Electronic Supplementary Material (ESI) for Nanoscale.

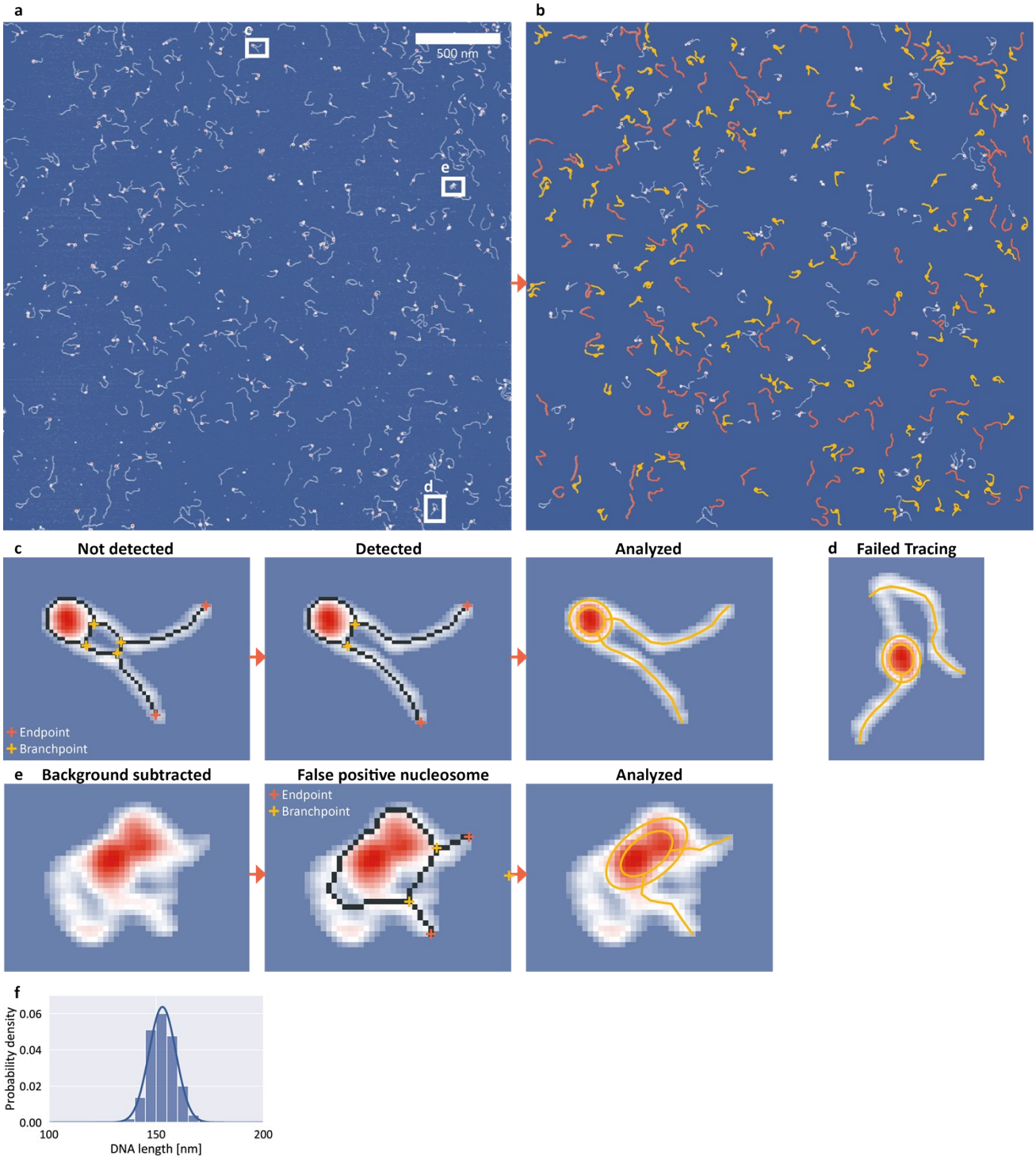
SUPPLEMENTARY INFORMATION for

High-throughput AFM analysis reveals unwrapping pathways of H3 and CENP-A nucleosomes

Sebastian F. Konrad[1], Willem Vanderlinden[1,*], Wout Frederickx[2], Tine Brouns[2], Björn Menze[3], Steven De Feyter[2], and Jan Lipfert[1,*]

[1]*Department of Physics and Center for Nanoscience, LMU Munich, Amalienstr. 54, 80799 Munich, Germany*
[2]*Department of Chemistry, KU Leuven, Celestijnenlaan 200F, 3001 Heverlee, Belgium*
[3]*Department of Informatics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany*

[*]*For correspondence: willem.vanderlinden@lmu.de and jan.lipfert@lmu.de*



Supplementary Figure 1 | Detection efficiency, manual molecule classification, and DNA length.

a, AFM topographic image of bare DNA and nucleosomes to assess the detection rate of the automated tracing. The total field of view is 3 µm x 3 µm and was recorded with 1.46 nm/pixel.

b, Analysis of the topographic image from panel **a**. 112 bare DNA strands (orange) and 131 nucleosomes (yellow) were detected and analyzed automatically with manual detection help as described in **c**. To quantify the detection rate and to assess imaged molecules that were not classified as either DNA or nucleosome in the first fully automated step, the image was inspected

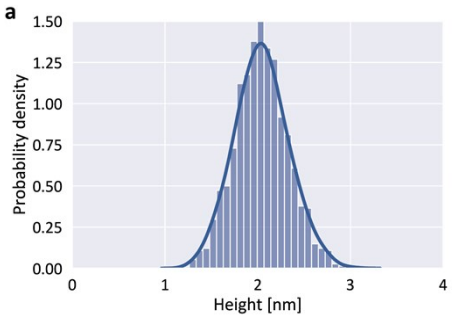
visually. The fully automated classification and tracing routine detected 95% of all manually analyzable DNA and nucleosomes in the field of view that did not show overlaps with other molecules.

c, Example how manual intervention can improve classification of molecules not assigned to bare DNA or nucleosome molecules in the first automatic classification step. In the case depicted here, too many branchpoints (four instead of two) are detected and thus the nucleosome was not automatically classified. After manually removing the overlapping pixels of the DNA arms, the nucleosome is properly classified and the automated analysis framework traces the structure parameters. Overall, such manual classification help for unclassified molecules enabled tracing of up to 98% of manually analyzable molecules.

d, Example of a false negative, i.e. a molecule that by visual inspection appears to constitute a valid nucleosome or DNA, but is not traced properly. The tracing fails due to the strong bending of the long arm. Overall, 2% of molecules that we identified as valid by visual inspection fall into this category.

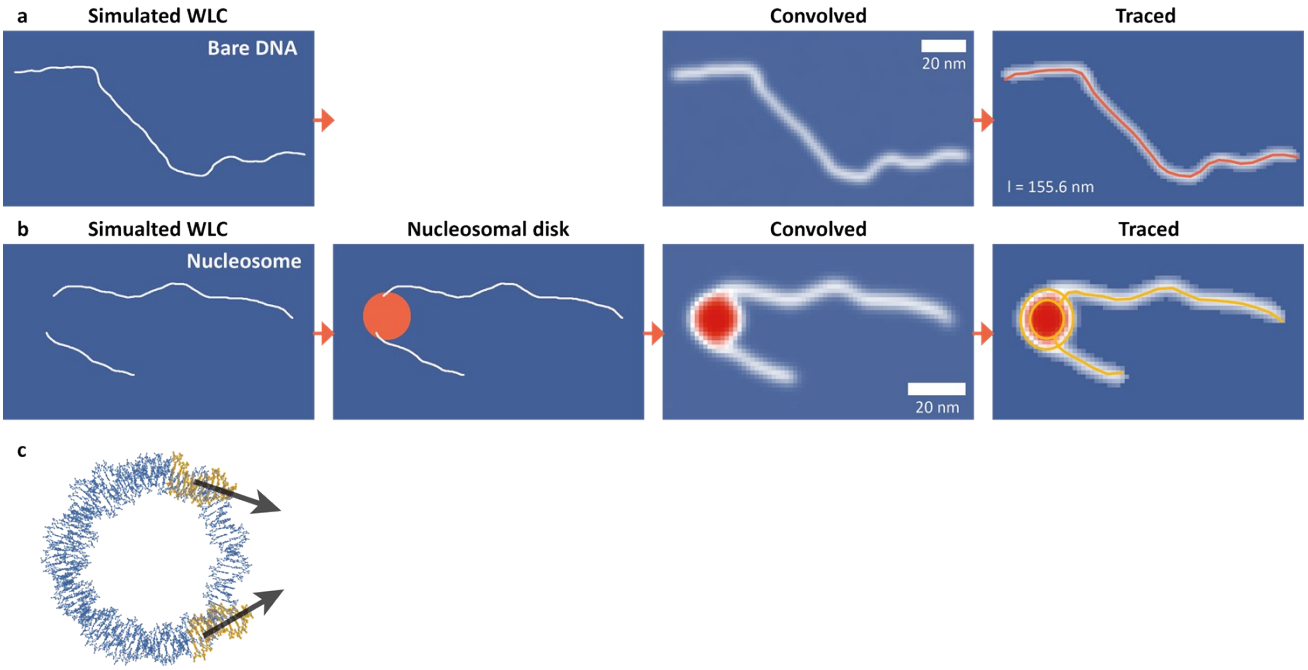
e, Example of a false positive, i.e. a molecule that is classified as a valid nucleosome or DNA by our algorithm, but excluded by visual inspection. In the image shown in **a**, a total of 9 false positive molecules were traced by the automated toolbox. For this work, false positives were removed for further analysis. The example molecule contains two branchpoints and two endpoints and is thus classified as a nucleosome.

f, Histogram of bare DNA lengths of all data sets measured in 200 mM NaCl presented in this work. We find a contour length of l_c = 152.9 ± 6.3 nm (mean ± std from 5651 molecules).



Supplementary Figure 2 | Nucleosome heights.

a, Raw nucleosome heights of the H3 nucleosome data set shown in Figure 2 (*N* = 1011).

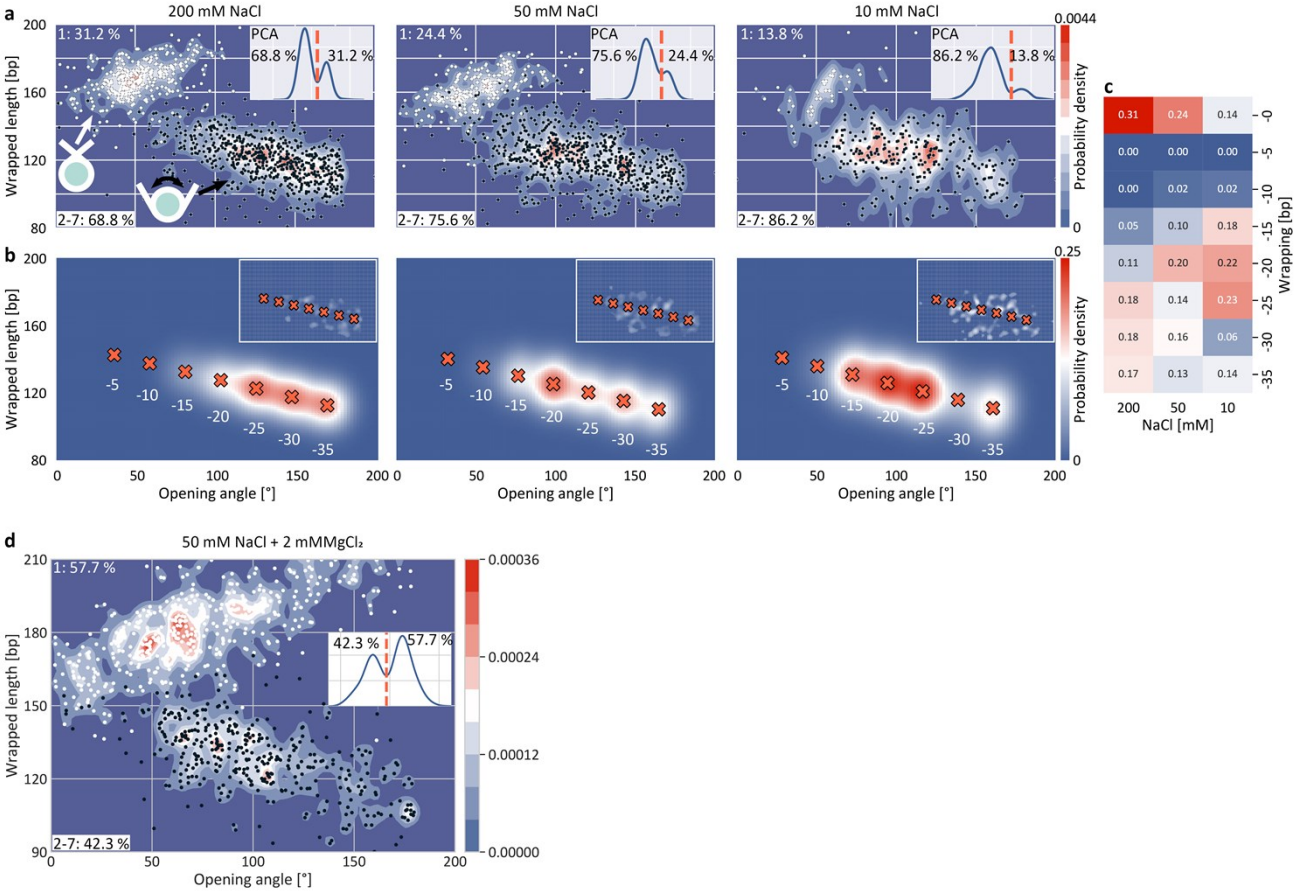


Supplementary Figure 3 | DNA and nucleosome simulations to quantify the effect of AFM tip convolution.

a, 486 bp DNA were simulated using the worm-like chain polymer model with a persistence length of 40 nm and a rise per base pair of 0.33 nm/bp. The simulated DNA backbone was dilated to a width of 2 nm and convolved with a Gaussian filter after applying random noise to mimic the effect of AFM tip convolution. The simulated DNA strands were then traced with our automated analysis pipeline using the same settings as used for experimental data. See Methods for details.

b, Simulation of nucleosomes consisted of generating two DNA arms (106 bp and 233 bp for the fully wrapped nucleosome) based on the worm-like chain model that protrude from the nucleosomal disc. The orientation of the protruding DNA arms was deduced from the crystal structure (panel **c**), for details see the section "AFM image simulations" in Methods. The nucleosome depicted here is unwrapped by 35 bp from the long arm side.

c, The crystal structure of the nucleosome core particle yields the orientation of the DNA arms and the nucleosomal opening angle of 66.5° for fully wrapped nucleosomes. Rendered from PDB 1KX5. For partially unwrapped nucleosomes the DNA orientation and opening angle are adjusted accordingly (Methods).



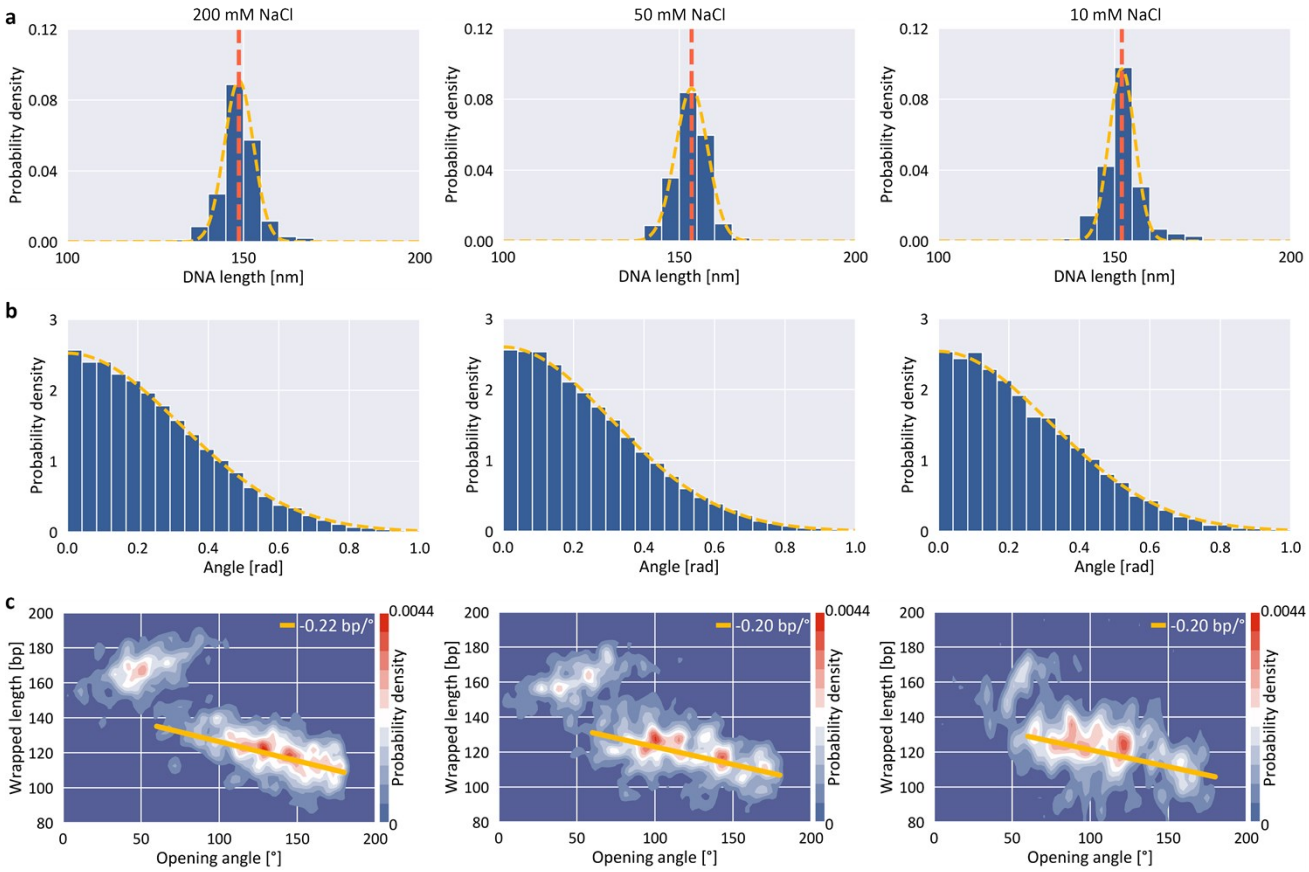
Supplementary Figure 4 | Effect of salt concentration on H3 nucleosome wrapping.

a, 2D kernel density estimate profiles of wrapped length versus opening angle. Black and white dots indicate individual nucleosomes. The PCA of wrapped length and volume yields the basis for separating the two major populations as shown in the inset. The data sets include *N* = 1011 nucleosomes for 200 mM NaCl, *N* = 934 at 50 mM NaCl, and *N* = 325 at 10 mM NaCl.

b, 2D Gaussian fits to the partially unwrapped nucleosomes. The Gaussian amplitudes yield the populations of the individual states of unwrapping and the insets show the fit residuals. 2D Gaussian fits to the partially unwrapped nucleosomes at 50 mM NaCl and 10 mM NaCl show the trend towards more compact wrapping for lower salt concentrations. All nucleosomes presented in this plot are from the same nucleosome reconstitution and were imaged with the same cantilever.

c, Heat map of the populations of the individual wrapping states for NaCl concentrations of 200 mM, 50 mM and 10 mM.

d, 2D kernel density estimate profile of wrapped length versus opening angle measured at a salt concentration of 50 mM NaCl and 2 mM $MgCl_2$. The data set contains *N* = 870 nucleosomes.



Supplementary Figure 5 | Effect of salt concentration on DNA contour and persistence length.

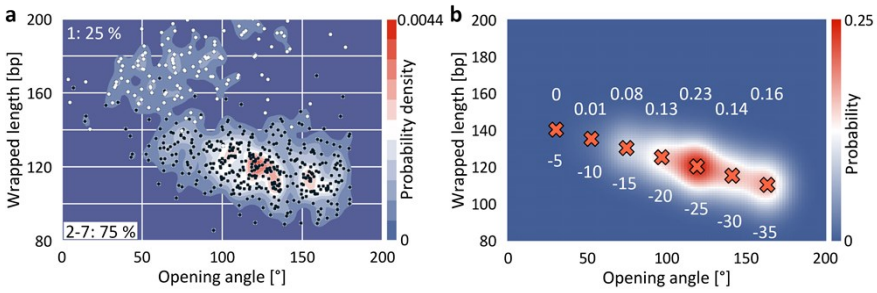
a, Bare DNA lengths of the data presented in Figure 4 for different salt concentrations. The Gaussian fits yield contour lengths of 148.7 ± 4.1 nm, 153.4 ± 4.6 nm and 152 ± 3.7 nm for NaCl concentrations of 200 mM, 50 mM and 10 mM respectively (*N* = 888, 797, and 294).

b, The DNA length is determined by tracing its contour with segments of 5 nm length. The distribution of angles α between consecutive 5 nm segments can be used to determine the persistence length by fitting

$$P(\alpha) \sim A.e^{\left(-\frac{\alpha^2}{2.z^2}\right)}$$

with $z = l_{segment} / l_{persistence}$. The fits yield persistence lengths of 49.5 nm, 53.0 nm and 50.3 nm for NaCl concentrations of 200 mM, 50 mM and 10 mM respectively (*N* = 24077, 22338, and 9995).

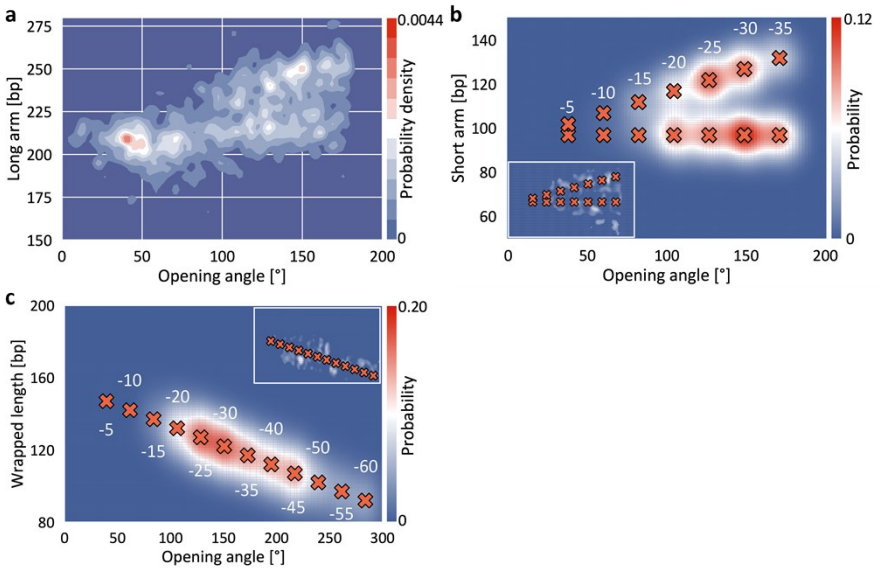
c, The slopes of wrapped length vs. opening angle for partially unwrapped nucleosomes are -0.22 bp/°, -0.20 bp/° and -0.20 bp/° for NaCl concentrations of 200 mM, 50 mM and 10 mM respectively. The slopes agree well with the slope predicted from simulated data (Figure 3b, -0.23 bp/°) indicating that nucleosomes attach to the surface in a flat geometry with the DNA gyres parallel to the surface for a broad range of ionic conditions.



Supplementary Figure 6 | Surface functionalization control.

a, 2D kernel density profile of nucleosome opening angle and wrapped length at 200 mM NaCl and 10 mM Tris (N = 558). The surface was functionalized using 0.001% w/v poly-L-lysine in contrast to the 0.01% w/v poly-L-lysine concentration used for the data presented in this work.

b, 2D Gaussian fits to the partially unwrapped nucleosomes. Numbers indicate the unwrapping in base pairs and the population sizes. The wrapping states for H3 nucleosomes measured at 200 mM NaCl agree for both poly-L-lysine concentrations (see H3 nucleosome population sizes measured at 0.01 % poly-L-lysine in Figure 5c).

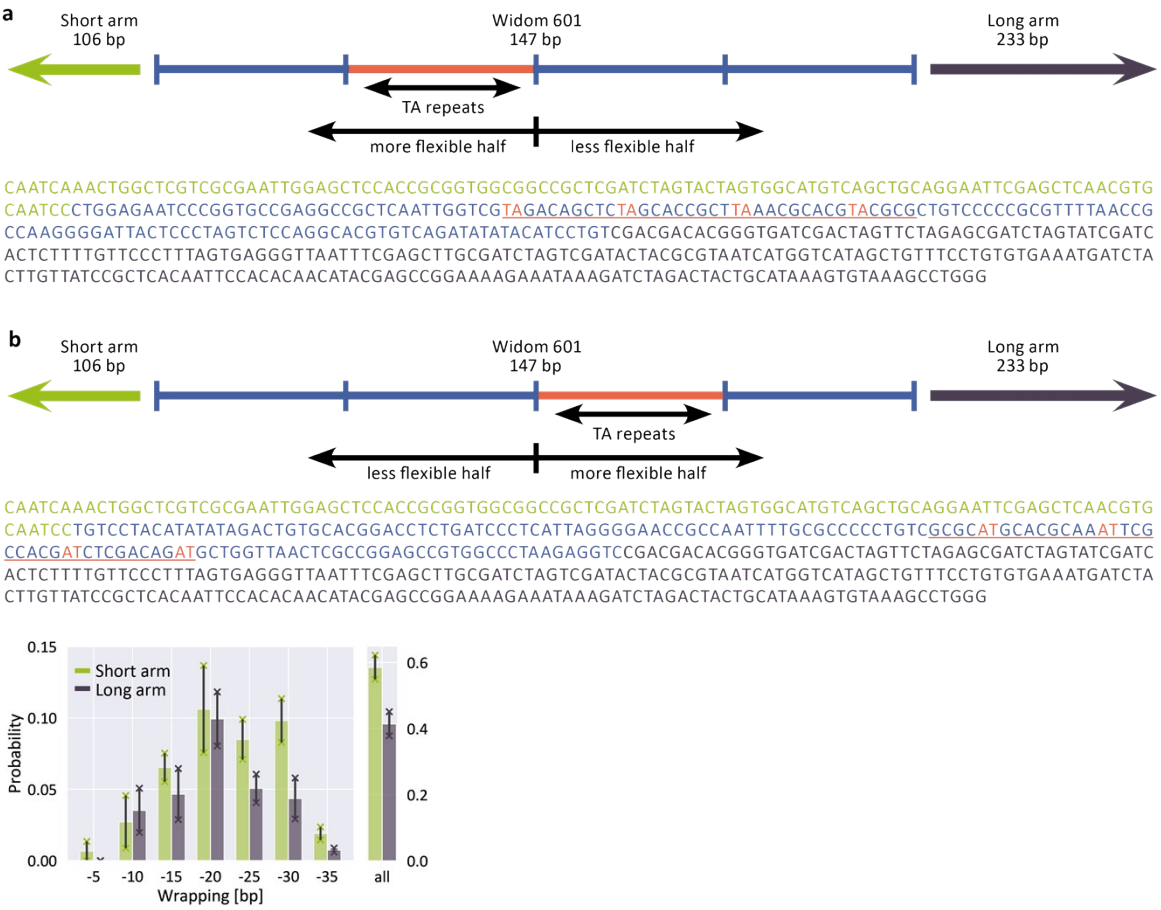


Supplementary Figure 7 | Determination of the occupancies of different wrapping states.

a, 2D kernel density profile (bandwidth = 2.5°, 2.5 bp) of long arm length and opening angle for the H3 nucleosome data set at 200 mM NaCl presented in Figure 4. The data for long arm length vs. angle are overall noisier than the data for the short arm length (Figure 4a) as expected from the broadening of the contour length distribution for a longer length.

b, Fit of 14 2D Gaussians to the density distribution of short arm length and opening angle for the partially unwrapped H3 nucleosomes (Figure 4a) measured at 200 mM NaCl (N = 696). The inset represents the difference between the measured and the fitted 2D density profile, i.e. the residuals of the fit. The occupancies determined from the fits are shown in Figure 4c.

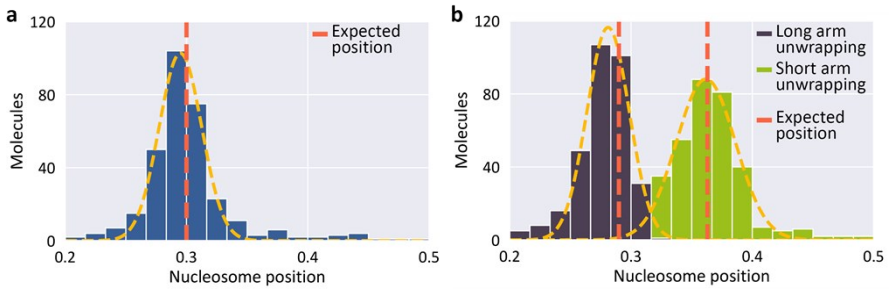
c, Fit of 12 2D Gaussians to the density distribution of wrapped length and opening angle for the partially unwrapped CENP-A nucleosomes (Figure 5b) measured at 200 mM NaCl (N = 1019). The inset represents the difference between the measured and the fitted 2D density profile, i.e. the residuals of the fit. The occupancies determined from the fits are shown in Figure 5c.



Supplementary Figure 8 | Non-palindromic nature of the W601 positioning sequence.

a, The 486 bp DNA construct contains the W601 nucleosome positioning sequence (147 bp) flanked by a short arm (106 bp) and a long arm (233 bp). Four TA repeats with a 10 bp periodicity on the short arm half of the W601 sequence induce the higher flexibility of the half of the W601 sequence on the short arm side[17].

b, As a control measurement the W601 sequence was flipped such that the more flexible half of the W601 sequence lies on the opposite side –the long arm side– compared to the data presented in Figure 4. Analysis of the unwrapping pathways of the partially unwrapped nucleosomes (69% of all nucleosomes) yields a preference for short arm opening over long arm opening with probabilities of (58.6 ± 3.7) % and (41.4 ± 3.7) % respectively (mean ± SEM from two data sets, *N* = 598 and 559).



Supplementary Figure 9 | W601 nucleosome positioning.

a, Positioning of the fully wrapped nucleosomes along the DNA for the H3 nucleosome dataset presented in Figure 4a. Yellow dashed line is a Gaussian fit (centered at 0.29 ± 0.02). The nucleosome position is calculated by dividing the short arm length by the sum of short arm and long arm length. From the DNA construct, the arms of a fully wrapped nucleosome are expected to be 106 bp and 233 bp (Figure 1a). However, since for fully wrapped nucleosomes the exiting DNA arms overlap, the length of the arms is underestimated by 10 bp each as described in Figs. 2 and 3. Thus, the expected nucleosome position of fully wrapped nucleosomes is (106 bp – 10bp) / (106 bp – 10 bp + 233 bp – 10 bp) = 0.30.

b, Positioning of the partially unwrapped nucleosomes along the DNA for the dataset presented in Figure 4a. The anti-cooperative unwrapping of H3 nucleosomes leads to different expected positions for nucleosomes that partially unwrap *via* the short arm or *via* the long arm. The distribution of partially unwrapped nucleosomes peaks at 120 bp of wrapping (Figure 2a) implying that one of the two arms has to unwrap 27 bp compared to canonical nucleosomes with 147 bp wrapped. Thus, the expected position for nucleosomes unwrapping *via* the long arm or *via* the short arm is (106 bp) / (106 bp + 233 bp + 27 bp) = 0.29 and (106 bp + 27 bp) / (106 bp + 27 bp + 233 bp) = 0.36 respectively. The fitted Gaussians center at 0.28 ± 0.02 and 0.36 ± 0.02 for the distributions unwrapping *via* the long arm and *via* the short arm respectively.