Histone tails cooperate to control the breathing of genomic nucleosomes

Jan Huertas,[†,‡,¶] Hans R Schöler,[‡,§] and Vlad Cojocaru[*,†,‡,¶]

†*In Silico Biomolecular Structure and Dynamics Group, Hubrecht Institute, Utrecht, The Netherlands*

‡*Department of Cellular and Developmental Biology, Max Planck Institute for Molecular Biomedicine, Münster, Germany*

¶*Center for Multiscale Theory and Computation, Westfälische Wilhelms University, Münster, Germany*

§*Medical Faculty, University of Münster, Münster, Germany*

E-mail: v.cojocaru@hubrecht.eu

Abstract

Genomic DNA is packaged in chromatin, a dynamic polymer that forms fibers variable in size and compaction. Chromatin's monomeric unit, the nucleosome, is formed by wrapping $\sim$ 146 basepairs of DNA around histone proteins. Dynamic transitions in the chromatin structure, which depend on intra- and inter-nucleosome dynamics and on chemical modifications of histones, are key to initiating gene regulation programs. The histone modifications occur in unstructured terminal regions known as tails and are key to epigenetic regulation of gene expression. Here we elucidate how the interplay between the tails of two histones, H3 and H2A modulate ample breathing motions of genomic nucleosomes. From 18 µs of atomistic molecular dynamics simulations of two genomic and one engineered nucleosomes, we observed two rare, large opening events,

one in each genomic nucleosome, and only small amplitude breathing in the engineered nucleosome. Clustering of the sampled structures revealed a strong correlation between DNA conformations with different degree of opening and histone tail conformations and positions. The ample nucleosome motions were caused by changes in the interaction profiles between the DNA and clusters of lysine and arginine residues in these tails, many of which are targets for epigenetic regulatory modifications. Given that histone tail modulated nucleosome breathing can change the structure and compaction of chromatin fibers, the mechanism we describe provides the basis for understanding chromatin dynamics at atomic resolution.

Introduction

In eukaryotic cells, long pieces of DNA are confined in a small nucleus by packing into a dynamic structure known as chromatin. The basic unit of chromatin, the nucleosome, is an octamer of proteins with $\sim$ 147 basepairs of DNA wrapped around.[1,2] Four histone proteins(H3, H4, H2A and H2B), each of them present twice, form this octamer. At the N-terminus of all the histones there is a disordered, positively charged region, known as the histone tail. Additionally, the H2A histones contain another tail at the C-terminus, with similar characteristics (from now on, the H2AC tail). Combined, the histone tails account for $\sim$28% of the mass of the core histone proteins.[1]

Histone tails interact with the DNA in a non-specific manner, usually protruding from the nucleosome superhelix and embracing the DNA. The role the histone tails have in controlling the gene expression at the genomic scale has been extensively analyzed.[3–5] Several studies described the role of H3 and H4 histone tails in the packing of nucleosomes.[6,7] Similarly, Vogler et al[8] demonstrated that the C-terminal tail of the histone H2A can affect nucleosome dynamics, repositioning, and also interact with other proteins that will modulate chromatin dynamics, such as linker histones or chromatin remodelers. The charged residues in this histone tails are the targets to the so-called "post-translational modifications": additions

of chemical groups (mainly acetyl or methyl groups) that will act as epigenetic markers to modulate the chromatin accessibility and the genes that are expressed at a given time point in a specific cellular context.[3] For example, the tri-methylation of the K9[9] or K27 residues of H3[10] marks regions of closed, inactive chromatin, thus contributing to gene silencing. Contrarily, the acetylation of lysines in H3[11] and H4[12] mark open, active regions of chromatin contributing to gene activation. Overall, the tails modulate chromatin structure by controlling both the inter- and intra-nucleosome interactions.

Investigating how the tails affect the structural dynamics of single nucleosomes with high resolution techniques is challenging. Experimental approaches such as single-molecule Förster Ressonance Energy Transfer have been used to describe in detail the dynamics of the histone tails. For instance, Lehmann et al.[13] recently confirmed the role of the N-terminal tail of H3 as a "close pin" that modulates nucleosome breathing and unwrapping. However, due to the high flexibility of the tails, the atomic resolution is difficult to achieve in vitro: only one of the many crystal structures of nucleosomes has resolved histone tails.[14]

To describe motions of single nucleosomes, molecular dynamics simulations represent a powerful alternative. By solving Newton's equation of motion, trajectories of biomolecular systems at high resolution can be obtained. Because of the long time scale required for large amplitude nucleosome motions and the bulky, disordered nature of the histone tails, the computational resources required for atomic resolution simulations are often prohibitive. Therefore, many studies used coarse grained representations of the nucleosome at different resolutions in such simulations.[15–22] This type of simulations have been used to study the role of DNA sequence in nucleosome unwrapping[17,19,20], protein-mediated nucleosome remodeling[22], or the role histone tails play in nucleosome mobility, intra-nucleosome interactions and nucleosome unwrapping.[15,18,21,23] However, the simplifications used in coarse grained models fail to accurately characterize interactions that depend on chemical properties of atoms.

Therefore, for a full characterization of the histone - DNA interactions and how they control nucleosome dynamics, all-atom molecular dynamics simulations are required. Such

simulations have been reported for engineered nucleosomes. However, in most of them, either the nucleosomes were incomplete or the sampling achieved was limited. For example, from simulations of nucleosomes without the complete histone tails, and using biased molecular dynamics, the histone-DNA interactions that are involved in nucleosome unwrapping were described.[24] Similarly, nucleosome opening has been observed in nucleosomes with partially truncated tails in simulations with high salt concentrations.[25] Other studies have shed light into the hydration patterns around the nucleosome core particle[26] and the local flexibility of nucleosomal DNA.[27–29] However, due to the absence of histone tails from the nucleosomes, these studies provide an incomplete description of nucleosome dynamics.

The structural flexibility of the histone tails was demonstrated in simulations of the histones alone or peptides mimicking the tails.[30,31] Some simulations of nucleosomes with complete histone tails at atomic resolution were reported. In these, the nucleosome structural flexibility observed was limited due to the short time scales or the small number of trajectories[32–38] . For example, some nucleosome breathing was observed by Shaytan et al[35], when simulating a nucleosome for 1µs. They observed that on this timescale, the histone tails collapsed on DNA in positions and conformations that were stable throughout the simulation. Bowerman and Wereszczynski[36] described the mechanisms by which H2A histone variants alter nucleosome dynamics from an ensemble of 4 simulations, each 600 ns long. Li and Kono[37] reported a 3.36µs ensemble of short ∼100 ns trajectories at high temperature (353 K), and observed nucleosomal opening when nucleosomes were truncated by removing either the H3 or H2AC tails.

All these studies have been performed mainly using two DNA sequences. The first one is a palindromic DNA sequence derived from human α-satellite DNA, used in the only crystal structure available of the nucleosome with resolved histone tails.[14] The second is the Widom 601 sequence, an artificial sequence selected for its strong positioning and high nucleosome stability[39] that is also present in several crystal structures.[40,41] For gene regulation in the cell, the structural flexibility and mobility of nucleosomes are highly relevant. Genomic nu-

cleosomes sample a variety of sequences, that are believed to behave more dynamically than the engineered nucleosomes.[42] Nevertheless, studies at atomic resolution confirming this and elucidating their structural flexibility are lacking. Recently, we studied with atomistic simulations two genomic nucleosomes, found in the enhancers of 2 genes, *ESRRB* and *LIN28B*, important for stem cell pluripotency.[43,44] These nucleosomes are bound by the transcription factor Oct4, a master regulator of stem cell pluripotency, during the conversion of skin cells to stem cells.[45,46] We observed a high degree of structural flexibility in one genomic nucleosome(Lin28b), involving nucleosome breathing and twisting motions.[38]

Here, from a total of 18 µs atomistic simulations of the engineered Widom and the genomic Lin28b and Esrrb nucleosomes, including the previously reported 9 µs performed with *Drosophila* histones and an additional ensemble of 9 µs with human histones, we demonstrate how the interplay between histone H3 and H2AC tails cooperate to control nucleosome breathing motions. We monitored 2 rare large opening events, one occurring in each genomic nucleosome, and found that nucleosome conformations with different degree of opening display specific H3 and H2AC conformations and positions. Moreover, nucleosome opening and closing is regulated by specific pattern of interactions between residues in H3 and H2AC tails that can act as epigenetic regulatory sites, and the DNA. Because the distribution of open and closed nucleosomes in chromatin fibers impacts their structure and compaction, the mechanism we describe here is key to understanding chromatin dynamics.

Materials and Methods

Nucleosome Modeling

The 146 bp Widom nucleosomes with *Drosophila melanogaster* and *Homo Sapiens* histones were built by homology modelling, using Modeller (https://salilab.org/modeller/), with the same procedure described in our previous report.[38] As templates, we used the structure of the *Drosophila melanogaster* nucleosome core (PDB ID: 2PYO), the structure of the Widom

601 nucleosome particle (PDB: 3LZ0), and the crystal structure of the X. laevis nucleosome that includes histone tails (PDB: 1KX5). For each nucleosome, 100 homology models were generated using a "slow" optimization protocol and a "slow" MD-based refinement protocol as defined in Modeller. We selected the lowest energy model and added an 11 bp fragment of B-DNA (generated with Nucleic Acid Builder from Ambertools 18[47]) to each linker DNA to generate the 168 bp nucleosomes.

The Esrrb and Lin28b enhancer sequences were the same as used in our previous work[38] (Suppementary Document S1), originally selected from data by Soufi et al.[45] We substituted the Widom sequence with the Lin28b and the Esrrb sequences using the "swapna" function in Chimera (https://www.cgl.ucsf.edu/chimera/). In the human Lin28b nucleosome, the sequence was shifted by 1 extra bp (2 compared to the construct presented in Soufi et al.[46]). This shift was introduced following our previous work[38] to optimize the accessibility of the Oct4 binding sites.

Molecular dynamics simulations

The MD protocol was described previously.[38] In short, each nucleosome was solvated in a truncated octahedron box of at least 12 Å of SPCE water molecules around the solute in any direction. Na^+ ions were added to neutralize the charge of the system, and K^+ and Cl^- ions were added up to a concentration of 150mM KCl. Then an energy minimization using the AMBER software[47] was performed, followed by 10.25 ns of equilibration with gradually decreasing positonal and base pairing restraints.[38,48] The equilibration and the following production runs were performed using NAMD.[49] The production runs were in the isobaric-isothermic (NPT, p = 1 atm T = 300 K) ensemble, using Langevin dynamics as a termostat and Nosé-Hoover and Langevin piston for pressure control. The AMBER force field parameters ff14SB[50], parmbsc1[51] and Li-Merz[52] were used for protein, DNA and ions, respectively.

Analysis of nucleosome dynamics

All simulations were fitted with a root mean-square fit of the heavy atoms of the histone core to the minimized structure of the Widom nucleosome. The histone core was defined by excluding the histone tails: residues 1–45 for hH3 and dH3, 1–32 for hH4 and dH4, 1-18 and 129 for hH2A, 1–17 and 115–124 for dH2A, and 1-33 for hH2B and 1–31 for dH2B.

The characterization of the breathing motions was done using the procedure first described by Öztürk et al.[53], that we also used in our previous work.[38] We first defined a coordinate system XYZ with the origin on the dyad. X was defined as the vector along the dyad axis. Y was defined as the cross product between X and a vector perpendicular to X intersecting it approximately at the center of the nucleosome. Finally Z was defined as the cross product between X and Y. Then the vectors v_3 and v_5 were defined along the 3' and 5' L-DNAs. The angle γ_1 was defined as the angle between the projector of these vectors in the XZ plane and the Z axis, whereas γ_2 was defined as the angle between the projection of the vectors on the XY plane and the Y axis.

The radius of gyration of the DNA (RoG) and the number of protein-DNA contacts were calculated using the cpptraj software.[54] Contacts were defined as protein non-hydrogen atoms closer than 4.5 Å. We split the DNA in two parts. The outer gyre was defined as 40 bp from each end of the nucleosomal DNA, and the inner gyre was defined to include all remaining base pairs. For the histones, contacts were split between contacts made by the histone core and the histone tails, using the core and tail definition described above.

Mean and minimal distances were measured using the weighted mean distance collective variable *distanceInv* implemented in the Colvar module[55] available in VMD.[56] The weighted distance between two groups of atoms is defined as follows: $d_{1,2}^{[n]} = (\frac{1}{N_1 N_2} \sum_{i,j} (\frac{1}{||d^{ij}||})^n)^{-1/n}$, where $||d^{ij}||$ is the distance between atoms i and j in groups 1 and 2 respectively, and n is an even integer. This distance will asymptotically approach the minimal distance when increasing n. For smaller values of n, weighted distance will be close to the true average distance. For measuring the position of the histone tails relative to DNA, we calculated the

weighted mean distances with $n = 10$ (from now on referred to as "mean distances") from the tails to two reference points: the L-DNA, defined as the last 15 basepairs of each DNA arm, and the dyad, defined as the 8 bp next to the dyad on the same side as the tail (3' and 5'). Minimal distances between individual residues in histone tails and the inner gyre or the outer gyre of DNA (as defined above), were calculated as weighted mean distances with $n = 100$.

Clustering of MD trajectories

MD trajectories of all nucleosomes with the same histones were combined, obtaining 2 groups of simulations: one with human histones (hH simulations) and one with *Drosophila* histones (dH simulations). hH and dH simulations were separated because a clustering of the combined ensemble of simulations based on the histone tails would not be possible due to the differences in histone sequences between the 2 organisms . The snapshots from each group of simulations were clustered based on the positional RMSD of the backbone heavy atoms of DNA, H3 and H2A. For the DNA, the outer gyre that didn't display big opening events (5' and 3' in hH and dH simulations respectively) was not considered in the clustering. For the histone H3, the RMSD of the entire histone was used for clustering, whereas for the H2A, the RMSD of the entire histone excluding the N-terminal tail was used. This resulted in 3 independent clustering distributions, referred to as "DNA clusters", "H3 clusters", and "H2AC clusters". The k-means clustering algorithm implemented in cpptraj was used, with 8 centroids. Finally, the percentage of snapshots from each DNA cluster in each H3 and H2AC cluster was computed, obtaining the distributions of the DNA cluster snapshots in the histone tail clusters to evaluate the correlation between DNA conformations and the H3 and H2AC conformations and positions.

Results

Breathing motions are more extensive in genomic nucleosomes

We performed 3 independent 1 µs simulations for each combination of DNA and histones (Table 1), which we grouped in ensembles of 3 µs. From these, we found that the highest structural flexibility of the nucleosomes was in the L-DNA arms (both 5' and 3') and the histone tails (Figure 1). This was reflected in the broader range of the number of contacts with DNA formed by the histone tails compared to the histone core (Table1, Supplementary Figure S1). All the tails were in direct contact to the core DNA. However, only the N-terminal tail of the histone H3 and the C-terminal tail of H2A (from now on, the H3 and H2AC tails) were in direct contact with both linker and core DNA.

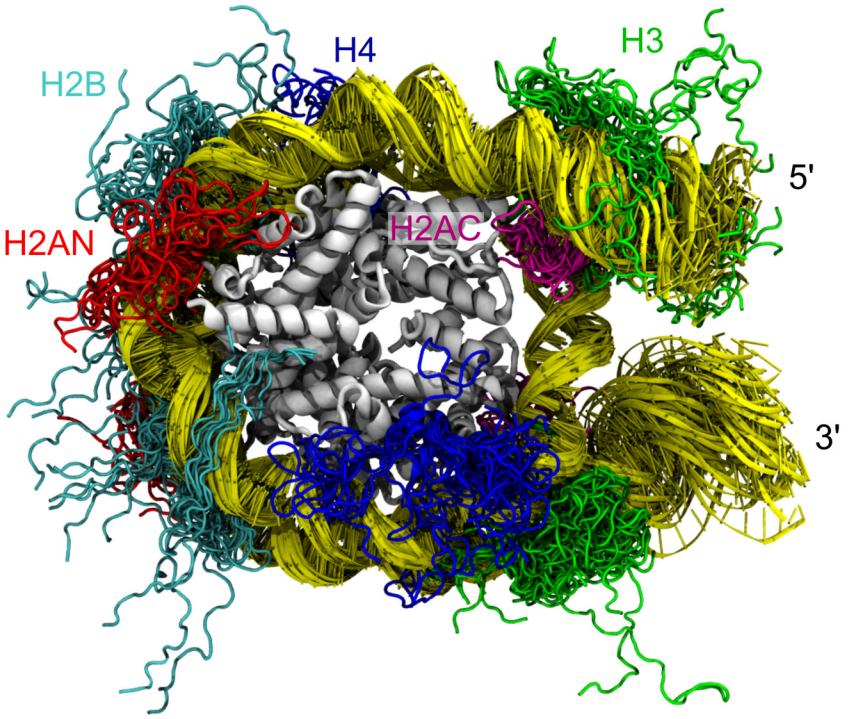


Figure 1: **Nucleosome motions on the µs time scale.** The representation was obtained by the superposition of every 5 ns of the 3 µs simulation ensemble of the Widom nucleosome with human histones. The histone core is depicted in gray, DNA in yellow, and the histone tails of H3, H4, H2A (both N and C terminal) and H2B are in green, blue, red, magenta and cyan respectively.

Table 1: Overview of the simulations performed

Histones	DNA	Replica	Time	RoG[1]	Number of contacts[2] with DNA		
					Histone Core	Histone Tails	
						inner gyre[3]	outer gyre[3]
Drosophila	Widom	1	1 µs	47.11–48.34	1914–2124	790-1514	891-1360
Drosophila	Widom	2	1 µs	47.62–49.58	1837–2151	683-982	743-1250
Drosophila	Widom	3	1 µs	48.08–49.03	1780–2019	769-1299	907-1421
Drosophila	Esrrb	1	1 µs	47.37–48.82	2010–2265	937-1388	764-1187
Drosophila	Esrrb	2	1 µs	47.26–48.34	1869–2161	675-1269	989-1295
Drosophila	Esrrb	3	1 µs	47.11–48.41	2013–2241	894-1375	783-1146
Drosophila	Lin28b	1	1 µs	48.00 49,91	1917–2170	837-1353	942-1354
Drosophila	Lin28b	2	1 µs	48.34–50.90	1928–2173	755-1718	639-1185
Drosophila	Lin28b	3	1 µs	47.63–49.26	1977–2171	917-1384	812-1066
Human	Widom	1	1 µs	47.93–49.26	2017–2224	762-1260	754-1031
Human	Widom	2	1 µs	47.26–48.54	1867–2116	1061-1673	717-1023
Human	Widom	3	1 µs	48.06–49.18	1807–2047	856-1442	716-1226
Human	Esrrb	1	1 µs	47.83–51.32	1708–2057	1147-1893	457-763
Human	Esrrb	2	1 µs	48.22–49.87	1844–2062	888-1631	777-1184
Human	Esrrb	3	1 µs	47.14–48.18	1926–2134	901-1438	666-1098
Human	Lin28b	1	1 µs	47.79–48.93	1907–2121	752-1123	828-1488
Human	Lin28b	2	1 µs	47.93–49.00	1852–2048	827-1284	830-1261
Human	Lin28b	3	1 µs	47.90–49.62	1792–2048	667-1549	739-1119

[1] RoG and the number of contacts are shown as the percentiles 5 to 95.
[2] A contact was defined as a non-hydrogen atom closer than 4.5 Å to another non-hydrogen atom.
[3] Inner and outer gyre of the nucleosomal DNA, see Methods for details.

To characterize the structural flexibility induced by the motions of the L-DNAs, we used the radius of gyration (RoG) and two angles, γ_1 and γ_2. These angles provide a complete description of the nucleosome breathing motions in the XZ (γ_1) and XY (γ_2) planes defined by a dyad-centered coordinate system[38,53] (Figure 2A, see Methods). The main peaks of the RoG distributions obtained from the simulations performed with human and *Drosophila* histones (hH and dH respectively) were analogous, indicating that the conformational space sampled in the 2 sets of simulations was similar. We found that the genomic nucleosomes Lin28b and Esrrb adopt more open conformations with the exception of the Esrrb^{dH} nucleosome (Figure 2B, Supplementary Figure S2). In contrast, both Widom nucleosomes sampled mainly a conformation with the RoG centered at $\approx$ 48.25 Å and a second, less populated conformation with RoG $\approx$ 47.5 Å. The latter was similar in RoG with the closed conforma-

tion sampled by the Esrrb^{dH} nucleosome. The Esrrb^{hH} nucleosome sampled three different regions of the conformational space. The conformations in two of them displayed RoG values resembling those sampled by the Widom nucleosomes, whereas the conformations in the third region were more open with higher RoG values (>50). The Lin28b nucleosomes did not sample the most closed conformations observed in the other nucleosomes and in general adopted more open conformations (RoG 48.25 - 49.5 Å). In addition, the Lin28b^{dH} sampled open conformations with RoG > 50 Å, matching those observed for the Esrrb^{hH} nucleosome (Figure 2B).

The most open conformations of the Esrrb^{hH} and Lin28b^{dH} nucleosomes were sampled in simulations 2 and 1 of each simulation group respectively and were characterized by higher RoG and lower number of contacts between the histone tails and the outer DNA gyre (Table 1, Supplementary Figure S1). The opening was induced by increased structural flexibility of the 3' and 5' L-DNA respectively. They involved opening in both XZ and XY planes, as well as opening in XY plane coupled with closing in XZ plane (Figure 2C). Except of these rare, large opening events, the sampling of the conformational space characterized by the histograms of the γ angles was similar between the hH and dH nucleosomes.

To this point, we showed that genomic nucleosomes breath with a larger amplitude than engineered nucleosomes on the simulated time scales. We monitored 2 large amplitude openings in the Esrrb^{hH} and Lin28b^{dH} nucleosomes.

Histone H3 and H2A tails sample a broad range of configurations.

Based on the visual analysis of the simulations, we hypothesized that the conformational diversity of the nucleosomal DNA, and some of the differences between the individual simulations are due to the conformational plasticity and positional fluctuations of the histone tails. To test this hypothesis, we first defined a specific combination of histone tail conformation and position as a histone tail configuration. Then, we characterized the configurations observed in the simulations, focusing on the H3 and H2AC tails because they are in contact

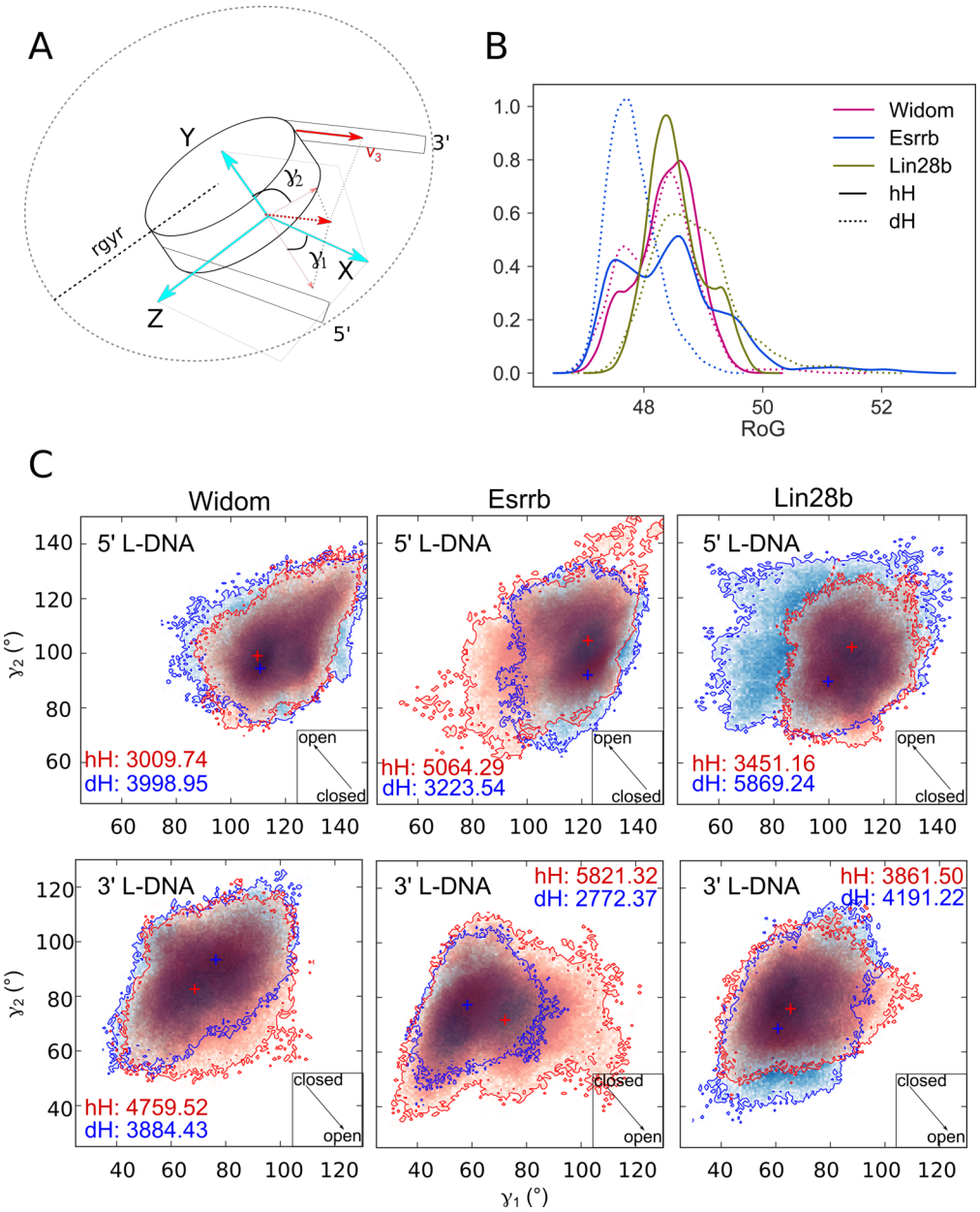


Figure 2: **Structural flexibility of the hH and dH nucleosomes. A)** Schematic representation of the RoG definition and the coordinate system used to describe nucleosome conformations. The angles γ_1 and γ_2, describing the nucleosome breathing motions were defined as described in Methods. An increase of γ_1 indicates opening at the 3' L-DNA, but closing at the 5' L-DNA, and viceversa for γ_2. **B)** RoG distribution from the 3 µs ensembles of simulations. Larger values indicate a lower degree of compaction in the nucleosomal DNA. Solid and dashed lines represent the hH and dH simulations, respectively. The Widom, Esrrb, and Lin28b nucleosomes are shown in magenta, blue and green respectively. **C)** Two-dimensional histograms depicting the conformational sampling of the 5' and 3' L-DNA arms in the space defined by the γ_1 and γ_2 angles. The arrow inserts in the lower-right corner indicate the direction of the opening, whereas the numbers are the areas of the histograms. The crosses indicate the most populated region of each histogram. The data for the hH and dH nucleosomes are in red and blue respectively.

with the L-DNA arms. For this, we monitored their position relative to the corresponding L-DNA and the DNA segment around the dyad and the number of contacts with the inner and outer gyre (see Methods for details) in the 3 µs ensembles of simulations.

We found that the H3 and H2AC tails neighboring the L-DNA arms with the highest flexibility (3' L-DNA of the Esrrb^{hH} and the 5' L-DNA of the Lin28b^{dH} nucleosome) sample a broad variety of configurations (Figure 3, Supplementary Figure S3, Supplementary Document S3). This diversity was not only due to different conformations and positions sampled in different simulations but also due to transitions occurring during the individual 1 µs simulations. Moreover, we observed correlations between tail configurations and nucleosomal DNA conformations. In general, when the H3 tail was in an intermediate position between the L-DNA and the dyad of the L-DNA, the tail acted as a bridge that kept the nucleosome closed. A large number of contacts to the outer gyre, and a short distance to the L-DNA, correlated with intermediately closed DNA conformations. When the H3 or H2AC tails were in the proximity of the dyad and further away from the L-DNAs forming only few contacts to the outer gyre, the nucleosomes opened.

During the opening of the Esrrb^{hH} nucleosome (Esrrb^{hH} simulation 1, between 750-850 ns), we observed an increase in the distance to the L-DNA and in the number of contacts to the dyad for the H3 tail. In this configuration, the H3 tail collapsed on the DNA around the dyad, allowing the DNA opening (Figure 3A). At the same time, the H2AC position relative to the dyad remained stable, but its contacts to the L-DNA were lost, further contributing to the opening (Figure 3B). Towards the end of the first simulation, the H2AC tail - L-DNA contacts were quickly reformed, closing the DNA.

On the other hand, in the largest opening event of the Lin28b^{dH} nucleosome (Lin28b^{dH} simulation 2, between 700-800ns), the H3 tail remained in a stable position during the opening: it did not collapse around the dyad, but in the center of the 5' L-DNA (Figure 3C). In this case, a transition in the configuration of the H2AC tail explained the opening of the DNA (Figure 3D). This transition involved an increase of the distance to the L-DNA,

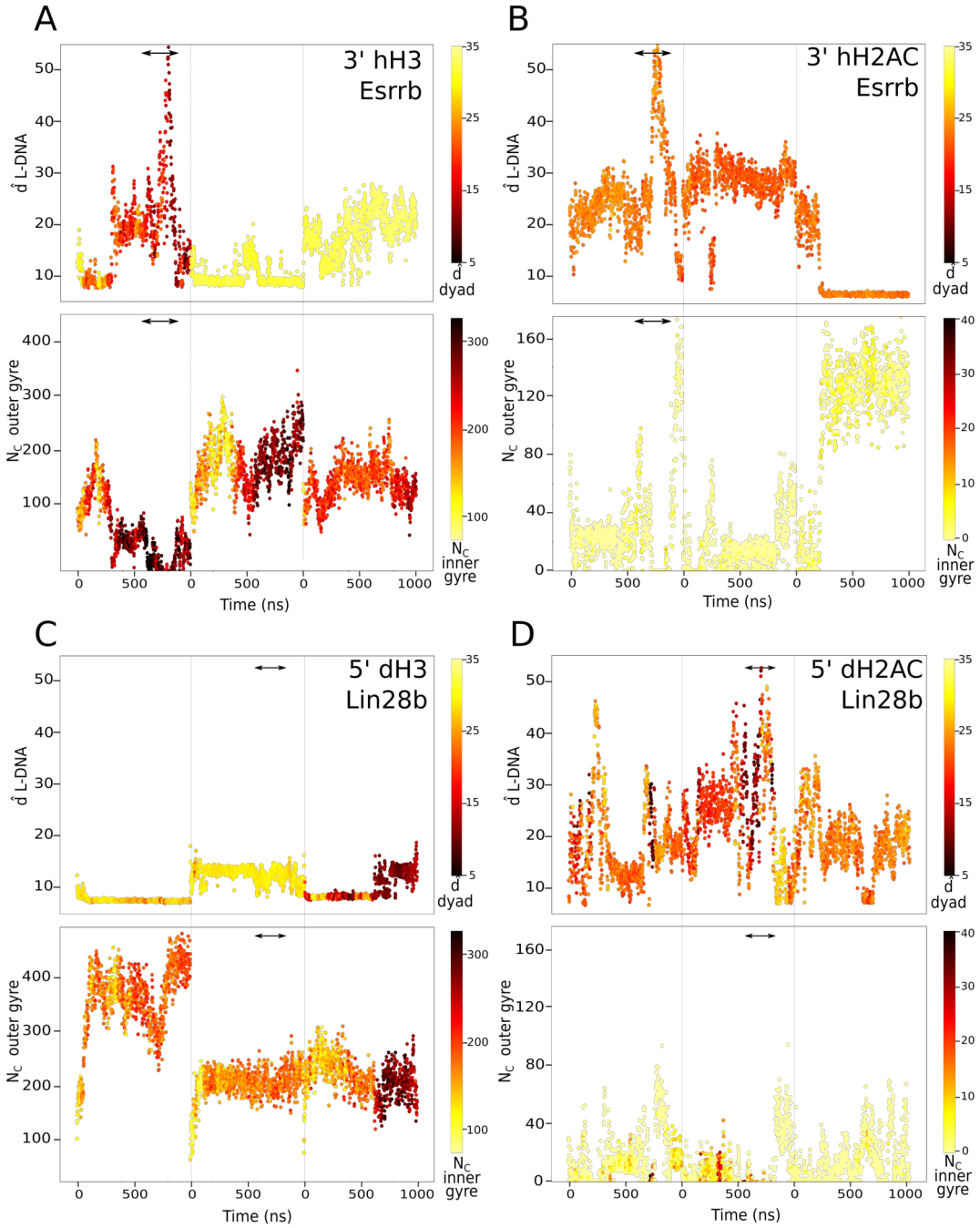


Figure 3: **Configurations of the H3 and H2AC tails.** Each panel shows the weighted mean distance to the L-DNA (top) and the number of contacts with the outer gyre (bottom) during the simulations. The points are colored by the weighted mean distance to the dyad (top) and the number of contacts to the inner gyre of the DNA (see Methods for details on the definition of inner and outer gyre). **A-B)** H3 (A) and H2AC (B) tails at the 3' end in the simulations of the Esrrb^{hH} nucleosome. **C-D)** H3 (C) and H2AC (D) tails at the 5' end in the simulations of the Lin28b^{dH} nucleosome

and a decrease of the number of contacts to the outer gyre. Again, the contacts with DNA reformed shortly after the opening occurred, closing the DNA.

From this analysis of the histone H3 and H2AC tail configurations, we show that these tails sampled a broad range of conformations and positions during the simulations. Furthermore, we observed that some nucleosome conformations correlate with specific positions of the tails.

Nucleosome conformations correlate with H3 and H2A tail configurations.

Next, we investigated whether there is a correlation between nucleosome conformations with different degree of opening and the configurations of the H3 and H2AC tails. For this, we performed a clustering analysis of the structural snapshots from the simulations. We merged all the hH simulations in one group, and all the dH simulations in another. Since the analysis was not aimed at revealing sequence specific effects on nucleosome dynamics, we could group together simulations with different DNA sequences. On the other hand, we could not merge all simulations in a single group because the human and *Drosophila* histone tails are not identical. For each group, we performed 3 different k-means cluster analyses, obtaining 8 clusters ranked by cluster size. To separate nucleosome conformations according to their degree of opening, we clustered based on the RMSD of the backbone of DNA bases of the inner gyre and the outer gyre on which the large opening occurred (see Methods). To separate nucleosome conformations according to histone tail positions, we clustered based on the RMSD of the non-hydrogen atoms of the H3 and H2A histones separately (excluding the N terminal tail of H2A).

We mapped the DNA clusters on the 2D histograms of the γ angles (Figure 4A,B, Supplementary Figure S4) and confirmed that the clusters spread across the sampled conformational space according to the degree of opening. Overall, the clusters obtained from the 2 groups of simulations displayed an almost symmetric distribution (Figure 4A,B, Supplementary Fig-

ure S3), demonstrating that the conformational sampling in the hH and dH simulations was similar. The largest clusters (DNA^{1h}, DNA^{1d}) contained the more closed conformations in both simulation groups, whereas the less populated clusters (DNA^{7h}, DNA^{8h} and DNA^{6d}) contained conformations open in both directions defined by γ_1 and γ_2.

Next, we evaluated how the DNA clusters correlate with the H3 and H2AC clusters (Figure 4 C,D). For this, we calculated the percentage of frames of each DNA cluster found in each H3 and H2AC cluster. To characterize the histone tail clusters, we calculated for each snapshot in every cluster the RoG of the tails and the number of contacts to the inner and outer gyre of the nucleosome (Figure 4 E,F).

The most remarkable correlation we found was between the most open conformations of the DNA and specific configurations of the H3 and H2AC tails (Figure 4C,D). In the hH simulations, the clusters containing the most open DNA conformations DNA^{7h} (Figure 5A) and DNA^{8h} correlated with the clusters $hH3^6$, and $hH2AC^4$ (Figure 4C). These clusters contain H3 and H2AC tail configurations with few contacts with the outer gyre of the DNA and a wide distribution of conformations (described by the RoG distribution), that contain the most extended tail conformations sampled (Figure 4E). Similarly, in the dH simulations, the cluster DNA^{6d} (Figure 5B) correlated with the H3 cluster $dH3^5$, and partially with $dH2AC^3$ (Figure 4D). The correlation with H2AC clusters was less clear when compared with the hH simulations. Again, $dH3^5$ contains H3 tail configurations with few contacts with DNA (both the inner and outer gyre) and extended conformations (higher RoG values) (Figure 4F). The H2AC clusters from the dH simulations contain rather similar H2AC configurations. This may be explained by the difference between H2AC in human and *Drosophila*: dH2AC is shorter and has less positive residues compared to hH2AC (Supplementary Document S2). $dH2AC^3$ contains dH2AC configurations with a slightly higher number of contacts with the inner gyre of DNA, being in proximity to the core nucleosomal DNA. These findings revealed that the higher amplitude opening of the two genomic nucleosomes was achieved when the H3 and H2AC tails adopted configurations with few interactions with the DNA, especially

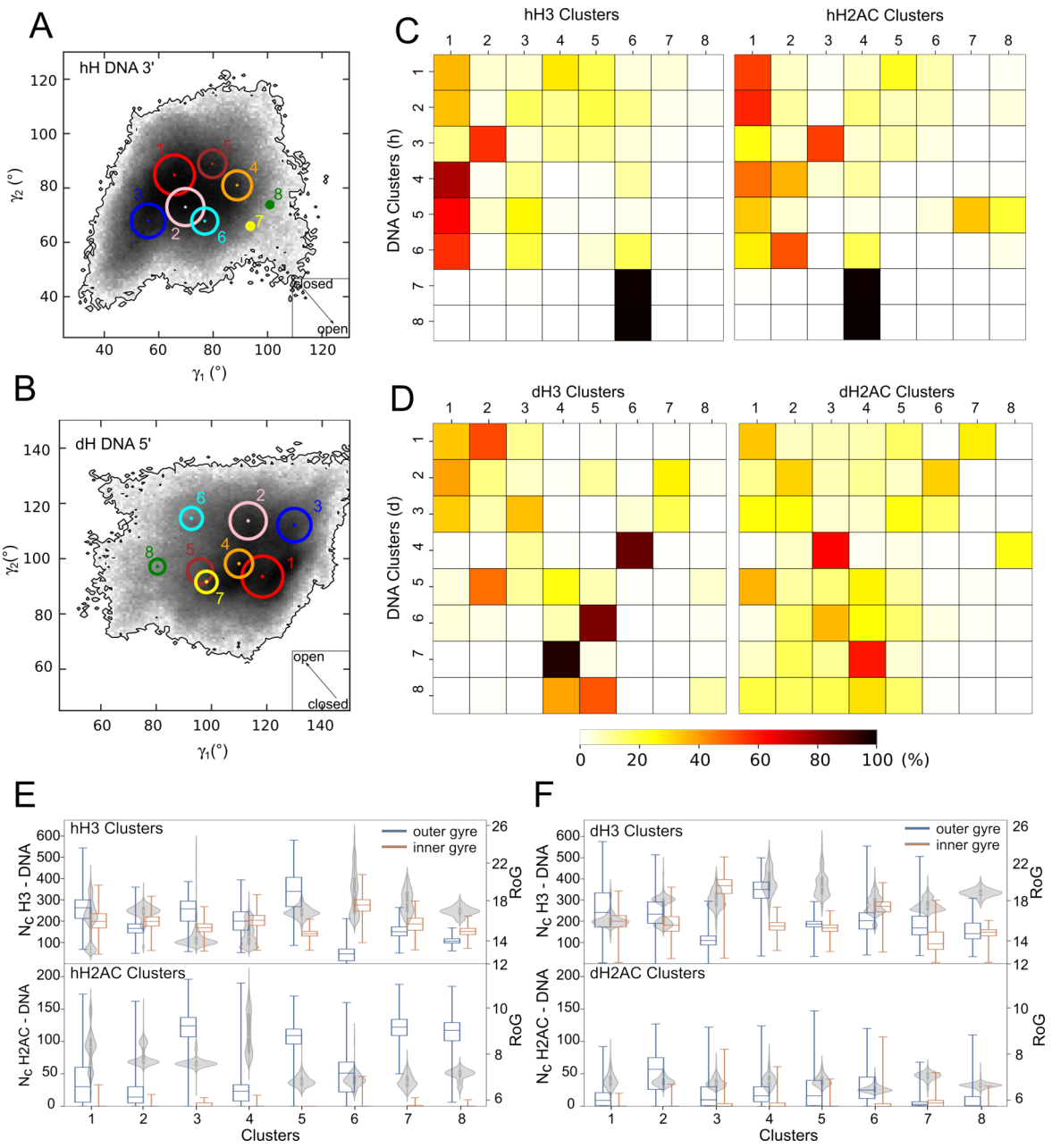


Figure 4: **Clustering of the DNA and histone tails. A-B**) Position of the DNA cluster centroids in the 2D histogram of the γ distribution for the 3' L-DNA in hH simulations (A) and 5' L-DNA in dH simulations (B). The size of the circles indicates the size of the clusters. **C-D)** Distribution of the frames from each DNA cluster on the tail clusters (H3 in the left matrix, H2AC in the right), for hH (C) and dH (D) simulations. For each DNA cluster (row), the cells are colored by the percentage of frames of that cluster that belong to the tail cluster (column). **E-F)** Features of H3 and H2AC configurations per cluster in the hH (E) and dH simulations (F). Box plots indicate the number of contacts between H3 (top) or H2AC (bottom) tail and the inner (red) and outer gyre (blue) of DNA per tail cluster. Violin plots represent RoG of the H3 (top) or H2AC (bottom) tail per cluster.

with the linker DNA.

We also observed correlations with histone tail configurations for the DNA clusters containing conformations that are open only in one direction (defined by one γ). From the hH simulations, clusters DNA^{4h} and DNA^{5h} (Figure 5C), that contain conformations open only in the direction defined by γ_1 correlate almost exclusively with $hH3^1$ and $hH3^3$ (Figure 4C). $hH3^3$ contains H3 configurations with a high number of contacts to the outer gyre, and remarkably, with compact tail configurations (low RoG values). Such conformations are also found in the $hH3^1$ cluster (Figure 4E). Therefore, when the nucleosome is closed in the XY plane (γ_2), the H3 tail adopts a compact, bridging conformation that makes contacts with both the inner and outer gyre of DNA (Figure 4E). In addition, we observed that the DNA conformations open only in the XZ plane (γ_1) correlated with H2AC configurations found in $hH2AC^1$ (Figure 4C). This is the largest hH2AC cluster and it has configurations with less contacts to the outer gyre than those found in other clusters (Figure 4E) . The differences between the DNA conformations found in clusters DNA^{4h} and DNA^{5h} may be explained by the alternative configurations of hH2A, because they correlate with hH2AC configurations in $hH2AC^2$ and $hH2AC^7$-$hH2AC^8$ respectively (Figure 4C).

The hH nucleosome conformations open only in the XY plane (γ_2) (DNA^{3h} , Figure 5C), correlated with histone tail configurations found in clusters $hH3^2$ and $hH2AC^3$ (Figure 4C). Both clusters contain conformations with high values of RoG which are characteristic for more extended tail conformations (Figure 4E). In addition, the configurations found in the $hH3^2$ cluster show only little variation in the positions of the tail relative to the DNA as revealed by the narrow distribution of the number of contacts with the inner and outer gyre of the DNA (Figure 4E). This indicates, that only specific H3 tail configurations are compatible with the nucleosome opening in the XY plane.

In the dH simulations, the cluster DNA^{8d} with conformations open only on the XZ plane (γ_1) (Figure 5D), correlates with the H3 clusters $dH3^4$ and $dH3^5$, but has not a significant correlation with any H2AC cluster. $dH3^5$ has tail configurations that correlate with the

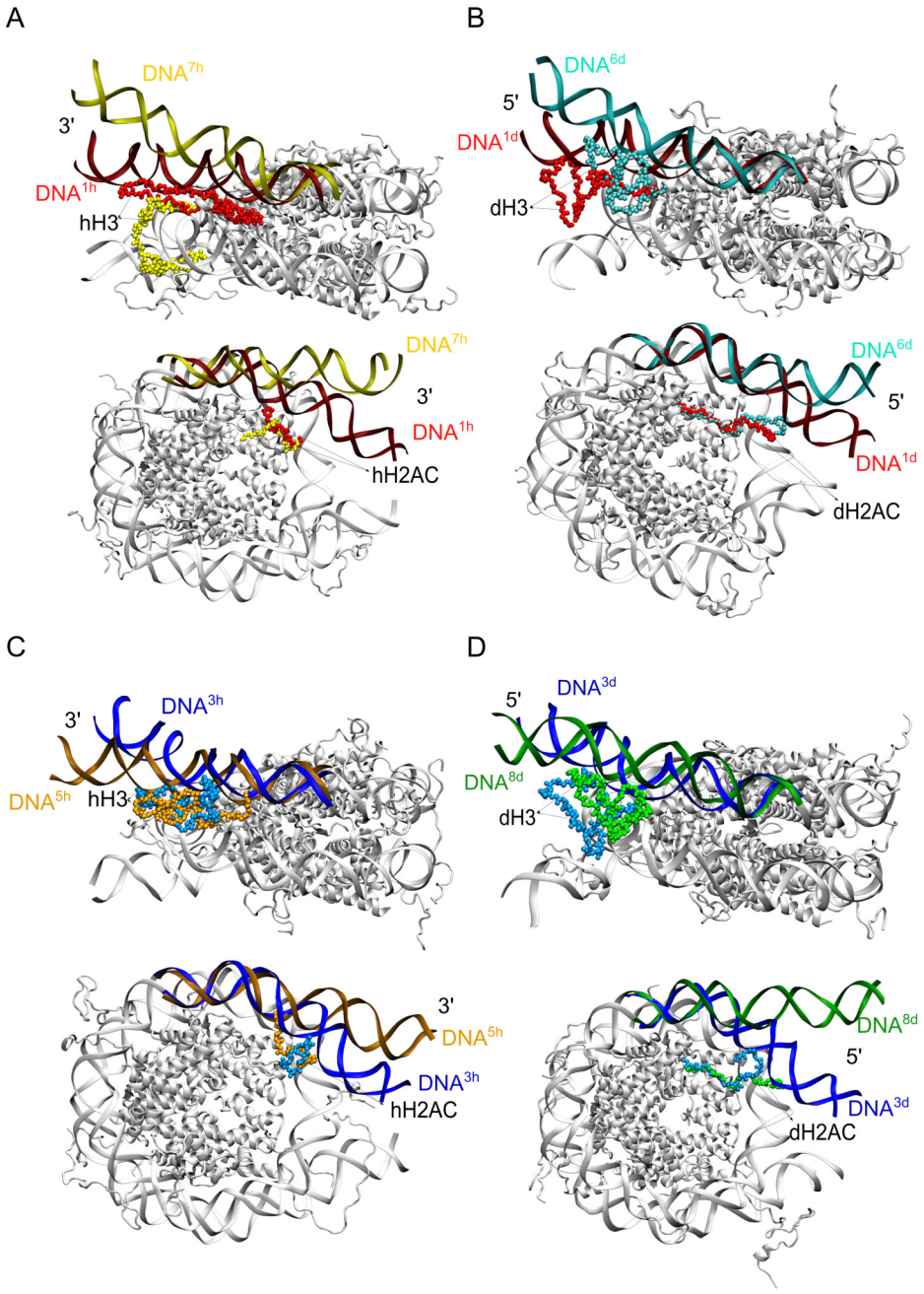


Figure 5: **Structures of DNA cluster representatives.** The L-DNA used for clustering is colored by the cluster color from Figure 4A-B. The L-DNA is in a dark shade, whereas the nearby histone tail is in a lighter shade of the same color. Each panel shows a side view of the nucleosome with the H3 tail (top), and a top view of the nucleosome with the H2AC tail (bottom). **A-B)** Conformations closed or open in both the XZ (γ_1) and XY planes (γ_2). For the hH simulations (A), clusters DNA^{1h} (closed) and DNA^{7h} (open). For the dH simulations (B), clusters DNA^{1d} (closed) and DNA^{6d} (open). **C-D)** Conformations open only in one plane, closed in the other. For the hH simulations (C), clusters DNA^{3h} (open in the XY plane) and DNA^{5h} (open in the XZ plane). For the dH simulations (D), clusters DNA^{3d} (open in the XY plane) and DNA^{8d} (open in the XZ plane). For clarity, the 3' end of the hH nucleosomes is shown in the same direction as the 5' end of the dH nucleosomes (views are inverted by 180° around the dyad axis).

most open DNA conformations in the dH simulations, whereas $dH3^4$ is the only H3 cluster that correlates with the DNA cluster DNA^{7d}, a cluster that has slightly more closed DNA conformations than DNA^{8d} (Figure 4D). $dH3^4$ contains dH3 tail configurations that have an extended conformation (high RoG values) with significantly more contacts to the outer gyre of the DNA than the configurations found in $dH3^5$. Taken together, these findings indicate that the conformations in DNA^{8d} represent a transition between conformations in DNA^{7d} and the most open conformations found in DNA^{6d}. The DNA cluster with conformations open only on the XY plane (γ_2) (DNA^{3d}, Figure 5D) was correlated mostly with the H3 clusters $dH3^1$ and $dH3^3$ and was not correlated with any specific dH2AC cluster (Figure 4D). The $dH3^3$ contains more extended conformations of the dH3 tail with more contacts with the core nucleosomal DNA than the L-DNAs (Figure 4F). The $dH3^1$ cluster is the largest dH3 cluster and therefore has more diverse dH3 configurations (Figure 4F). These findings further confirm that the loss of a large fraction of the contacts between H3 and H2AC tails and the L-DNAs is required for nucleosome opening to occur.

The most closed DNA conformations found in the clusters DNA^{1h}, DNA^{1d} (Figure 5A) have H3 tail configurations that are either found in the largest H3 clusters or distributed among different clusters (Figure 4C,D). These configurations are either extended or more compact and have a large number of contacts both with L-DNAs and core DNA (Figure 4E,F). Therefore, the nucleosomes are maintained in a closed conformation by H3 mediated bridging interactions between the corresponding L-DNA and the core nucleosomal DNA. In the hH simulations, closed nucleosomes have the H2AC in configurations with extended conformations and with a moderate number of contacts with the L-DNA (Figure 4C,E). On the other hand, in the dH simulations, the configuration of the H2AC tail does not have features specific for closed nucleosomes (Figure 4D,F)

Overall, the clustering analysis demonstrated that nucleosome conformations are strongly correlated with specific configurations of the H3 and H2AC tails. In particular open nucleosome conformations correlate with tail configurations characterized by a decrease in the

number of contacts with the DNA, specially to the adjacent L-DNA arm.

Nucleosome opening is induced by H3 and H2AC tail dynamics

To this point, we described the correlations between DNA conformations with different degree of opening and specific H3 and H2AC tail configurations. However, it remains unclear whether the motions of the H3 and H2AC tails amplify or induce nucleosome breathing. To elucidate this, we analyzed the distributions of the structural snapshots from the DNA, H3, and H2AC clusters during the simulations and compared it with the nucleosome opening profile given by the time evolution of the RoG of the nucleosomal DNA (Figure 6).

We found that reversible conformational transitions in the nucleosomal DNA occur frequently within 1 µs. However, transitions to largely open conformations are rare. In the hH simulations, open DNA conformations found in clusters DNA^{7h} and DNA^{8h} are sampled only in the second simulation of the Esrrb^{hH} nucleosome in the 700-800 ns time interval (Figure 6A). In the dH simulations, the most open conformations are found in cluster DNA^{6d} and DNA^{8d} and are sampled mainly in the first and second simulations of the Lin28b nucleosome (Figure 6B). The largest opening in the second simulations of the Lin28b^{dH} nucleosome was more gradual than that observed in the Esrrb^{hH} nucleosome.

On the other hand, configurational transitions in the H3 tail during a single simulation were rare. With few exceptions, the configuration of the H3 tail was stable in each simulations and differed between different simulations. This finding highlights the importance of grouping simulations in ensembles for investigating histone tail dynamics. For the H2AC tail, we found an important difference between the hH and dH simulations. The longer hH2AC displays configurations that are more stable on the 1 µs timescale comparing to dH2AC.

Interestingly, both large openings of the nuclesomal DNA were preceeded by configurational transitions in both H3 and H2AC tails. The opening sampled in the hH simulations occurred after both H3 and H2AC tails adopted specific conformations (clusters $hH3^6$ and $hH2AC^4$) (Figure 6A). For the opening observed in the Lin28b^{dH} nucleosome, the H3 tail

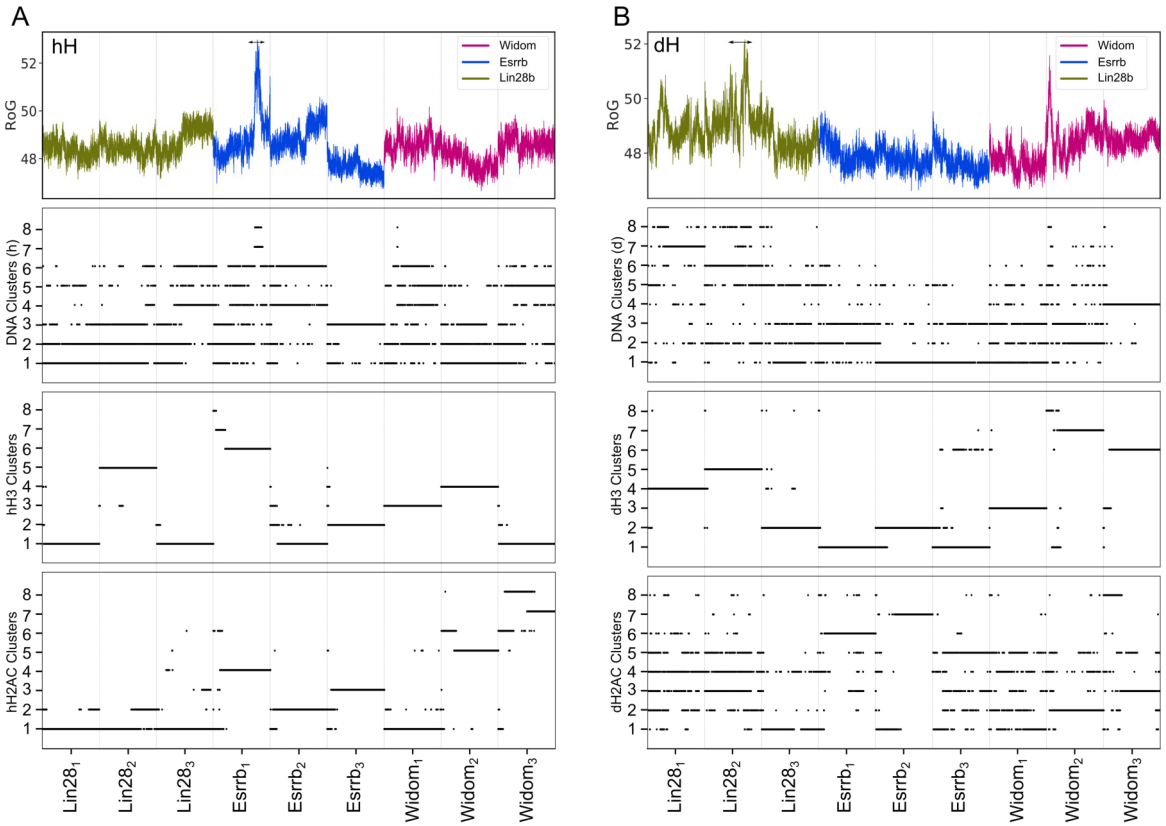


Figure 6: **Structural transitions in the DNA, H3 and H2AC tails.** Structures from each DNA, H3, and H2AC cluster mapped on the simulation time series. **A**) The 3' end of the hH nucleosomes. **B**) The 5' end of the dH nucleosomes. In each panel, the top plot shows the nucleosome RoG and the 3 plots below show to which DNA, H3, and H2AC clusters each simulation snapshot belongs to. Each point corresponds to one single snapshot.

was in a configuration from the $dH3^5$ cluster over the entire simulation. This configuration was not sampled in any other simulation. On the other hand, the H2AC tail transitioned to conformations from clusters $dH2AC^6$ and $dH2AC^8$ before the largest amplitude DNA opening.

Simulation 1 of the Widom^{dH} nucleosome (Table 1, Figure 6B) confirms that an interplay between both tails is required for a large opening of the nucleosomal DNA. In this particular simulation the H3 tail sampled configurations of the $dH3^3$ cluster. This is the dH3 cluster with less contacts to the outer gyre, and a relatively high RoG distribution (Figure 4F), suggesting that the dH3 tail was in extended conformations, compatible with DNA opening. Nevertheless, since the dH2AC tail didn't explore conformations from the clusters that would allow opening of the DNA ($dH2AC^6$ and $dH2AC^8$), we see that the Widom^{dH} nucleosome stays in one of the most closed conformations of the simulated ensemble.

From the projection of the clusters of nucleosome conformations and histone tail configuration on the simulation time series, we found that large amplitude nucleosome opening happens after both the H3 and H2AC tails adopt particular configurations. Moreover, when the tails transition from these configurations to others, the nucleosomes closed.

Nucleosome opening is modulated by epigenetic regulatory residues

Last, we aimed to elucidate the mechanism by which configurational changes in the H3 and H2AC tails induce conformational changes in the nucleosomal DNA. Taking advantage of the atomic resolution of our simulations we analyzed how specific chemical interactions between the histone tails and the DNA impact on nucleosome conformational dynamics.

For this, we monitored the minimal distance between positively charged lysine and arginine residues in the H3 and H2AC tails and the inner and outer gyres of the DNA (Figure 7). The interaction of these aminoacids with DNA is key to protein-DNA binding in general. In nucleosomes, such residues are chemically modified and involved in epigenetic regulation of gene expression. The modifications are also expected to change the impact on nucleosome

dynamics.

Because the H3 tail has a large number of lysines and arginines and their interaction with the DNA is often redundant, we separated the residues of this tail in 4 groups: the "tip" group includes the residues at the tip of the tail R2, K4, the "center-tip" group includes residues R8, K9, K14, the "center-anchor" group includes residues R26 and K27, and the "anchor" group includes the residues anchoring the L-DNA to the inner gyre of the DNA K36, K37, R40 and R42.

We found that the mechanisms of the two large opening events we observed were similar as they involved the loss of many interactions of these residues with the outer gyre of the DNA. However, we also found some differences between these 2 events.

The open Esrrb^{hH} nucleosome conformations were correlated with H3 tail configurations in cluster $hH3^6$. In these configurations, most of the H3 tail interactions with the outer DNA gyre were lost. In fact, only R42 from the anchor group maintained a minimal distance to the outer gyre that is compatible with hydrogen bond interactions. The tip and center-tip interactions were never formed in the simulation in which the large nucleosome opening occurred, whereas the center-anchor deattached from the outer gyre long before the breathing, demonstrating that the loss of these interactions is a prerequisite for the opening. (Supplementary Figure S5). Some of the residues from the tip and the center groups (R2, R8, R26) shifted position to interact with the inner gyre. Interestingly, the lysines involved in epigenetic regulation have no interaction with DNA in these configurations (Figure 7A). The H2AC tail configurations found in these open nucleosomes (cluster $dH2AC^4$) do not display any interactions with DNA (Figure 7B). The interactions between the H2AC tail and the outer gyre of the DNA were lost before the opening occurred. In contrast, when the Esrrb^{hH} nucleosome remained closed, K126, K128, and K130 formed stable interactions with the outer gyre (Supplementary Figure S6).

The open Lin28b^{dH} nucleosome conformations were correlated mainly with H3 tail configurations from cluster $dH3^5$. These configurations kept a series of interactions between the

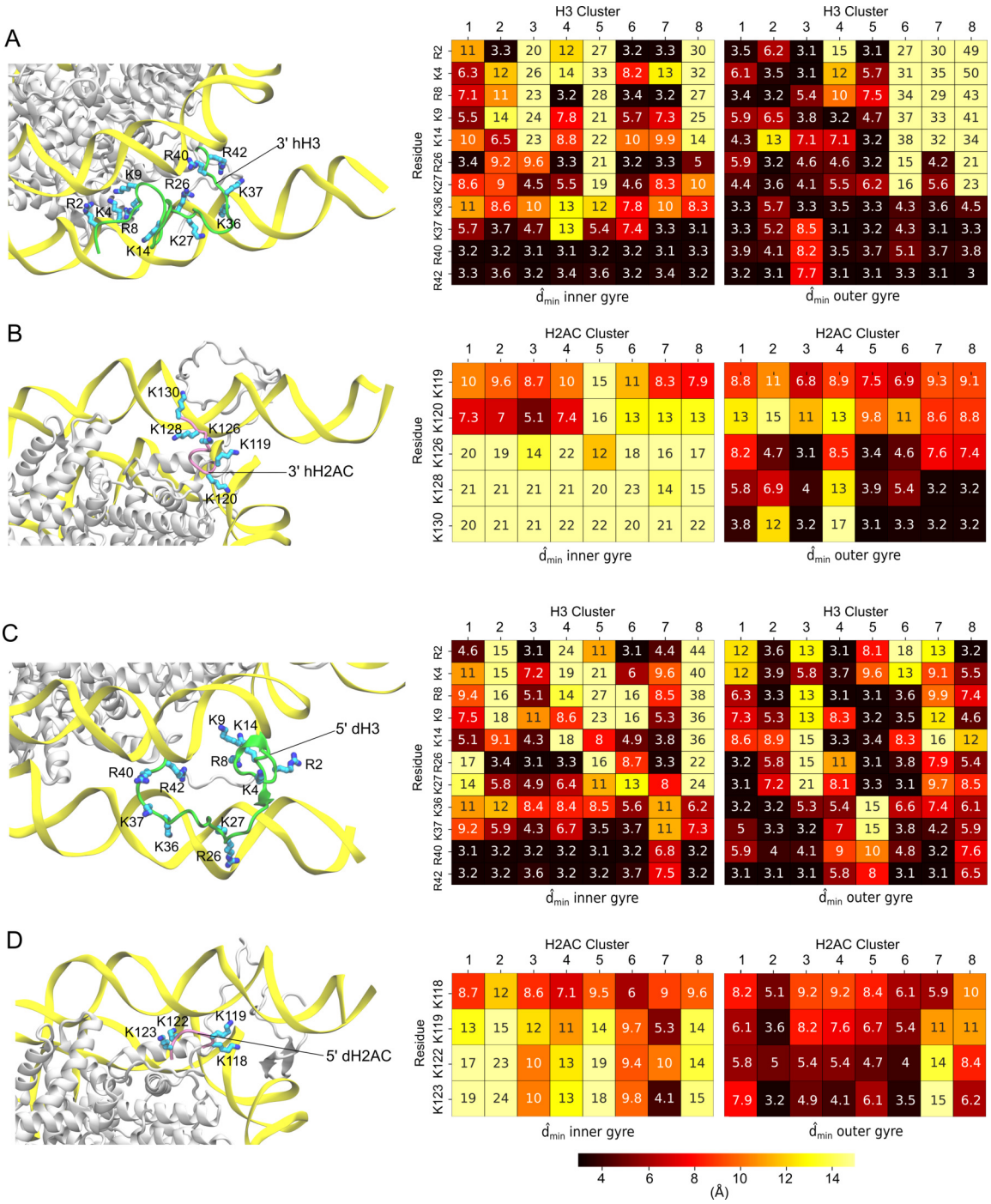


Figure 7: **Interactions of epigenetic regulatory residues from the histone tails with DNA**. Each panel shows a structural illustration of the histone tail (left) and 2 heatmaps with minimal distances of the analyzed residues to the inner (middle) and outer gyre of the DNA (right). **A-B**) the hH simulations. **C-D**) the dH simulations. The H3 (A,C) and H2AC (B,D) tails are shown as green and mauve ribbons with the analyzed aminoacids colored by atom name. The DNA is in yellow ribbons and the core histones are in white. The heatmaps show the analyzed residues on the vertical and the DNA clusters on the horizontal. The numbers are the median value of the minimal distance of each residue to the DNA in the corresponding DNA cluster.

residues in the center groups (R8, K9, K14, R26, K27) and the outer DNA gyre. On the other hand, the residues of the tip and anchor groups lost all their interactions with the DNA while some residues of the anchor group shifted to interact with the inner DNA gyre (K37, R40, R42) (Figure 7C, Supplementary Figure S5). There is no clear correlation between the dynamics of the dH2AC tail - DNA interactions and nucleosome opening (Figure 7D), but even with high fluctuations, we see the residues losing the contacts with the outer gyre before the big opening event in the Lin28b^{dH} second simulation (Supplementary Figure S6).

The hH3 configurations found in nucleosomes open only in one direction (γ_1 or γ_2) (clusters $hH3^1$, $hH3^2$ and $hH3^3$) maintain some of the interactions with the outer gyre of DNA. In the configurations found in cluster $hH3^3$ which correlate with nucleosomes open in the XZ plane (γ_1), the anchor group interacts mainly with the inner gyre of the DNA (Figure 7A). On the other hand, in the configurations found in clusters $hH3^2$, $hH3^4$, and $hH3^5$ which correlate with nucleosomes open in the XY plane (γ_2), the anchor group interacts with the outer gyre of the DNA, whereas some of the residues in the tip and center groups do not interact with DNA (Figure 7A).

A similar pattern is observed for the dH3 configurations found in nucleosomes open only in one direction. Configurations in clusters $dH3^4$ and $dH3^5$ found in nucleosomes open in the XZ plane (γ_1) have no interactions between the anchor group and the outer DNA gyre. On the other hand, in configurations in clusters $dH3^1$ and $dH3^3$ found in nucleosomes open in XY plane (γ_2), residues in the tip and center groups do not interact with the outer gyre whereas residues in the anchor group do (Figure 7C).

Finally, as expected, the H3 and H2AC configurations associated with the most closed nucleosome conformations have the most interactions between the positively charged residues and the outer gyre of the DNA.

Therefore, we conclude that the amplitude of nucleosome breathing motions is inversely proportional with the number of interactions between positively charged residues in the histone H3 and H2AC tails and the nearby L-DNA arm. On the other hand, the residues

involved in these interactions are found in clusters of charged residues, and the absence of interactions involving one particular residue may be compensated by other interactions involving nearby residues from the same cluster. Notably, a large opening of the nucleosome does not necessarily require the loss of all interactions between the tails and the L-DNA.

Discussion

Chromatin fibers are not as compact as it has been long thought.[57] In less compact fibers, the ratio between open and closed nucleosomes will impact on their folding and dynamics. Therefore, deciphering the mechanisms by which single nucleosomes open and close is key to understand chromatin dynamics.

Here we showed how the interplay between the histone H3 and H2AC tails controls the opening and closing of genomic nucleosomes. From a total of 18 µs of atomistic molecular dynamics simulations, we observed 2 large opening events in 2 different genomic nucleosomes and only low amplitude breathing motions in an engineered nucleosome. Our findings suggest that large amplitude opening occurs on shorter time scales in genomic nucleosomes, supporting previous findings[38] that genomic nucleosomes are more flexible and mobile compared to nucleosomes with DNA sequences engineered for strong positioning[39] such as the Widom 601 sequence.

Previously, we observed that the larger structural flexibility observed in the Lin28b nucleosome with *Drosophila* histones was in agreement with nucleosome thermal stability experiments, which revealed that the Lin28b nucleosome disassembled at lower temperatures than the Esrrb and Widom nucleosomes.[38] In the simulations we report here, we also observed a high degree of structural flexibility leading to a large opening event in the other genomic nucleosome, Esrrb, when simulated with human histones. This nucleosome, in our simulations with *Drosophila* histones was remarkably stable. We propose that the differences between the simulations of the Esrrb nucleosome with human and *Drosophila* histones are not due to

the small differences in histone sequences but rather due to the sampling of histone tail dynamics (see discussion below). The large opening for Lin28b occurred only in the simulations with *Drosophila* histones, but the Lin28b nucleosome adopted more open conformations in all our simulations, suggesting that it is indeed the least compact nucleosome. Even if more rigid than Lin28b, the Esrrb nucleosome also disassembled at lower temperatures in experiments than the Widom nucleosome, and it displayed a higher variablity among different replicates of the experiments.[38] Caution is needed when compraing simulations to this type of experiments since, although both nucleosome disassembly in experiments and nucleosome breathing in the simulations are measures of nucleosome flexibility and stability, they are different in amplitude and time scale of the motions involved.

It has been proposed that the time scale for the complete nucleosome unwrapping is in the order of seconds, with intermediate states forming after hundreds of milliseconds.[58,59] While atomistic molecular dynamics simulations are very powerful in revealing the details of nucleosome dynamics, currently the timescales that can be simulated for such a system are in the µs range.[35,36,38,60] This results in simulations that are not converged and achieve only a limited sampling of the conformational space. Tools like enhanced sampling techniques[60,61] can help cope with this limitation, but have difficulties to estimate the energies involved in the nucleosome motions.[60] A powerful approach to achieve a more complete sampling of the conformational space is the grouping of multiple simulation replicas in ensembles like we report here. In each replica, the nucleosome samples a different subregion of the phase space and two replicas are never identical because of the randomizations involved in the velocity assignment at the beginning of the simulations, so merging the replicas expands the phase space explored. Due to the lack of the convergence of individual simulations, it is not surprising that the large opening events we observed were rare. These events are both natural extensions of the low amplitude breathing dynamics of the nucleosomes as demonstrated by the γ_1-γ_2 histograms. Furthermore, the small breathing motions were similarly sampled by the different individual replicas. To this date, the openings we report here have the

largest amplitude among all nucleosome breathing motions reported in atomistic simulations with complete nucleosomes at physiological salt concentrations and temperature. Larger amplitude opening at atomic resolution has only been described in simulations of incomplete nucleosomes from which entire histones were removed.[62]

In addition, the presence of the histone tails in our simulations enabled us to reveal mechanistic details of nucleosome breathing that previous studies have been unable to. Atomistic simulations of tail-less nucleosomes helped understand the histone core and its interactions with DNA, but provided an incomplete description of the breathing and unwrapping of nucleosomes.[24,26–29] We report a regulation mechanism for nucleosomal breathing that involves the histone H3 and H2AC tails and specific interactions between residues in these tails and the DNA. The general behavior of the histone tails H3 and H2AC is similar between *Drosophila* and human histones, despite the differences in sequence observed for H2AC. The tails generally collapse into the DNA, and big changes in the tail conformation and binding happen rarely in our 1µs long simulations.

Histone tails have been previously reported to be important for nucleosome and chromatin dynamics. For example, H3 and H4 tails have a role in genome packing.[6,7] However, the precise mechanisms by which tails impact nucleosome structural flexibility remained obscure. Coarse grained simulations have been used for studying sequence-dependent motions, nucleosome remodeling, and other events that depend on histone tails. Nevertheless, in this type of simulations, the histone tails are simplified and the detailed chemistry of the interactions involved is are often neglected, leading to an incomplete representation of the protein-DNA interactions.[15,18,21,23]

Studying nucleosome structural flexibility with atomistic simulations presents some particular limitations. In addition to the challenges to simulate large amplitude motions of nucleosomes, there are challenges to accurately describe the dynamics of histone tails because of their disordered nature. The simulations of intrinsically disordered proteins has been proven particularly difficult, because the force-fields used for structured proteins often

do not provide a correct sampling of their conformational space.[63] However, despite these limitations the force fields we used represent the state-of-the-art, and have been shown to reproduce the dynamics of intrinsically disordered proteins with reasonable accuracy[63]). In addition, force fields from this family have been used already in the past to understand and reproduce experimental data, such as the destabilization of the nucleosome by mutations on the histone H2A.[64] Moreover, the use of force fields specially developed for such proteins is not an option given the structured nature of the core histones.

The main limitation of the atomistic studies to date is that they have been performed using non genomic nucleosomes, baring artificial sequences such as the Widom or the human α-satellite DNA. This combined with insufficient simulation times may explain why large amplitude opening events similar to the ones we report here were not observed.

We demonstrated here that position of the histone H3 and H2AC tails relative to both the core and the L-DNA controls opening and closing of the DNA. A role for these tails in nucleosome opening has been previously suggested. The deletion of either the H3 or H2AC tails lead to large amplitude nucleosome opening, as shown in experimental studies[8,65,66] and in high temperature simulations.[37] However, the mechanisms by which these tails control the opening was not revealed in these studies. Here, from unbiased simulations, we found that when both tails are present, they have to adopt configurations compatible with nucleosome opening simultaneously. When only one of these tails was in a configuration compatible with DNA opening we observed only partial breathing of the L-DNA in one plane, but no large opening. Our findings suggest that both tails have to move cooperatively to achieve a large amplitude opening of the nuclesome.

The atomic resolution of our simulations enabled us to propose how the interaction between specific residues in the H3 and H2AC tail modulate nucleosome breathing motions. This is of utmost importance, because many of these residues are crucial for epigenetic regulation. Post-translational, chemical modifications of these residues mark chromatin regions as active or inactive. These regions establish which genes are expressed and which are not

at a given time in a given cellular context. Notably, most of these chemical modifications are placed on lysines. Lysines in the H3 tail are located in clusters of positive residues that interact with DNA. We observed that for a large opening of nucleosomes, the majority of these interactions, especially those with the L-DNA have to be absent. In addition, the large opening was a consequence of the absence of the interactions to the L-DNA, suggesting that the nucleosome is kept closed when a majority of these interactions are formed.

The absence of interactions of each individual residue could be compensated by interactions of residues from the same cluster. In case one residue is chemically modified in a certain chromatin context, other positive residues (other lysines or arginines) will still be available for interacting with DNA to keep nucleosomes closed. For example, K9 and K27, both methylated in inactive chromatin (known as heterochromatin) did not establish any specific interaction with DNA independent of the nucleosome conformation in our simulations. However, arginines flanking these lysines displayed a pattern of interactions that were nucleosome conformation dependent. On the other hand K36 is a residue that is acetylated in active chromatin is located in the anchor region between the L-DNA and the inner gyre of the DNA and formed interactions with the DNA in closed but not opened nucleosomes conformations. Nevertheless, K36 is also flanked by 3 other residues in this anchor group, 2 arginines and another lysine which may form interactions with DNA that are redundant with those formed by K36. Therefore, our findings suggest that the epigenetic regulatory residues on histone H3 and H2AC tails may be accessible for chemical modification independent of the nucleosome conformations. The impact of such modifications on intra-nucleosome motions may manifest in more subtle ways than just by neutralizing the positive charge of the tail residues.

Studying the effect of single tail modifications on intra-nucleosome dynamics is challenging experimentally. Atomistic molecular dynamics simulations with modified residues or histone variants such as the ones done by Ikebe et al.[61], Winogradoff et al.[67] or Bowerman and Wereszczynski[36], using complete genomic nucleosomes, represent a powerful alternative.

Conclusions

Here we demonstrated how the interplay between the histone tails modulates the nucleosome conformational dynamics. Large opening of nucleosomes is possible only when the H3 and H2AC tail adopt specific configurations that involve only few contacts with the linker DNA arms of the nucleosome. In addition, we showed that such large openings are more frequent in genomic nucleosomes compared with engineered nucleosomes with DNA sequences optimized for strong positioning such as the Widom sequence. Finally, we harvested the potential of atomistic molecular dynamic simulations to reveal nucleosome and histone tail dynamics and mechanistic insights on how these are interconnected. Our findings are based on unprecedented sampling of nucleosome dynamics in atomistic simulations from which we revealed two large amplitude opening events of two genomic nucleosomes. Future simulations of larger amplitude motions of single or connected nucleosomes on longer timescales will be needed to describe events such as ample nucleosome breathing, nucleosome repositioning or disassembly. From such simulations, the impact of the interplay between histone tails and linker DNA in modulating intra- and inter-nucleosomes interactions as well as how this triggers changes in the chromatin state may be inferred.

Acknowledgement

The authors thank Caitlin M MacCarthy for the support and discussion. J.H. is part of the International Max Planck Research School-Molecular Biomedicine, Münster, Germany. This work was supported by funds of the Max Planck Society and The Royal Netherlands Academy of Arts and Sciences. Computer resources were provided by the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) (project ID 12622, STRUCNUCREC running on the GCS Supercomputer SuperMUC at the Leibniz Supercomputing Centre (www.lrz.de), to J.H. and V.C.).

Supporting Information Available

The following files are available free of charge.

- Huertas_et_al_Supplementary_info.pdf: PDF file containing Supplementary Documents S1 and S2, and Supplementary Figures S1 to S6
- Supplementary_Document_S3.xlsx: Excel file containing all the distances and contacts of the H3 and H2AC tails, over time, in all simulations.

References

(1) Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **1997**, *389*, 251–260.

(2) Richmond, T. J.; Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **2003**, *423*, 145–150.

(3) Bowman, G. D.; Poirier, M. G. Post-translational modifications of histones that influence nucleosome dynamics. *Chemical Reviews* **2015**, *115*, 2274–2295.

(4) Tolsma, T. O.; Hansen, J. C. Post-translational modifications and chromatin dynamics. *Essays in Biochemistry* **2019**, *63*, 89–96.

(5) Zhou, K.; Gaullier, G.; Luger, K. Nucleosome structure and dynamics are coming of age. *Nature Structural and Molecular Biology* **2019**, *26*, 3–13.

(6) Chen, Q.; Yang, R.; Korolev, N.; Liu, C. F.; Nordenskiöld, L. Regulation of Nucleosome Stacking and Chromatin Compaction by the Histone H4 N-Terminal Tail–H2A Acidic Patch Interaction. *Journal of Molecular Biology* **2017**, *429*, 2075–2092.

(7) Zheng, C.; Lu, X.; Hansen, J. C.; Hayes, J. J. Salt-dependent intra- and internucleosomal interactions of the H3 tail domain in a model oligonucleosomal array. *Journal of Biological Chemistry* **2005**, *280*, 33552–33557.

(8) Vogler, C.; Huber, C.; Waldmann, T.; Ettig, R.; Braun, L.; Izzo, A.; Daujat, S.; Chassignet, I.; Lopez-Contreras, A. J.; Fernandez-Capetillo, O.; Dundr, M.; Rippe, K.; Längst, G.; Schneider, R. Histone H2A C-terminus regulates chromatin dynamics, remodeling, and histone H1 binding. *PLoS Genetics* **2010**, *6*, 1–12.

(9) Lachner, M.; O'Carroll, D.; Rea, S.; Mechtler, K.; Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **2001**, *410*, 116–120.

(10) Barski, A.; Cuddapah, S.; Cui, K.; Roh, T. Y.; Schones, D. E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **2007**, *129*, 823–837.

(11) 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.; Aiden, E. L. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **2014**, *159*, 1665–1680.

(12) Shogren-Knaak, M.; Ishii, H.; Sun, J. M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **2006**, *311*, 844–847.

(13) Lehmann, K.; Felekyan, S.; Kühnemuth, R.; Dimura, M.; Tóth, K.; Seidel, C. A. M.; Langowski, J. Dynamics of the nucleosomal histone H3 N-terminal tail revealed by high precision single-molecule FRET. *Nucleic Acids Research* **2020**, *48*, 1551–1571.

(14) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 A resolution. *Journal of Molecular Biology* **2002**, *319*, 1097–1113.

(15) Schlick, T.; Perišić, O. Mesoscale simulations of two nucleosome-repeat length oligonucleosomes. *Physical Chemistry Chemical Physics* **2009**, *11*, 10729.

(16) Collepardo-Guevara, R.; Portella, G.; Vendruscolo, M.; Frenkel, D.; Schlick, T.; Orozco, M. Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: A multiscale computational study. *Journal of the American Chemical Society* **2015**, *137*, 10205–10215.

(17) Freeman, G. S.; Lequieu, J. P.; Hinckley, D. M.; Whitmer, J. K.; De Pablo, J. J. DNA shape dominates sequence affinity in nucleosome formation. *Physical Review Letters* **2014**, *113*, 1–5.

(18) Kenzaki, H.; Takada, S. Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. *PLoS Computational Biology* **2015**, *11*, 1–20.

(19) Niina, T.; Brandani, G. B.; Tan, C.; Takada, S. Sequence-dependent nucleosome sliding in rotation-coupled and uncoupled modes revealed by molecular simulations. *PLoS Computational Biology* **2017**, *13*, 1–22.

(20) Culkin, J.; de Bruin, L.; Tompitak, M.; Phillips, R.; Schiessel, H. The role of DNA sequence in nucleosome breathing. *European Physical Journal E* **2017**, *40*.

(21) Lequieu, J.; Schwartz, D. C.; de Pablo, J. J. In silico evidence for sequence-dependent nucleosome sliding. *Proceedings of the National Academy of Sciences* **2017**, *114*, 201705685.

(22) Brandani, G. B.; Takada, S. Chromatin remodelers couple inchworm motion with twist-defect formation to slide nucleosomal DNA. *PLoS Computational Biology* **2018**, *14*, 1–25.

(23) Zhang, B.; Zheng, W.; Papoian, G. A.; Wolynes, P. G. Exploring the Free Energy Landscape of Nucleosomes. *Journal of the American Chemical Society* **2016**, *138*, 8126–8133.

(24) Ettig, R.; Kepper, N.; Stehr, R.; Wedemann, G.; Rippe, K. Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of dna unwrapping. *Biophysical Journal* **2011**, *101*, 1999–2008.

(25) Winogradoff, D.; Aksimentiev, A. Molecular Mechanism of Spontaneous Nucleosome Unraveling. *Journal of Molecular Biology* **2019**, *431*, 323–335.

(26) Materese, C. K.; Savelyev, A.; Papoian, G. A. Counterion atmosphere and hydration patterns near a nucleosome core particle. *Journal of the American Chemical Society* **2009**, *131*, 15005–15013.

(27) Bishop, T. C. Molecular dynamics simulations of a nucleosome and free dna. *Journal of Biomolecular Structure and Dynamics* **2005**, *22*, 673–685.

(28) Ruscio, J. Z.; Onufriev, A. A computational study of nucleosomal DNA flexibility. *Biophysical Journal* **2006**, *91*, 4121–4132.

(29) Garai, A.; Saurabh, S.; Lansac, Y.; Maiti, P. K. DNA Elasticity from Short DNA to Nucleosomal DNA. *Journal of Physical Chemistry B* **2015**, *119*, 11146–11156.

(30) Potoyan, D. A.; Papoian, G. A. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *Journal of the American Chemical Society* **2011**, *133*, 7405–7415.

(31) Erler, J.; Zhang, R.; Petridis, L.; Cheng, X.; Smith, J. C.; Langowski, J. The Role of Histone Tails in the Nucleosome: A Computational Study. *Biophysical Journal* **2014**, *107*, 2911–2922.

(32) Roccatano, D.; Barthel, A.; Zacharias, M. Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation. *Biopolymers* **2007**, *85*, 407–421.

(33) Biswas, M.; Voltz, K.; Smith, J. C.; Langowski, J. Role of histone tails in structural stability of the nucleosome. *PLoS Computational Biology* **2011**, *7*, e1002279.

(34) Chakraborty, K.; Kang, M.; Loverde, S. M. Molecular Mechanism for the Role of the H2A and H2B Histone Tails in Nucleosome Repositioning. *Journal of Physical Chemistry B* **2018**, *122*, 11827–11840.

(35) Shaytan, A. K.; Armeev, G. A.; Goncearenco, A.; Zhurkin, V. B.; Landsman, D.; Panchenko, A. R. Coupling between Histone Conformations and DNA Geometry in Nucleosomes on a Microsecond Timescale: Atomistic Insights into Nucleosome Functions. *Journal of Molecular Biology* **2016**, *428*, 221–237.

(36) Bowerman, S.; Wereszczynski, J. Effects of MacroH2A and H2A.Z on Nucleosome Dynamics as Elucidated by Molecular Dynamics Simulations. *Biophysical Journal* **2016**, *110*, 327–337.

(37) Li, Z.; Kono, H. Investigating the Influence of Arginine Dimethylation on Nucleosome Dynamics Using All-Atom Simulations and Kinetic Analysis. *The Journal of Physical Chemistry B* **2018**, *122*, 9625–9634.

(38) Huertas, J.; MacCarthy, C. M.; Schöler, H. R.; Cojocaru, V. Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding. *Biophysical Journal* **2020**, 634386.

(39) Lowary, P. T.; Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology* **1998**, *276*, 19–42.

(40) Vasudevan, D.; Chua, E. Y. D.; Davey, C. A. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *Journal of molecular biology* **2010**, *403*, 1–10.

(41) Chua, E. Y.; Vasudevan, D.; Davey, G. E.; Wu, B.; Davey, C. A. The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Research* **2012**, *40*, 6338–6352.

(42) Tóth, K.; Böhm, V.; Sellmann, C.; Danner, M.; Hanne, J.; Berg, M.; Barz, I.; Gansen, A.; Langowski, J. Histone- and DNA sequence-dependent stability of nucleosomes studied by single-pair FRET. *Cytometry Part A* **2013**, *83*, 839–846.

(43) Adachi, K.; Kopp, W.; Wu, G.; Heising, S.; Greber, B.; Stehling, M.; Araúzo-Bravo, M. J.; Boerno, S. T.; Timmermann, B.; Vingron, M.; Schöler, H. R. Esrrb Unlocks Silenced Enhancers for Reprogramming to Naive Pluripotency. *Cell Stem Cell* **2018**, *23*, 266–275.e6.

(44) Zhang, J. et al. LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency. *Cell Stem Cell* **2016**, *19*, 66–80.

(45) Soufi, A.; Donahue, G.; Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **2012**, *151*, 994–1004.

(46) Soufi, A.; Garcia, M. F.; Jaroszewicz, A.; Osman, N.; Pellegrini, M.; Zaret, K. S. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **2015**, *161*, 555–568.

(47) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005**, *26*, 1668–88.

(48) Jerabek, S.; Ng, C. K. L.; Wu, G.; Arauzo-Bravo, M. J.; Kim, K.; Esch, D.; Malik, V.; Chen, Y.; Velychko, S.; MacCarthy, C. M.; Yang, X.; Cojocaru, V.; Schöler, H. R.; Jauch, R. Changing POU dimerization preferences converts Oct6 into a pluripotency inducer. *EMBO reports* **2017**, *18*, e201642958.

(49) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **2005**, *26*, 1781–802.

(50) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.

(51) Ivani, I. et al. Parmbsc1: A refined force field for DNA simulations. *Nature Methods* **2015**, *13*, 55–58.

(52) Li, P.; Song, L. F.; Merz, K. M. Systematic parameterization of monovalent ions employing the nonbonded model. *Journal of Chemical Theory and Computation* **2015**, *11*, 1645–1657.

(53) Öztürk, M. A.; Pachov, G. V.; Wade, R. C.; Cojocaru, V. Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome. *Nucleic Acids Research* **2016**, *44*, 6599–6613.

(54) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084–3095.

(55) Fiorin, G.; Klein, M. L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Molecular Physics* **2013**, *111*, 3345–3362.

(56) Humphrey, W.; Dalke, A.; Schulten, K. Visual Molecular Dyanmics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(57) Ou, H. D.; Phan, S.; Deerinck, T. J.; Thor, A.; Ellisman, M. H.; O'Shea, C. C. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **2017**, *357*.

(58) Li, G.; Levitus, M.; Bustamante, C.; Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. *Nature Structural and Molecular Biology* **2005**, *12*, 46–53.

(59) Chen, Y.; Tokuda, J. M.; Topping, T.; Meisburger, S. P.; Pabit, S. A.; Gloss, L. M.; Pollack, L. Asymmetric unwrapping of nucleosomal DNA propagates asymmetric opening and dissociation of the histone core. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, 334–339.

(60) Kono, H.; Sakuraba, S.; Ishida, H. Free energy profiles for unwrapping the outer superhelical turn of nucleosomal DNA. *PLOS Computational Biology* **2018**, *14*, e1006024.

(61) Ikebe, J.; Sakuraba, S.; Kono, H. H3 Histone Tail Conformation within the Nucleosome and the Impact of K14 Acetylation Studied Using Enhanced Sampling Simulation. *PLoS Computational Biology* **2016**, *12*, 1–13.

(62) Kameda, T.; Awazu, A.; Togashi, Y. Histone Tail Dynamics in Partially Disassembled Nucleosomes During Chromatin Remodeling. *Frontiers in Molecular Biosciences* **2019**, *6*, 1–10.

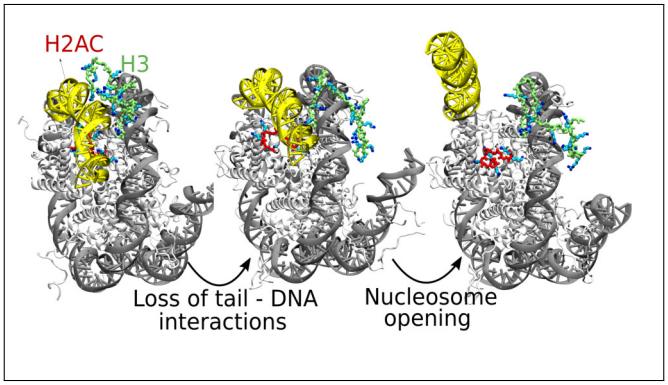
(63) Duong, V. T.; Chen, Z.; Thapa, M. T.; Luo, R. Computational Studies of Intrinsically Disordered Proteins. *Journal of Physical Chemistry B* **2018**, *122*.

(64) Lehmann, K.; Zhang, R.; Schwarz, N.; Gansen, A.; Mücke, N.; Langowski, J.; Toth, K. Effects of charge-modifying mutations in histone H2A α3-domain on nucleosome sta-

bility assessed by single-pair FRET and MD simulations. *Scientific Reports* **2017**, *7*, 13303.

(65) Nurse, N. P.; Jimenez-Useche, I.; Smith, I. T.; Yuan, C. Clipping of flexible tails of histones H3 and H4 affects the structure and dynamics of the nucleosome. *Biophysical Journal* **2013**, *104*, 1081–1088.

(66) Andresen, K.; Jimenez-Useche, I.; Howell, S. C.; Yuan, C.; Qiu, X. Solution scattering and FRET studies on nucleosomes reveal DNA unwrapping effects of H3 and H4 tail removal. *PLoS ONE* **2013**, *8*, 1–9.

(67) Winogradoff, D.; Echeverria, I.; Potoyan, D. A.; Papoian, G. A. The acetylation landscape of the H4 histone tail: Disentangling the interplay between the specific and cumulative effects. *Journal of the American Chemical Society* **2015**, *137*, 6245–6253.

Graphical TOC Entry