Biophysical Journal, Volume *121*

Supplemental information

Quantifying epigenetic modulation of nucleosome breathing by high-throughput AFM imaging

Sebastian F. Konrad, Willem Vanderlinden, and Jan Lipfert

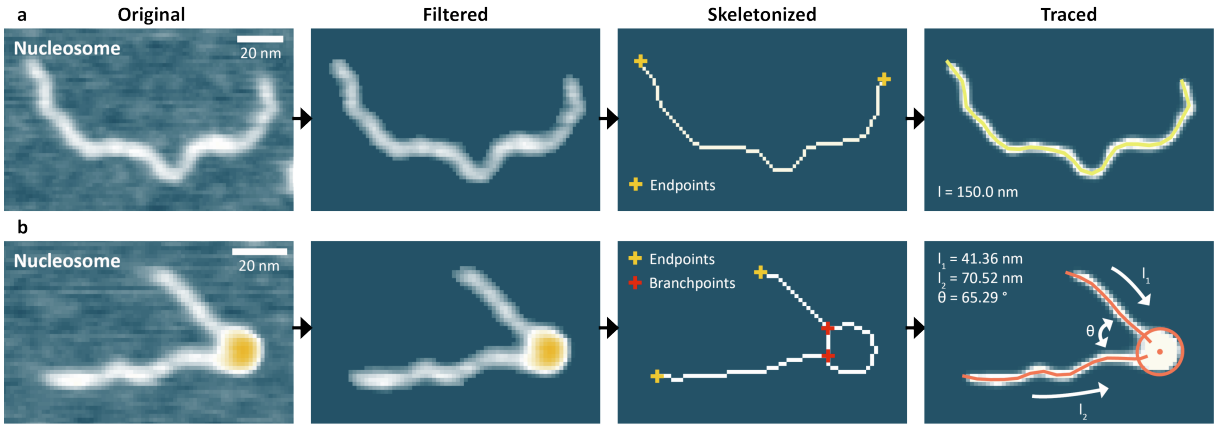
SUPPLEMENTARY INFORMATION FOR

Quantifying epigenetic modulation of nucleosome breathing by high-throughput AFM imaging

Sebastian F. Konrad[1], Willem Vanderlinden[1], Jan Lipfert[1,*]

[1]*Department of Physics and Center for Nanoscience, LMU Munich, Amalienstr. 54, 80799 Munich, Germany*

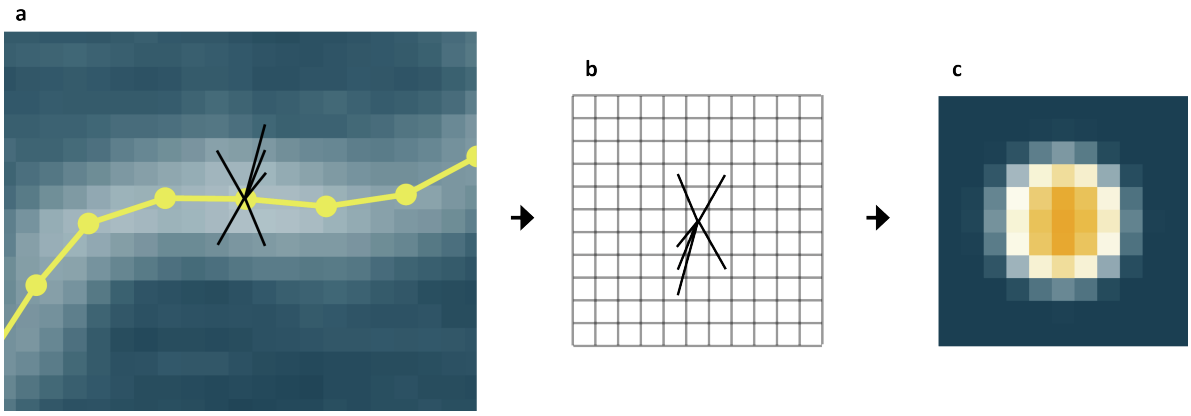
[*]Correspondence: Jan.Lipfert@lmu.de



Supplementary Figure 1 | Tracing of bare DNA and nucleosomes.

a, AFM topographic image of a bare DNA strand with the different tracing steps visualized. First, the original image is filtered by applying a Gaussian filter and removing the background with a fixed threshold value. Subsequently, the filtered image is skeletonized. The skeletonized backbone of the molecules serves as the basis for classification: whereas the skeleton of bare DNA has exactly two endpoints and no branchpoints – points that have more than two neighbors – the skeleton of nucleosomes contains exactly two endpoints and two branchpoints. Finally, the length of the bare DNA molecule is traced after applying the deconvolution.

b, AFM topographic image of a nucleosome with the different tracing steps involved. After initial filtering, the molecule is skeletonized for classification. Finally, the nucleosome is traced with regards to the arm lengths, the opening angle and the volume.



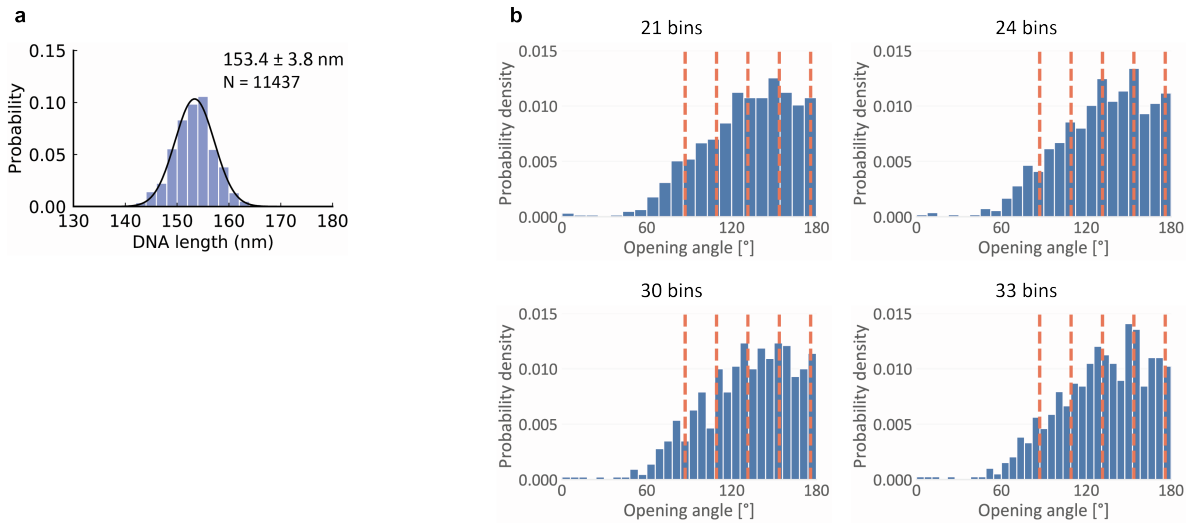
Supplementary Figure 2 | Tip shape estimation for image deconvolution.

The tip shape estimation is based on analysis of the bare DNA images present in each AFM field of view. DNA molecules are identified in the images as described in ("***AFM image analysis***").

a, In the first step, DNA strands are traced without deconvolution (The trace and trace points are shown schematically in yellow). Note that the spacing of trace points used is much smaller than what is shown for clarity: The DNA trace is approximated by a spline interpolation to provide more trace points (spacing of <0.2 nm) to reduce error in estimating the tip shape due to a too coarse grained trace. For every image pixel within 8 nm from a DNA trace point, the vector between the pixel center and the nearest trace point is computed (the vectors are shown as black lines).

b, The height values for the pixels in panel **a** are summed up on an initially empty grid, whereby the height value with vector (x, y) to the nearest trace point contributes to position (-x, -y) on the grid. Repeating the process for all DNA strands in the image (typically 100-200 DNA strands per image) results in tens of thousands height values to fill the 2D grid with.

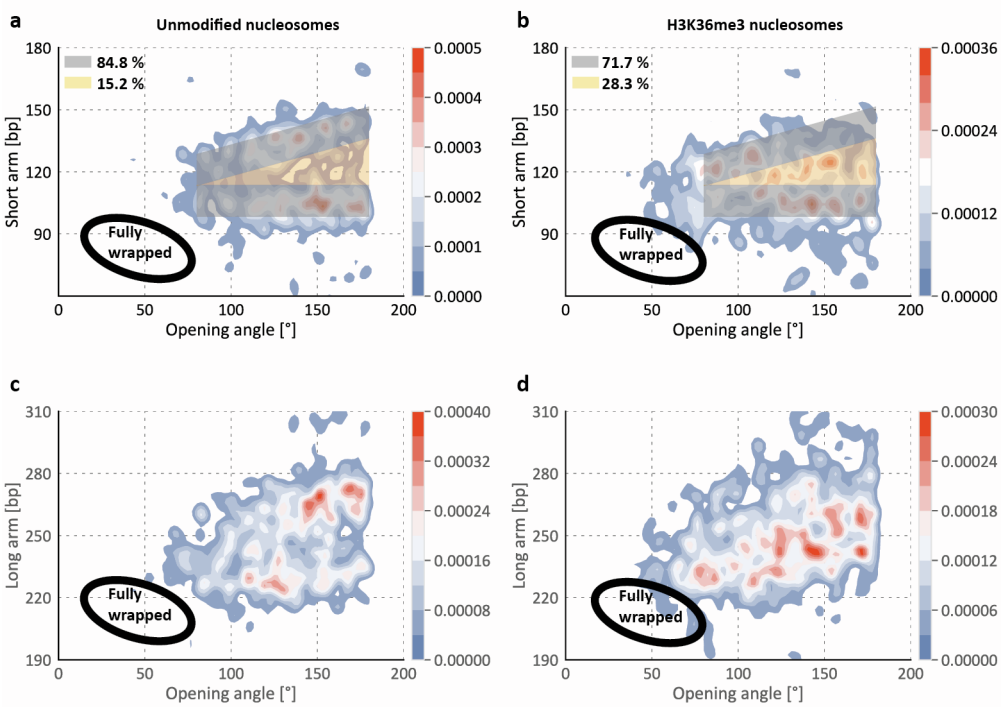
c, After adding up all pixel height values at the positions specified by the vectors to the nearest DNA trace point on the grid, the grid pixels are normalized based on the number of values added per grid pixel. The resulting distribution approximates the tip shape that was used in imaging the respective image and is used as the point spread function for image deconvolution. Each pixel is 1.46 nm x 1.46 nm.



Supplementary Figure 3 | Bare DNA lengths and opening angle distributions with different bin sizes

a, Histogram of bare DNA lengths combined for all data sets used in this work. We find a contour length of l_c = 153.4 ± 3.9 nm (mean ± std from 11437 molecules) corresponding to a length per bp of 0.316 ± 0.008 nm, in agreement with previous measurements by AFM[1,2], and solution X-ray scattering[3].

b, Histograms with different bin sizes for the opening angle distribution from Fig. 1g. The histogram in Fig. 1g contains a total of 27 equally spaced bins over 180°, i.e. 6.67° per bin. This bin size was chosen as a compromise between the detail that can be seen in the distribution and the noise that comes with the smaller sample sizes per bin when increasing the total number of bins. For the different bin sizes, the 5 bp periodicity of nucleosome unwrapping can still be seen. However, for larger bins the distinct peaks get washed out and for smaller bins the sample sizes are too small at angles <100° such that noise dominates the distribution.



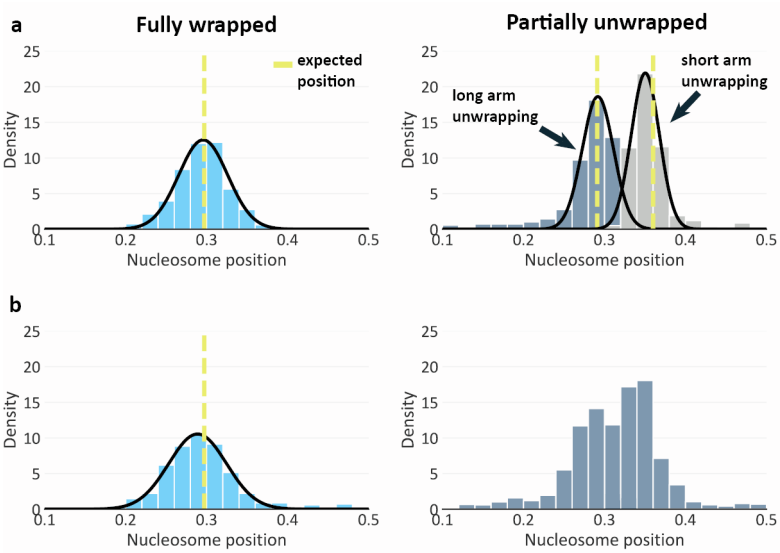
Supplementary Figure 4 | Quantification of anti-cooperative unwrapping.

a, 2D kernel density profile (bandwidth = 2.5°, 2.5 bp) of short arm length and opening angle for partially unwrapped canonical nucleosomes (N = 1035). Nucleosomes that unwrap anti-cooperatively, *i.e.* with either the long arm or the short arm unwrapping, are expected in the dark area: Nucleosomes that unwrap from the long arm side only are expected to have a short arm length of ~106 bp (based on the DNA construct used). This population corresponds to the lower dark area that is centered at a 106 bp short arm length. In contrast, for nucleosomes unwrapping from the short arm side only, the length of the short arm should increase with a slope of ~0.24 bp/° according to a wrapping of ~147 bp in ~1.7 turns around the histone octamer. The population unwrapping from the short arm side, therefore, corresponds to the upper dark area. The width of the dark areas on the y-axis was chosen to be 13 bp, which is based on the STD that we find for the length distribution of bare DNA (see Supplementary Figure 3). Nucleosomes that unwrap from both sides simultaneously are expected in the yellow area. For this particular data set of canonical nucleosomes, 84.8 % of the nucleosomes are in the regime of anti-cooperative unwrapping and 15.2 % in the regime of stochastic unwrapping. The values obtained in this analysis serve as basis for the quantification of anti-cooperativity in Fig. 4c of the main text. From the geometric construction of the areas, the dark area makes up 75 % of the total area and the yellow area makes up 25 % of the total area. The black ellipse indicates the position of fully wrapped nucleosomes that are omitted for clarity.

b, 2D kernel density profile (bandwidth = 2.5°, 2.5 bp) of short arm length and opening angle for partially unwrapped H3K36me3 nucleosomes (N = 1155). 71.7 % of the nucleosomes are in the regime of anti-cooperative unwrapping and 28.3 % in the regime of stochastic unwrapping.

c, 2D kernel density profile (bandwidth = 2.5°, 2.5 bp) of long arm length and opening angle for partially unwrapped wild type nucleosomes (same data set as in panel **a**). Due to the increased length of the long arm compared to the short arm (233 bp vs. 106 bp from the construct), the distribution is noisier compared to the short arm vs. opening angle distribution. Yet, the anti-cooperativity of unwrapping can still be clearly observed.

d, 2D kernel density profile (bandwidth = 2.5°, 2.5 bp) of long arm length and opening angle for partially unwrapped H3K36me3 nucleosomes (same data set as in panel **b**).

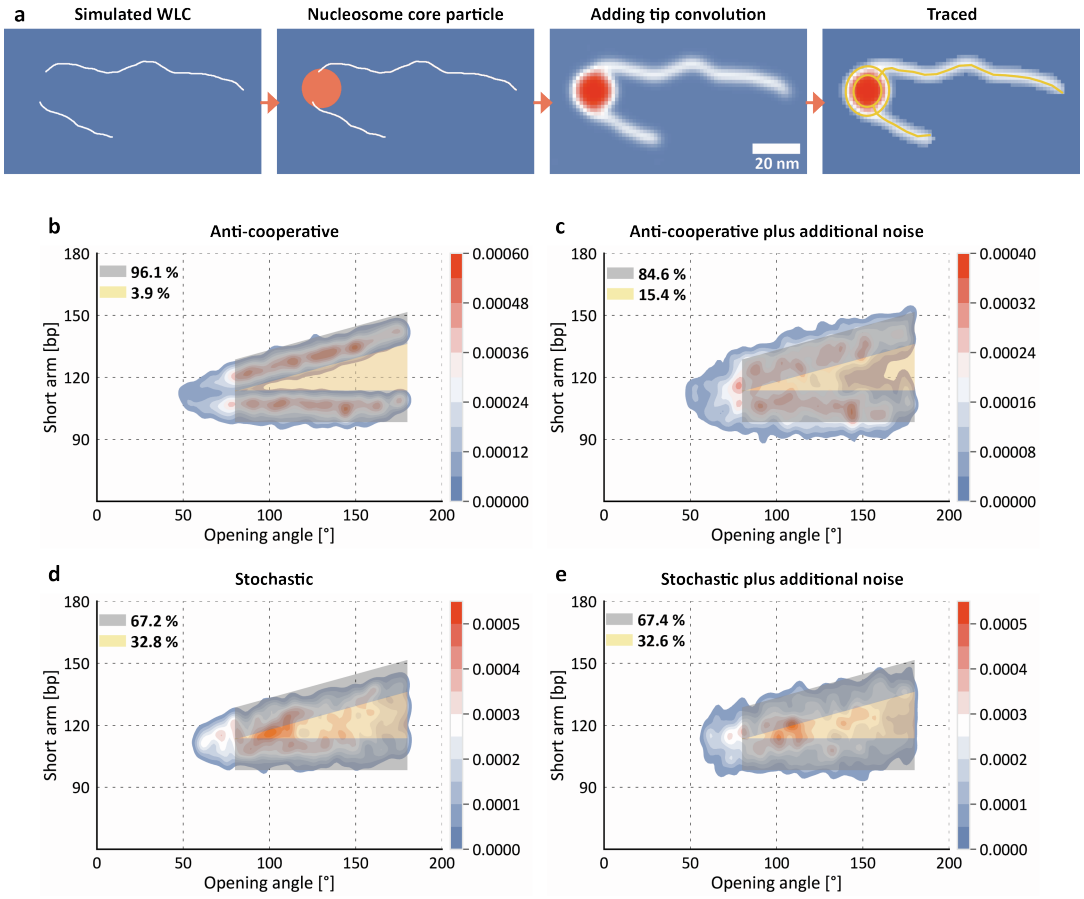


Supplementary Figure 5 | Nucleosome positioning.

a, Nucleosome positioning for a sample data set of unmodified nucleosomes (N = 1300). The nucleosome position is calculated by dividing the short arm length by the sum of short arm and long arm length. For fully wrapped nucleosomes (left panel), from the DNA construct, the arms are expected to be 106 bp and 233 bp (Fig. 1b). 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 Fig. 2 and in the main text. Thus, the expected nucleosome position of fully wrapped nucleosomes is (106 bp – 10bp) / (106 bp – 10 bp + 233 bp – 10 bp) = 0.30.

For partially unwrapped nucleosomes (right panel), the expected values are based on our finding that the distribution of partially unwrapped nucleosomes peaks at ~25 bp unwrapped (for wild type nucleosomes) and that it happens anti-cooperatively, *i.e.* from one side only. Therefore, there are two expected positions that can be calculated from short arm unwrapping and long arm unwrapping: the expected position for nucleosomes unwrapping *via* the long arm or *via* the short arm is (106 bp) / (106 bp + 233 bp + 25 bp) = 0.29 and (106 bp + 25 bp) / (106 bp + 25 bp + 233 bp) = 0.36 respectively.

b, Nucleosome positioning for a sample data set of H3K36me3 nucleosomes (N = 1732). Similarly to the unmodified nucleosomes, the H3K36me3 nucleosomes are positioned well. Due to the stochastic unwrapping of H3K36me3 nucleosomes it is not possible to separate the partially unwrapped nucleosomes (right histogram) into nucleosomes that solely unwrap from the short arm and those that unwrap solely from the long arm.



Supplementary Figure 6 | Simulation of anti-cooperative and stochastic unwrapping.

a, Simulation of nucleosomes consisted of placing a nucleosomal disk and simulating protruding DNA arms. The position and initial directionality of the protruding DNA arms was deduced from the nucleosome crystal structure (PDB 1KX5). The lengths of the DNA arms are 106 bp for the short arm and 233 bp for the long arm initially and are varied based on the state of unwrapping that is simulated (see Methods). Consecutively, the DNA was dilated to its expected width of 2 nm and a Gaussian filter was applied to mimic the effect of tip convolution. Finally, the synthetic AFM image is traced with our automated image analysis pipeline.

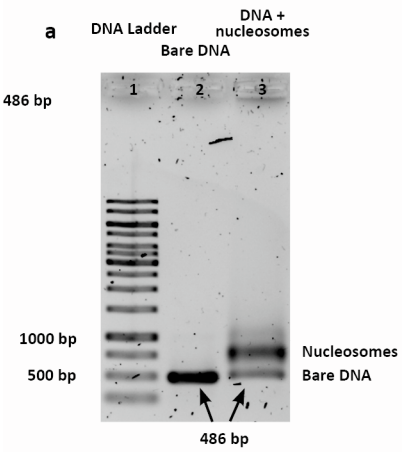
b, 2D Kernel density plot for simulated nucleosomes (N = 2072). To simulate anti-cooperative unwrapping, the length of either the short or the long arm was always kept constant (106 bp or 233 bp for the short and the long arm respectively) and the length of the other arm was increased in 5 bp steps up to a maximum unwrapping of 35 bp. Unwrapping is simulated to occur from each arm in 50 % of the cases and the sizes of the individual unwrapping populations are based on the probability for each population as experimentally measured for unmodified nucleosomes (Figure 3a).

c, 2D Kernel density plot for simulated nucleosomes (N = 2072). The plot comprises the same nucleosomes as shown in b. However, additional Gaussian distributed noise with σ = 5 bp was added

to the short arm length to better capture expected imaging errors that might occur during experiments.

d, 2D Kernel density plot for simulated nucleosomes (N = 1469). Compared to the anti-cooperative unwrapping of plots b and c, this plot shows nucleosomes that unwrap stochastically. To simulate stochastic unwrapping, the length of both arms was randomly increased in 5 bp steps for the individual unwrapping steps. For example, for simulation of states of partial unwrapping of 10 bp in total, the length of the short and long arm was randomly either increased by [+0, +10], [+5, +5] or [+10, +0] respectively. This procedure was again repeated up to a maximum unwrapping of 35 bp in total.

e, 2D Kernel density plot for simulated nucleosomes (N = 1469). The plot comprises the same nucleosomes as shown in d. Similar to c, additional Gaussian noise with σ = 5 bp was added to the short arm length to better capture imaging errors that might occur during experiments.



Supplementary Figure 7 | Electrophoretic mobility shift assay confirms nucleosome assembly.

Agarose gel stained with ethidium bromide. Gel electrophoresis was performed for a DNA size ladder (lane 1; "GeneRuler 1 kb", ThermoFisher Scientific), the bare 486 bp DNA used for reconstitution (lane 2), and the nucleosome sample (lane 3) that was obtained from nucleosome reconstitution (see Methods). The nucleosome sample contains both bare DNA (486bp) and unmodified nucleosomes (the second band that travelled more slowly).

Supporting References

1 T. Brouns, H. De Keersmaecker, S. F. Konrad, N. Kodera, T. Ando, J. Lipfert, S. De Feyter and W. Vanderlinden, *ACS Nano*, 2018, **12**, 11907–11916.

2 C. Rivetti, M. Guthold and C. Bustamante, *J. Mol. Biol.*, 1996, **264**, 919–932.

3 T. Zettl, R. S. Mathew, S. Seifert, S. Doniach, P. A. B. Harbury and J. Lipfert, *Nano Lett.*, 2016, **16**, 5353–5357.