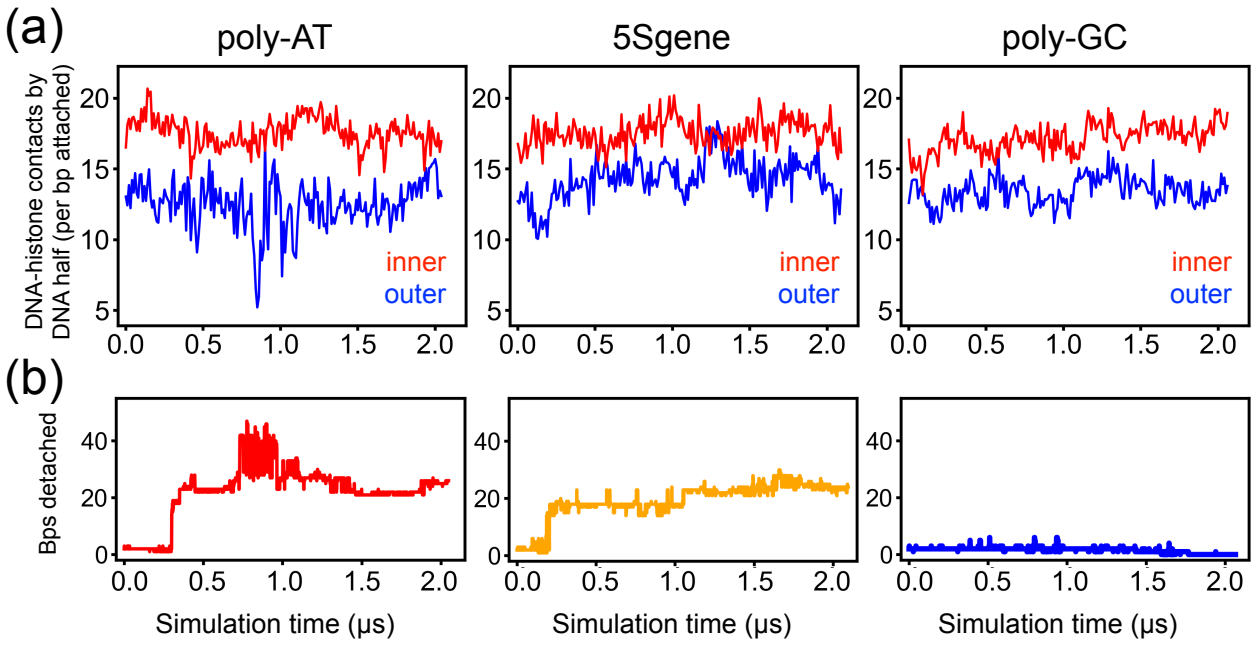
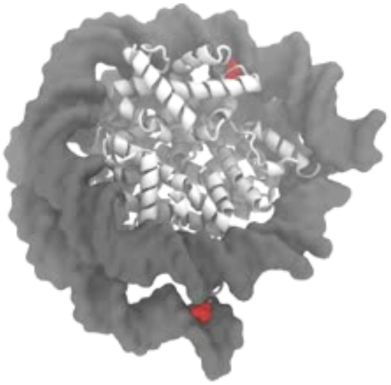


Figure S9 : DNA–histone contacts and DNA detachment. (A) The number of contacts that a histone core forms with the inner or the outer regions of the DNA per bp attached as a function of simulation time. (B) Base pairs detached as a function of simulation time. The plots in panel B match those provided in Fig. 2 of the main text. Each column refers to one nucleosome system in 1 M $MgCl_2$ solution, DNA sequences provided as labels along the top.

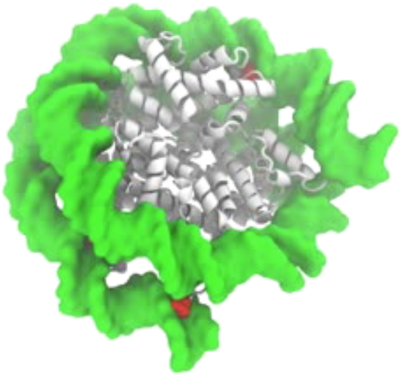
Time	Ensemble	Position restraints k	Bp restraints k	EN k within DNA	EN k between DNA and protein	EN k within protein
0 - 0.05	NVT	1000	4180	0	0	0
0.05 - 0.5	NVT	0	4180	418	418	418
0.5 - 1.0	NPT	0	4180	41.8	418	418
1.0 - 1.5	NPT	0	4180	0	418	418
1.5 - 3.0	NPT	0	4180	0	41.8	41.8
3.0 - 3.5	NPT	0	418	0	0	0

Table S4: Structural relaxation of all-atom models. The table specifies the duration (in ns), the thermodynamic ensemble, and the type and the force constant of the harmonic restraints for each step of the structural relaxation simulation. The force constant k is specified for individual restraints in units of kJ mol^{-1} nm^{-2}. EN denotes elastic network. Absolute positions of the restraints in the first step (spring constant in column 3 of the table), and the rest lengths defining the elastic networks (columns 5, 6 and 7), were determined from crystallographic coordinates (PDB ID: 3LZ0 [5]). Each base-pairing restraint (k in column 4) had an equilibrium distance $r_{\mathrm{b}} = 0.285$ nm, applied to the pairs of non-hydrogen atoms involved in the hydrogen bonds defining base pairs. EN's were composed of any pair of non-hydrogen atoms located within 0.5 nm of each other and not forming a direct chemical bond, excluding all atoms of the DNA bases.

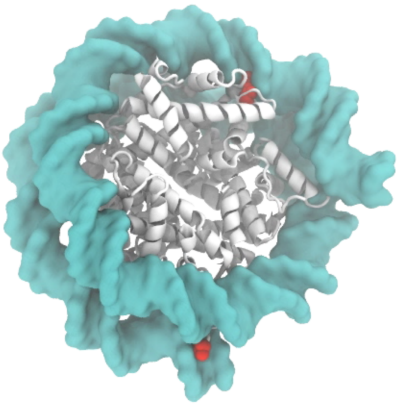
Captions to Supplementary Movies



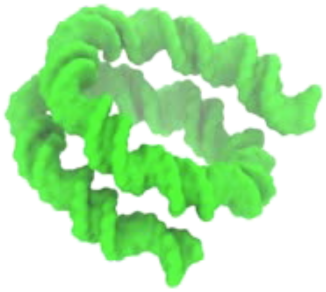
Movie 1 : Widom 601L nucleosome in 0.150 M NaCl. Animation illustrating the 2 μs MD trajectory of a nucleosome featuring the Widom 601L DNA sequence in 0.150 M NaCl. Colors highlight the histone core (white), surrounding DNA (dark gray), and histone residues H3 His39 and H2A Arg77 (red). This movie corresponds to the top panel of Fig. 1b.



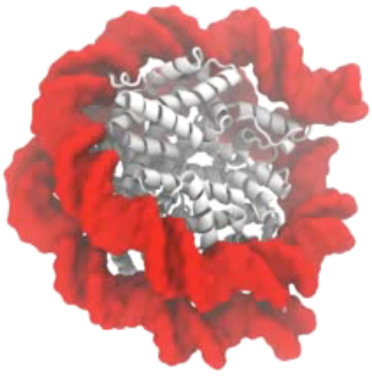
Movie 2 : Widom 601L nucleosome in 1.0 M $MgCl_2$. Animation illustrating the 2 μs MD trajectory of a nucleosome featuring the Widom 601L DNA sequence in 1.0 M $MgCl_2$. Colors highlight the histone core (white), surrounding DNA (green), and histone residues H3 His39 and H2A Arg77 (red). This movie corresponds to the middle panel of Fig. 1b.



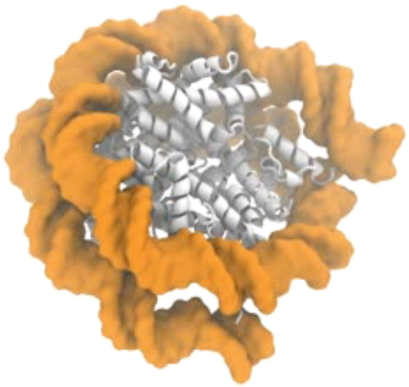
Movie 3 : Widom 601L nucleosome in 3.0 M NaCl. Animation illustrating the 2 μs MD trajectory of a nucleosome featuring the Widom 601L DNA sequence in 3.0 M NaCl. Colors highlight the histone core (white), surrounding DNA (cyan), and histone residues H3 His39 and H2A Arg77 (red). This movie corresponds to the bottom panel of Fig. 1b.



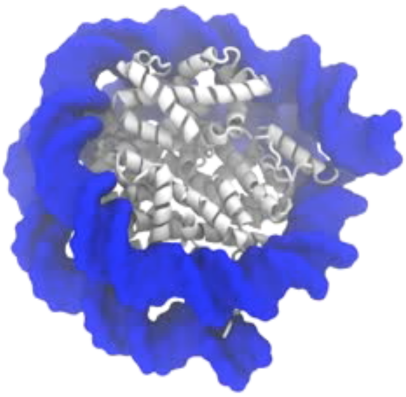
Movie 4 : Widom 601L DNA only in 1.0 M $MgCl_2$. Animation illustrating the 0.30 μs MD trajectory of the Widom 601L DNA sequence in 1.0 M $MgCl_2$, starting in a super-helically wound state, excluding the histone core. DNA shown in green.



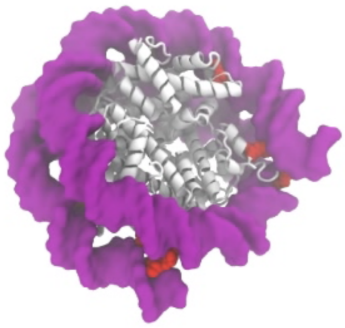
Movie 5 : poly-AT nucleosome in 1.0 M $MgCl_2$. Animation illustrating the 2 μs MD trajectory of a nucleosome featuring the poly-AT DNA sequence (*i.e.* ATAT...) in 1.0 M $MgCl_2$. Colors highlight the histone core (white) and surrounding DNA (red). This movie corresponds to the top panel of Fig. 2a.



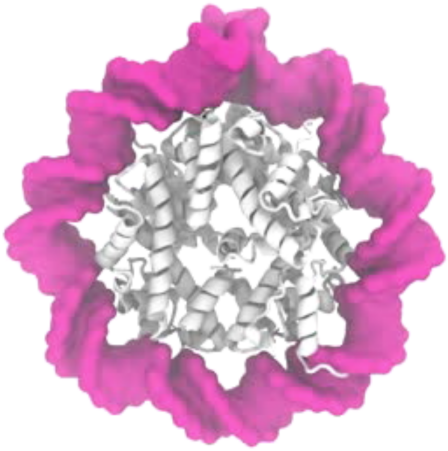
Movie 6 : 5S RNA gene nucleosome in 1.0 M $MgCl_2$. Animation illustrating the 2 μs MD trajectory of a nucleosome featuring the 5S RNA gene DNA sequence in 1.0 M $MgCl_2$. Colors highlight the histone core (white) and surrounding DNA (orange). This movie corresponds to the middle panel of Fig. 2a.



Movie 7 : poly-GC nucleosome in 1.0 M $MgCl_2$. Animation illustrating the 2 μs MD trajectory of a nucleosome featuring the poly-GC DNA sequence (*i.e.* GCGC...) in 1.0 M $MgCl_2$. Colors highlight the histone core (white) and surrounding DNA (blue). This movie corresponds to the bottom panel of Fig. 2a.



Movie 8 : Widom 601L nucleosome in 1.0 M $MgCl_2$, shifting histone residues away from DNA. Animation illustrating the 0.21 μs MD trajectory of a nucleosome featuring the Widom 601L DNA sequence in 1.0 M $MgCl_2$ in which periodically-placed residues were shifted away from DNA and held away. Colors highlight the histone core (white), surrounding DNA (dark purple), and the histone residues shifted away from nucleosomal DNA—H3 His39, Arg40, Arg49 and H2A Arg77, Arg30 (red). This movie corresponds to Figs. 4 and S4.



Movie 9 : The central 85 bp of Widom 601 in 1.0 M $MgCl_2$. Animation illustrating the 2 μs MD trajectory of a nucleosome featuring only the central 85 bp of the Widom 601 DNA sequence in 1.0 M $MgCl_2$. Colors highlight the histone core (white) and surrounding DNA (magenta). This movie corresponds to Fig. 5a.

References

[1] R. Salomon-Ferrer, D. A. Case, R. C. Walker, An overview of the Amber biomolecular simulation package, WIREs Comput Mol Sci 3 (2013) 198–210.

[2] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, et al., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1 (2015) 19–25.

[3] D. E. Shaw, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, et al., Anton, a Special-purpose Machine for Molecular Dynamics Simulation, Comm. ACM 51 (2008) 91.

[4] J. S. Mitchell, J. Glowacki, A. E. Grandchamp, R. S. Manning, J. H. Maddocks, Sequence-dependent persistence lengths of DNA, J. Chem. Theory Comput. 13 (2017) 1539–1555.

[5] D. Vasudevan, E. Y. Chua, C. A. Davey, Crystal structures of nucleosome core particles containing the '601' strong positioning sequence, J. Mol. Biol. 403 (2010) 1–10.