Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

Biophysical *Journal*
Article

Biophysical Society

Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding

Jan Huertas,[1,2,3] Caitlin M. MacCarthy,[2] Hans R. Schöler,[2,4] and Vlad Cojocaru[1,2,3,*]

[1]In Silico Biomolecular Structure and Dynamics Group, Hubrecht Institute, Utrecht, the Netherlands; [2]Department of Cellular and Developmental Biology, Max Planck Institute for Molecular Biomedicine, Münster, Germany; [3]Center for Multiscale Theory and Computation, Westfälische Wilhelms University, Münster, Germany; and; and [4]Medical Faculty, University of Münster, Münster, Germany

ABSTRACT Transcription factor (TF) proteins bind to DNA to regulate gene expression. Normally, accessibility to DNA is required for their function. However, in the nucleus, the DNA is often inaccessible, wrapped around histone proteins in nucleosomes forming the chromatin. Pioneer TFs are thought to induce chromatin opening by recognizing their DNA binding sites on nucleosomes. For example, Oct4, a master regulator and inducer of stem cell pluripotency, binds to DNA in nucleosomes in a sequence-specific manner. Here, we reveal the structural dynamics of nucleosomes that mediate Oct4 binding from molecular dynamics simulations. Nucleosome flexibility and the amplitude of nucleosome motions such as breathing and twisting are enhanced in nucleosomes with multiple TF binding sites. Moreover, the regions around the binding sites display higher local structural flexibility. Probing different structures of Oct4-nucleosome complexes, we show that alternative configurations in which Oct4 recognizes partial binding sites display stable TF-DNA interactions similar to those observed in complexes with free DNA and compatible with the DNA curvature and DNA-histone interactions. Therefore, we propose a structural basis for nucleosome recognition by a pioneer TF that is essential for understanding how chromatin is unraveled during cell fate conversions.

SIGNIFICANCE Pioneer transcription factors (TFs) are proteins that bind to DNA in closed chromatin and mediate chromatin opening during cell fate conversions. In closed chromatin, the DNA is wrapped around histone proteins in nucleosomes, which partially occlude the DNA features required for protein binding. Nevertheless, how pioneer factors recognize DNA on nucleosomes remains unknown. Here, we describe how structural dynamics mediate the binding of Oct4, a pioneer TF, to the nucleosome and propose a structural basis for the Oct4-nucleosome interaction. Our findings reveal how pioneers such as Oct4 are able to engage wrapped DNA. Moreover, we demonstrate that molecular simulations can be used as a tool to screen for configurations of TFs that are compatible with nucleosome binding.

INTRODUCTION

Transcription factors (TFs) are proteins that directly or indirectly bind to DNA to promote transcription. In most cases, accessibility to DNA is a prerequisite for TF binding. However, in the nucleus, the DNA is packed into chromatin, often making it inaccessible for TF binding. The fundamental unit of chromatin is the nucleosome, which is formed by wrapping 146 DNA basepairs (bps) 1.65 times around the histone octamer, a group of four DNA binding proteins (each present twice) that have a globular domain and a disordered tail (1). Packing DNA in this way leaves it partially occluded on the histone interface and by the second gyre of DNA.

Interestingly, a series of TFs are able to bind to closed, inactive chromatin, i.e., they recognize their binding sites even in the presence of histones. These factors, known as pioneer TFs, can help open chromatin, increase DNA accessibility, and support binding of other TFs. Among the TFs shown to bind to closed chromatin, several regulate transitions between cell fates (2).

Particularly relevant for inducing pluripotency are Oct4, Sox2, and Klf4. These three TFs together are known for their capacity to convert differentiated cells into pluripotent cells, a process known as reprogramming (3), and have been characterized as pioneer TFs (4,5). Structurally, Oct4 has two DNA binding domains (the POU (name derived from the Pituary-specific TF Pit1, Octamer binding TFs Oct1 and Oct2, and neural TF Unc-86) specific domain, POU_S, and

Submitted September 27, 2019, and accepted for publication December 23, 2019.

*Correspondence: v.cojocaru@hubrecht.eu

Editor: Tamar Schlick.

https://doi.org/10.1016/j.bpj.2019.12.038

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

the POU homeodomain, POU_{HD}) that together recognize an octamer sequence in free DNA (6). In free DNA, Oct4 cooperates with Sox2 to recognize composite motifs (7) via mechanisms involving structural dynamics that we previously characterized (8–10). However, the interaction of these TFs with nucleosomes remains obscure. A recent survey of TF-nucleosome binding properties was performed for many different classes of TFs, revealing how the nucleosomal context modifies the orientation and position of the TF binding compared to free DNA (11). In the case of Oct4, the canonical configuration known from the structures of Oct4 bound to free DNA (6) is incompatible with nucleosome binding because it involves binding of the POU_S and POU_{HD} on opposite sides of DNA. A partial motif recognition model has been proposed, in which only one of the domains is bound sequence specifically to the nucleosome (5).

Although nucleosomes can assemble on just about any sequence, certain DNA sequences have structural features that lead to a higher affinity for the histone core proteins. For example, stiff DNA and polyA or T tracts are unfavorable for nucleosome formation, but bendable DNA as well as DNA with AA/TT or GC dinucleotides in 10 bp periodicity prove favorable (12). Accordingly, DNA sequence also influences nucleosome stability and mobility (13). However, how the DNA sequence modulates nucleosome dynamics (e.g., local flexibility, compaction, breathing, twisting) at atomic resolution is not understood.

Currently, atomic-resolution structures of a TF-nucleosome complex do not exist. Moreover, the interplay between TF binding and sequence-dependent nucleosome dynamics is not understood. This prevents rational design of pioneer TFs for improving the efficiency and, perhaps, the direction of cell fate transitions. Therefore, it is of utmost importance to elucidate the features of nucleosomes that facilitate TF binding and to reveal how TF binding modulates nucleosome dynamics.

All-atom molecular dynamics (MD) simulations provide a powerful approach to study nucleosome dynamics. By solving Newton's equations of motion, the motions and interactions of biomolecules at atomic resolution can be studied. Nucleosome dynamics have been simulated (14–23), and motions such as the opening and closing of the linker DNA (L-DNA) in tail-less nucleosomes (24), the twisting involved in nucleosome repositioning (19), and the behavior of the L-DNAs in the presence of linker histones (25) have been described.

Nevertheless, almost all of these previous atomic-resolution MD simulations (26) were started from the most complete crystal structure of the nucleosome (27) using a palindromic DNA sequence derived from human α-satellite DNA. Interestingly, coarse-grained MD simulations in which several nonhydrogen atoms are grouped in beads have been performed with different DNA sequences (28–32). Many of these studies used the Widom 601 artificial sequence (29,33), an artificial sequence selected for its stability (34) that is present in many crystal structures of the nucleosome (35,36). Because of the reduced representation, the computational cost of these simulations is much lower. Therefore, questions that require longer and a higher number of simulations can be addressed. As a result, models of sequence-specific nucleosome dynamics revealing the role of the sequence in nucleosome breathing and unwrapping (29,37) and protein-mediated remodeling (38) have been reported. However, the descriptions obtained from coarse-grained simulations are limited because important details regarding the interactions that modulate nucleosome dynamics are missing. For example, the hydrogen bonds driving the interaction of histone tails with the DNA grooves and the fine structural properties of DNA modulating TF binding could not be studied. Therefore, atomic-resolution methods are indispensable for revealing the interplay between TF binding and nucleosome dynamics.

To study the Oct4-nucleosome interaction, we selected two native nucleosome sequences shown to be bound by Oct4 by Soufi et al. during reprogramming (4). The sequences were discovered by overlapping data sets of two types of sequencing experiments: the Oct4 binding sites from chromatin immunoprecipitation (ChIP) with the nucleosome positioning from micrococcal nuclease (MNase) digestion. We selected these sequences for their role in pluripotent cells, clarity of nucleosome boundaries, and number of TF binding sites. They are found in the enhancers of the *ESRRB* and *LIN28B* genes. Esrrb is a TF that regulates cellular pluripotency (39), whereas Lin28b is an RNA-binding protein expressed in pluripotent stem cells and involved in development and maintenance of the pluripotent state (40). The Esrrb sequence has just one obvious POU_S binding site, whereas Lin28b has at least two sites that Oct4 can interact with, one for the POU_S and one for the POU_{HD}. The Oct4-Lin28b binding strength and position were established by electrophoretic mobility shift assays and footprinting (5).

We first compared the dynamics on the microsecond timescale of the two native nucleosomes with those of an artificial nucleosome with the Widom 601 sequence from atomic-resolution MD simulations. This nucleosome provides a good reference for studying the dynamics of nucleosomes with native sequences because of its strong positioning that leads to limited structural flexibility. Moreover, the dynamics of this nucleosome has been studied before (36). We found differences in the nucleosome flexibility, as well as in the amplitude of different types of motions such as nucleosomal breathing, twisting, or enhanced fluctuations among the nucleosomes, that correlated with the number of TF binding sites. We validated these findings using temperature-induced nucleosome disassembly. Then, we built models of Oct4 bound to mononucleosomes and showed that some noncanonical configurations, but not all tested configurations, of Oct4-nucleosome complexes were stable and not affected by DNA curvature or DNA-histone interactions. Our findings describe nucleosome dynamics that can facilitate Oct4-nucleosome binding and provide a

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

model for the structural basis of Oct4-nucleosome recognition.

MATERIALS AND METHODS

Nucleosome modeling

The 146 bp Widom *Drosophila melanogaster* nucleosome was built by homology modeling, using as templates the structures of the *D. melanogaster* nucleosome core (Protein Data Bank, PDB: 2PYO), the structure of the Widom 601 nucleosome core particle (PDB: 3LZ0) and the high-resolution structure of the nucleosome core particle including tails (PDB: 1KX5). Using Modeller (https://salilab.org/modeller/), 100 homology models were generated using a "slow" optimization protocol, followed by a "slow" MD-based refinement protocol. The model with the lowest discrete optimized protein energy was then selected. Then, fragments of B-DNA generated with NAB (Nucleic Acid Builder available in Ambertools 18 (41)) were added to each end to generate the 168 bp nucleosome.

The Esrrb and Lin28b enhancer sequences with different Oct4 binding profiles were selected from data by Soufi et al. (4). The Lin28b sequence was selected as a representative of a nucleosome with multiple TF binding sites, whereas the Esrrb sequence was chosen for having a single binding site of Oct4. To model these natural nucleosomes, we used the MNase data that were used by Soufi et al. to superimpose on Oct4 binding data obtained from ChiP sequencing (ChIP-Seq) (42). We substituted the Widom sequence with the Lin28b and the Esrrb sequences using the "swapna" function in Chimera (https://www.cgl.ucsf.edu/chimera/) upon fixing the center of the MNase peak to be the dyad point. Because the MNase data did not have single bp resolution, the Lin28b sequence was shifted by one bp to obtain a conformation with a better exposure of the Oct4 binding sites.

Modeling of the Oct4-nucleosome complex

We used the following structures to model Oct4 bound to nucleosomes: 1) Oct4 bound in the canonical configuration as found in the structure of the homodimer bound to the palindromic PORE motif (6), 2) Oct4 bound as a homodimer to the partially palindromic MORE motif, and 3) free Oct4 from MD simulations. The canonical configuration of Oct4 was obtained by modeling the human Oct4, using as a template the final snapshot from the previously published trajectories of Oct4 bound to free DNA (9). The MORE configuration was obtained by stripping one monomer from the first snapshot after equilibration of the published trajectories of Oct4 bound as a homodimer in the MORE configuration (10). Finally, the MD configurations of Oct4 were obtained after 100 ns of MD simulations of Oct4, starting from the canonical configuration upon stripping the DNA (see details of the simulation in the next section of Materials and Methods).

To model the binding of Oct4 to the nucleosome, we assumed that the interactions between Oct4 residues and DNA bases (direct readout) will be the same as in the free DNA. This assumption was based on ChIP-Seq data that revealed the same recognition motif for Oct4 on free and wrapped DNA. First, we superimposed the Oct4 binding sites from the free and nucleosomal DNA, using only bases that form direct readout interactions with Oct4. Then, we removed the free DNA from the resulting model. For the MD configurations, we used the hypothetical Oct4-nucleosome model with Oct4 in the canonical configuration as template on which we superimposed the selected MD configuration. Snapshots with no clashes between Oct4 and the histones or the DNA were selected for further MD simulations.

Molecular dynamics simulations

The procedure below was applied for the simulations of free Oct4 alone, free nucleosomes, and Oct4-nucleosome complexes. Each species was solvated in a truncated octahedron box of TIP3P water molecules, with a layer of at least 12 Å of water around the solute in any direction. Na^+ ions were added to counter the negative charges. KCl was added to the system up to a concentration of 150 mM. The solvated systems were then optimized by energy minimization with the AMBER software (41). After that, the systems were equilibrated in NAMD (43) for 10.25 ns using restraints that were removed in a stepwise manner, using a protocol described in detail in our previous simulations of protein-DNA systems (10). After equilibration, the system was simulated using NAMD in the isobaric-isothermic (NPT, $p = 1$ atm, T = 300 K) ensemble, using a standard protocol with Langevin dynamics to control the temperature and Nosé-Hoover and Langevin piston to control pressure. The Merz ion parameterization and the ff14SB (44) and the parmbsc1 force fields (45) were used for ions, protein, and DNA, respectively. Oct4 alone was simulated 5×100 ns to obtain conformations compatible with nucleosome binding. Each nucleosome was simulated for 3×1 μs, whereas the Oct4-nucleosome complexes were submitted to three independent equilibrations, and one equilibrated structure for every configuration was simulated for 100 ns.

Analysis of nucleosome dynamics

The MD trajectories were processed by imaging and performing a root mean-square fit of the heavy atoms of the histone core (excluding the histone tails: residues 1–45 for H3, 1–32 for H4, 1–17 and 115–124 for H2A, and 1–31 for H2B) to the minimized structure of the Widom nucleosome. Atomic fluctuations of the heavy atoms of the DNA were calculated and averaged by residue using cpptraj (46)The fluctuations were squared and weight by $(8/3)\pi^2$ to obtain a value similar to a crystallographic B-factor. To facilitate the comparison, the average value of atomic fluctuations of the DNA core (146 bp) was calculated, and all B-factors were normalized by that factor. Then, DNA residues were colored using VMD (47), with blue and red indicating normalized B-factors smaller than 1 or bigger than 1, respectively.

We calculated the elastic deformation energy $E(X) = \sum_i K_i \Delta X_i^2$ (kcal/mol) of the minimized average structure of nucleosomal DNA from each ensemble of simulations as implemented in the MCDNA tool (https://mmb.irbbarcelona.org/MCDNA). For each bp step in its unique tetrameric environment (i), a 6×6 stiffnes matrix K_i was derived from the ABC microsecond-long MD simulations (48) with the parmbsc1 force field (45). K_i is the inverted covariance matrix in helical space, which was calculated using six DNA helical parameters (twist, roll, tilt, rise, slide, and shift). $\Delta X_i = X_i - X_i^0$, where X_i is the vector with the helical parameters as components calculated from the average nucleosomal DNA structure and X_i^0 is the same vector obtained using the average values of the helical parameters from the ABC simulations.

Distance root mean-square deviations (dRMSDs), the RMSD matrices, and the radius of gyration were also calculated using cpptraj. In the fitted trajectories, the dRMSD and the mass-weighted radius of gyration were obtained for the heavy atoms of the DNA. Calculations were done for DNA fragments of 168, 146, 138, and 120 bp. To obtain the fragments smaller than 168 bp, we excluded 11, 15, and 24 basepairs from each L-DNA from the analysis. The average structure of the DNA backbone of all three 3 μs ensembles was used as reference to prevent bias coming from the reference structure. Similarly, the RMSD matrices were obtained by calculating the RMSD of the heavy atoms of the DNA between each of the two snapshots for the different DNA fragments. To characterize the regions of local structural flexibility, the DNA was fragmented in regions of 8 bp each, starting from the dyad, and the dRMSD was calculated for each fragment, again using the average of the DNA backbone of all three ensembles as reference.

To characterize the breathing motions of the nucleosomes, we applied the procedure described by Öztürk et al. (25). First, we defined a coordinate system XYZ with the origin on the dyad. X was defined along the dyad axis, Y as the cross products between a vector v_1 defined along the dyad axis and a vector v_2 defined to be approximately perpendicular to v_1 and to intersect v_1 as close as possible to the center of the nucleosome, and Z

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

as the cross product between *X* and *Y*. Then, the angles γ_1 and γ_2 were defined as the angle between the projection of the vector v_N defined along the helical axis of the linker DNA on the *XZ* plane and the *Z* axis, and γ_2 as the angle formed between the projection of v_N on the *XY* plane and the *Y* axis.

Principal component analysis (PCA) was performed in cpptraj using the nonhydrogen atoms of the DNA and the rigid part of the histones (excluding the histone tails). The covariance matrix was calculated and diagonalized to extract the first 25 eigenvectors and eigenvalues. Then, the trajectory was projected on the first five modes, and the minimum and maximum projection values for each mode were extracted. Finally, pseudotrajectories along each mode were generated to analyze the motions of the system. The first five modes represent 84.27, 78.19, and 87.02% of all the motion of the Widom, Esrrb, and Lin28b nucleosomes, respectively.

To quantify the twisting motions of the nucleosomes, and distinguish them from other motions, a bp-centered method was developed (Fig. S7 *A*). For every bp, the *X* axis was defined by a vector along the line that connects the N1 atom of the pyrimidine with the N9 of the purine. Then, a vector connecting the geometrical center of the C_2 atoms of both bases and the geometrical center of the C_4 atoms of both bases was defined. *Z* was set along the cross product between this vector and *X*. Then, *Y* is the cross product between *X* and *Z*. The origin O of the coordinate system was defined at the intersection between this vector and *X*, which is near the bp center. The coordinate system was updated in each snapshot, the updated origin being O′ and the updated axes *X*′, *Y*′, and *Z*′. Finally, we calculated the angle formed between *Y* and *Y*′ for each bp and the displacement *d* as the distance between O and O′.

Analysis of Oct4-nucleosome complexes

To characterize the stability of the Oct4-nucleosome complexes, we used the last, unrestrained step of the equilibrations (5 ns) and the production runs (100 ns). The binding of Oct4 was evaluated using two criteria. On one hand, the number of stable contacts between Oct4 and the DNA (defined as the number of Oct4 nonhydrogen atoms closer than 4.5 Å to the DNA that are maintained for at least 75% of the simulated time) was determined using cpptraj. As direct and indirect readout interactions, we considered the contacts with the DNA bases and backbone, respectively. On the other hand, we monitored the specific hydrogen bonds formed by Q44 of the POU_S and N143 of POU_{HD} with adenine bases. These are direct readout interactions selected based on the Oct4-DNA structure (6,49). The percentage of simulation time in which the donor-acceptor distance was smaller than 4 Å was calculated using VMD.

Mononucleosome preparation

Full-length *D. melanogaster* histones were individually purified and refolded in a ratio of 1:1:1:1 according to the protocol published by Luger et al. (50). Unlabeled, high-concentration nucleosomes were assembled using the salt gradient dialysis method described in the same publication. Assembled nucleosomes were purified from free DNA and free histones using a 10–30% sucrose gradient spun at 30,000 × *g* and 4°C in a Beckman Coulter Optima L-100 XP swing bucket rotor (SW-41; Brea, CA) for 18 h. Fractions were collected from the bottom of the gradient, screened on native gels, pooled, and quantitated by densitometry using a DNA standard curve and Quantity One software (Bio-Rad, Hercules, CA).

Thermal stability assay

Nucleosome thermal stability was assessed according to the protocol outlined in Taguchi et al. (51). Briefly, nucleosomes were diluted to 2.5 μM in low salt buffer (25 mM Hepes (pH 7.8), 50 mM NaCl, 0.5 mM EDTA, and 10% glycerol) along with SYPRO® Orange diluted 1:1000 (Cat. #S5692; Sigma-Aldrich, St. Louis, MO) in a 20 μL final reaction volume. Reaction components were carefully mixed, briefly spun down, and immediately measured. Samples were analyzed in triplicate over the temperature range 25–95°C in a 96-well microplate using the TAMRA detection settings on an ABI 7900HT (Applied Biosystems, Foster City, CA). Wells containing water or 2.5 μM bovine serum albumin standard (Cat. # 5000206; Bio-Rad) were included as controls. Raw data were collected from the channel with the maximum signal among all the samples, normalized, and replotted to find the T_{max}. In Table 1, we present the average and standard deviation of eight replicates of this experiment. First, we performed a set of three technical replicates (same nucleosome assembly, different wells), and then an additional set of five technical replicates on a different date from a different nucleosome assembly.

RESULTS

The nucleosome with multiple transcription factor binding sites is more dynamic

We performed three 1 μs MD simulations for each of the Widom-, Esrrb-, and Lin28b-containing nucleosomes. Given that 1 μs is insufficient for convergent sampling of the conformational space of the nucleosome, we combined three individual simulations of each nucleosome into one ensemble and performed all analyses on the three ensembles obtained.

To distinguish between the two L-DNAs, we used the genomic 5′-3′ orientation and referred to them as 5′ L-DNA and 3′ L-DNA. For the Widom nucleosome, we defined the 5′ and 3′ ends as the ends labeled with negative and positive residue numbers respectively in the structure PDB: 3LZ0 (35). To characterize the nucleosome flexibility, we first fitted the structured regions of the core histones (excluding the histone tails; see Materials and Methods) from each snapshot on the starting model of the Widom nucleosome and calculated the atomic fluctuations of the DNA residues (Fig. 1, *A–C*). Then, we converted the fluctuations in β-factor analogs to crystallographic β-factors and normalized the β-factors by the mean β-value of the core DNA (see Materials and Methods). We found that the most mobile regions are the L-DNAs. Remarkably, their motions propagate inside the core nucleosomal DNA (the 146 bp that define the nucleosome particle) and were asymmetric. The 5′ L-DNA displayed higher flexibility in regions located deeper in the core DNA. Moreover, we detected

TABLE 1 Overview of the Simulated Nucleosomes

Sequence	Origin	Oct4 Binding Sites	AT Content (%)	Simulation Time	Temperature of Dissociation	Elastic Distortion Energies[a] (kcal/mol)
Widom	Artificial	0	42.26	3 × 1 μs	88.51 ± 1.60	322.93
Esrrb	Natural	1	45.24	3 × 1 μs	87.24 ± 1.49	364.92
Lin28b	Natural	>1	59.52	3 × 1 μs	83.90 ± 1.08	366.53

[a]Energies calculated from the average structure of the MD simulation ensembles.

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

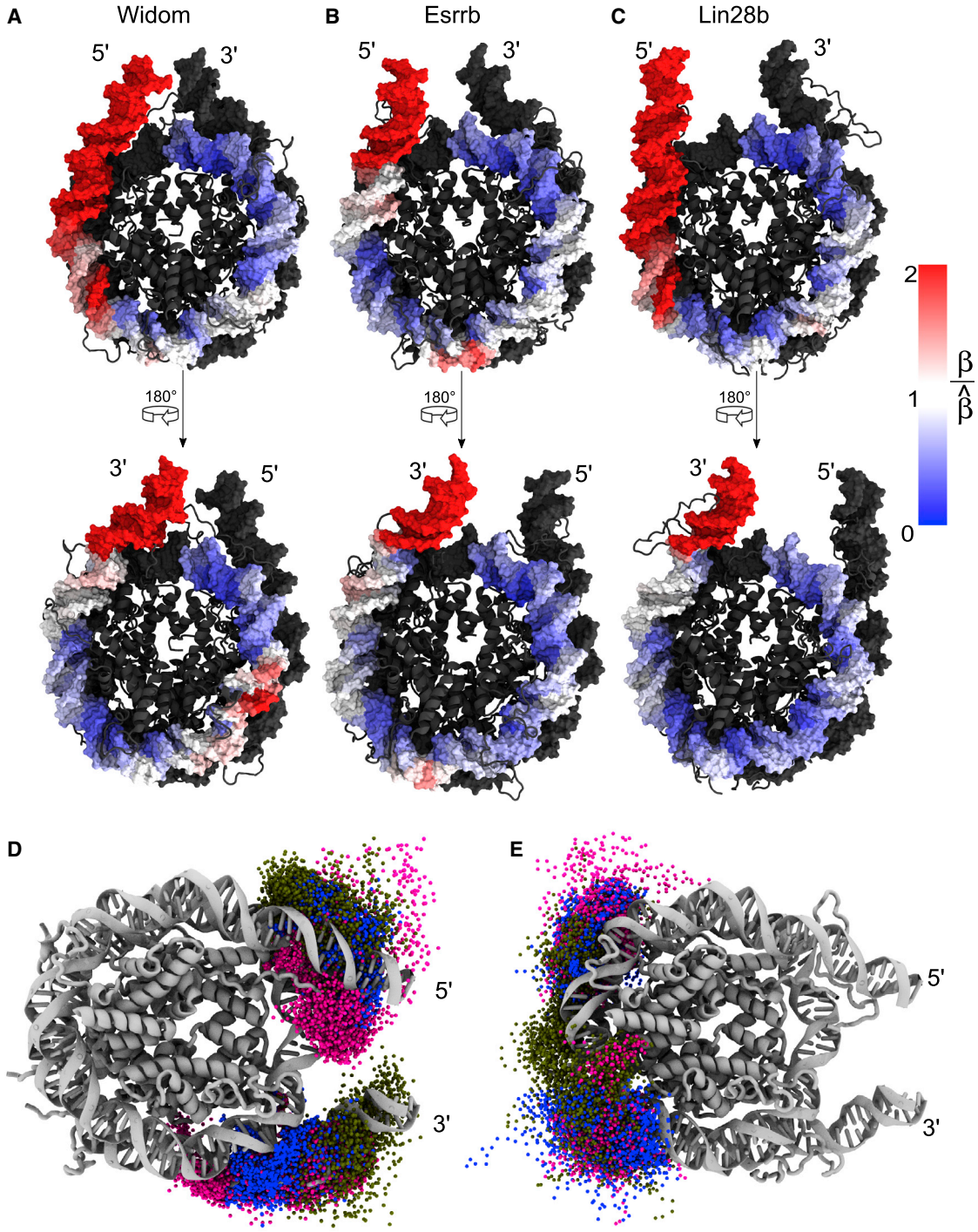


FIGURE 1 Asymmetric nucleosome flexibility. (*A–C*) Normalized, relative atomic fluctuations of the nonhydrogen DNA atoms calculated from the simulation ensembles of the three nucleosomes are shown: (*A*) Widom, (*B*) Esrrb, and (*C*) Lin28b. First, the absolute atomic fluctuations were calculated, and the average (mass weighted) per residue was then squared and weighted by $8/3\pi^2$ to obtain factor β analogs to crystallographic B-factors. The nucleosomes were colored by $\frac{\beta}{\widehat{\beta}}$, where $\widehat{\beta}$ is the mean β of the core nucleosomal DNA (146 bp). The residues with fluctuations smaller and larger than the average fluctuations of the core DNA are in blue and red, respectively, with darkest red corresponding to $\frac{\beta}{\widehat{\beta}} = 2$ to avoid the masking of fluctuations by the very high values found in the L-DNA. (*D–E*) Phase space sampling of the core histone tails during the simulation ensembles is shown. The positions of all C_α atoms of the H3 (*D*) and H2B (*E*) tails are shown at every 20 ns. The colors indicate the three nucleosomes—Widom (*magenta*), Esrrb (*blue*), and Lin28b (*green*)—and are maintained throughout this article. To see this figure in color, go online.

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

small enhanced flexibility patches inside the core DNA, independent of the L-DNA motions.

The structural flexibility of the L-DNA arms and its asymmetry could be partly attributed to the differences in the conformational sampling of the unstructured tails of the core histones (Figs. 1, *D* and *E* and S1). In particular, the H3 tail involved in a large number of epigenetic modifications sampled significantly different regions of the phase space in the three nucleosomes (Fig. 1 *D*), accounting partly for the motions involving the L-DNAs and their asymmetry. Moreover, in the areas of increased local flexibility inside the core DNA, we observed fewer contacts with the tails of the core histones located near the area (Figs. 1 *E* and S1). For example, the H2B tails located on the opposite side of the dyad sample different regions of the phase space in the simulations of the Widom and Esrrb nucleosomes comparing to Lin28b, explaining the higher flexibility of that region in the two nucleosomes.

To study the differences in the structural flexibility of the three nucleosomes, we calculated the dRMSD of the DNA backbone (Fig. 2 *A*). The dRMSD measures the deviation of all internal interatomic distances from their values in a reference structure and is independent of the initial fitting of the trajectory snapshots to the reference structure. The dRMSD values indicate how far different structures deviate from the reference, whereas their spread (width of the dRMSD histograms) is a measure of intrinsic flexibility. We used the average structure of the DNA backbone from all simulations as the reference, ensuring that no reference dependent bias in the dRMSD values was introduced.

To correlate the nucleosome flexibility with the compaction of the nucleosomes, we also calculated the radius of gyration of the DNA. Higher radius of gyration values reflect a decrease in nucleosome compaction. We found a hierarchy of dynamics among the nucleosomes. Lin28b, having the highest AT content (Table 1) and multiple Oct4 and other TF binding sites, was the most mobile. It was followed by the artificial Widom nucleosome with no Oct4 binding sites. The least mobile was Esrrb, which has a similar AT content as Widom and contains a single Oct4 binding site (Fig. 2 *A*; Table 1). A measurable population of the Lin28b nucleosome adopted conformations with high dRMSD values, indicating it undergoes larger conformational transitions compared to the other two. This hierarchy of structural flexibility was correlated with the compaction of the nucleosomes, with Lin28b sampling less-compact conformations (Fig. 2 *A*). The Esrrb nucleosome was the most compact (Fig. 2 *A*), whereas Widom adopted two distinct populations, one as compact as Esrrb and one as compact as the most compacted population of Lin28b (Fig. 2 *A*).

Comparing the three independent simulations of each nucleosome, we observed that Lin28b and Widom sampled distinct regions of the conformational space in each simulation, whereas the conformations of Esrrb were more similar among triplicates (Fig. S2 *A*). This confirmed that 1 μs is not sufficient for a converged sampling of the conformational space and that merging the simulations into ensembles is

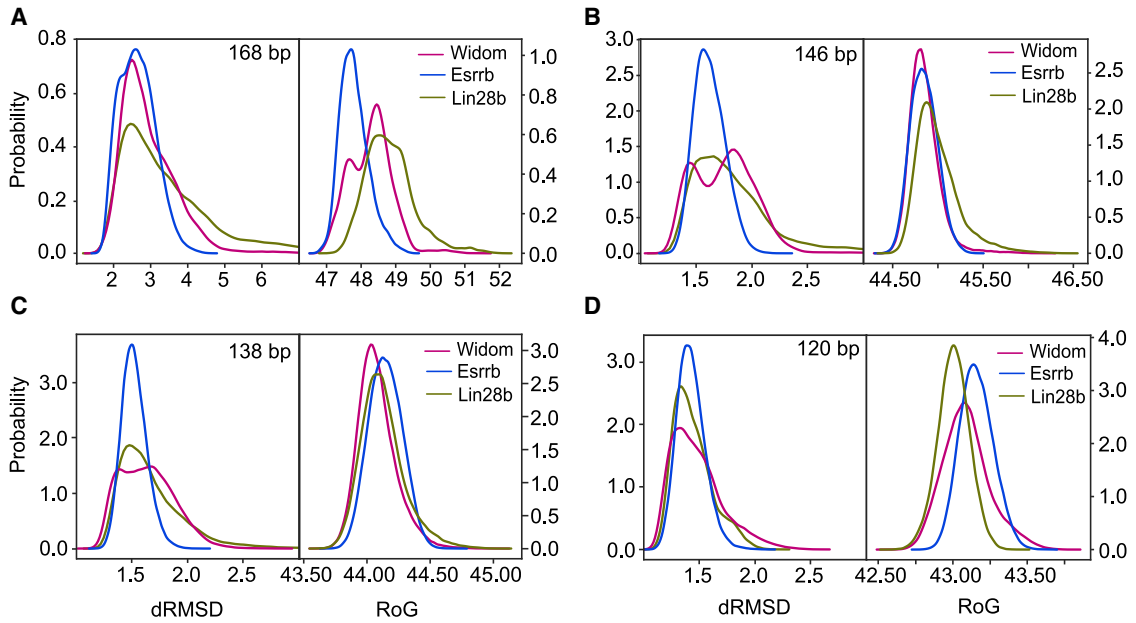


FIGURE 2 Sequence-dependent nucleosome flexibility. Distributions of the RMSD of interatomic distances in the DNA sugar-phosphate backbone (dRMSD, *left*) and radius of gyration (RoG, *right*) from the simulation ensembles of the three nucleosomes. The distributions were calculated for different regions of DNA: (*A*) the entire simulated DNA (168 bp), (*B*) the core nucleosomal DNA (146 bp) obtained by excluding 11 bp of L-DNA at each end, (*C*) nucleosomal DNA obtained by excluding 15 bp of DNA at each end (138 bp), and (*D*) nucleosomal DNA corresponding to the complete inner DNA turn and half of the outer DNA turn obtained by excluding 23 bp of DNA at each end (120 bp). To see this figure in color, go online.

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

necessary to obtain a more complete view of the nucleosome flexibility. The intermediate flexibility of the Widom nucleosome is due to motions mainly at the 5′ end that occurred in the second simulation, which included a short-lived large opening at the beginning of the simulation and a steady opening in the second half of the simulation (Figs. S2 *A* and S3 *A*). In addition, a partial opening at the 3′ end occurred after 750 ns in the third simulation (Figs. S2 *A* and S3 *A*). The higher structural flexibility of the natural Lin28b nucleosome is not only rooted in the different sampling across different simulations but also in transitions occurring in the same simulation (Fig. S2 *A*). Lin28b underwent large concerted opening-closing motions in the first two simulations and a smaller-scale closing in the third simulation mostly at the 5′ end (Fig. S3 *A*). The largest opening was observed in the second simulation, occurring after 100 ns and with a lifetime of ~700 ns, with the highest amplitude between 700 and 800 ns (Fig. S3 *A*).

We then computed the average structure of each ensemble of simulations and analyzed the elastic deformation energy of the nucleosomes. The nucleosome with the artificial Widom sequence has the lowest energy (Table 1), in agreement with its known strong nucleosome positioning. Both nucleosomes with native sequences have a higher energy, suggesting that they are more flexible. The high value obtained for the Esrrb nucleosome indicates that its dynamics may be more similar to those of the Lin28b nucleosome, and its stability in the simulations may be due to insufficient sampling.

To validate the observed differences in motion between these nucleosomes, we performed a thermal stability assay on reconstituted nucleosomes (51) (n = 8). We found that Widom had the highest mean temperature of dissociation in the majority of experiments (Fig. S4; Table 1). Esrrb had an intermediate temperature of dissociation between Widom and Lin28b. In some experiments, Esrrb was more stable than Widom (Fig. S4). Therefore, the low flexibility of Esrrb in the simulations reflects the dynamics within a subpopulation of its ensemble of conformations. Lin28b had the lowest temperature of dissociation, requiring the least energy for disassembly. This confirmed our findings from the simulations that Lin28b, a nucleosome with high AT content and multiple TF binding sites, is the most mobile among the three nucleosomes studied.

Differences in nucleosome structural flexibility are not restricted to the linker DNAs

To analyze the propagation of DNA flexibility into the core region, we calculated the distributions of dRMSD for different fragments of DNA. First, we compared the structural flexibility of the 146 bp DNA fragments that formally define the core nucleosomes. Interestingly, the core of the artificial Widom nucleosome displayed two distinct peaks, indicating the sampling of two distinct conformational populations (Fig. 2 *B*). However, in the case of the core fragments, these two populations contain equally compact conformations because the radius of gyration has a very narrow, single peaked distribution (Fig. 2 *B*). The core of Esrrb has a uniform, narrow dRMSD distribution reflecting its lower flexibility (Fig. 2 *B*). Remarkably, the core of Lin28b displayed a wide dRMSD histogram and adopted less-compact conformations (Fig. 2 *B*). Therefore, the increased flexibility of the Lin28b nucleosome is propagated beyond the L-DNA arms, and it involved decompaction of the core nucleosomal DNA. The most decompacted conformations were observed between 500 and 800 ns of the second simulation (Fig. S3 *B*).

In general, the differences observed between the 146 bp core fragments are similar to those between the 138 bp DNA fragments, which are the largest fragments devoid of L-DNA motion propagation. However, the two conformational populations of the Widom nucleosome are less clearly separated (Fig. 2 *C*), suggesting that these populations differ mainly in the L-DNAs and the adjacent 7 bp. The 138 bp fragment of Lin28b is still more mobile than the similar fragments of Widom and Esrrb (Fig. 2 *C*). This flexibility was mainly apparent in the first two simulations (Fig. S3 *C*). The main difference between the 146 and 138 bp fragments is in their compactness. The 138 bp fragments are very similar and show no decompaction in the case of Lin28b (Figs. 2 *C* and S3 *C*). Therefore, the overall decompaction of Lin28b was due to conformational dynamics in the L-DNA arms and the adjacent 7 bp from the nucleosomal DNA.

Finally, the hierarchy of the structural flexibility and compaction of the three nucleosomes changes when the inner 120 bp fragments were compared (Fig. 2 *D*). The flexibility of these fragments was lower and more similar between the nucleosomes, as shown by the narrower dRMSD distributions (Fig. 2 *D*). Interestingly, the Widom fragment shows the widest dRMSD distribution (Fig. 2 *D*), suggesting that the inner part of this nucleosome was somewhat more mobile than the corresponding part of the two natural nucleosomes (Figs. 2 *D* and S3 *D*). In addition, the 120 bp fragment of Lin28b was the most compact (Figs. 2 *D* and S3 *D*), indicating that a higher overall flexibility of the nucleosome is coupled with a less mobile and more compact region deep in the core around the dyad.

The amplitude of nucleosome breathing motions depends on DNA sequence

To understand the origin of the sequence-dependent nucleosome flexibility and compaction, we characterized the different types of motions we observed. These were breathing, in which the L-DNAs open and close either parallel or perpendicular to the nucleosomal superaxis, changing its accessibility; DNA overtwisting; and local fluctuations.

To characterize the breathing motions, we defined two angles, γ_1 and γ_2, that describe the motions of the L-DNA on two orthogonal planes (see also Materials and Methods). γ_1 describes a scissors-like motion of the L-DNAs in which the

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

dyad can be viewed as the point connecting the two blades, whereas γ_2 describes a motion perpendicular to the plane of the blades. From the γ_1-γ_2 histograms (Fig. 3 *C*), we concluded that all three nucleosomes sampled different but overlapping regions of the conformational space in the three independent simulations (represented in *magenta*,

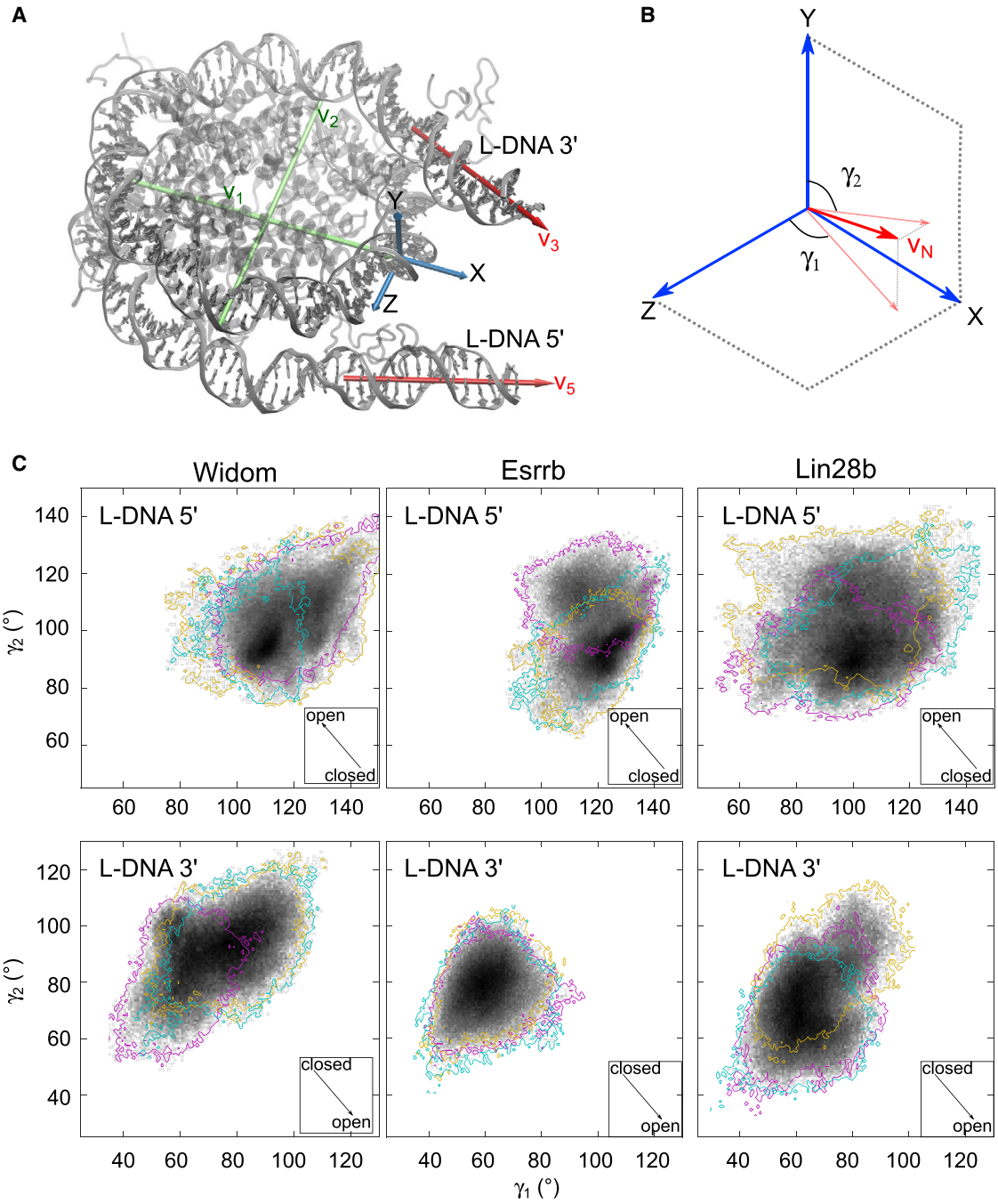


FIGURE 3 Motions of the linker DNA arms. (*A*) The coordinate system definition is given. The *XYZ* reference system was defined as follows (see Materials and Methods for details): *X* was defined along the dyad axis, *Y* as the cross products between a vector v_1 defined along the dyad axis and a vector v_2 defined to be approximately orthogonal to v_1 and to intersect v_1 as close as possible to the center of the nucleosome, and *Z* as the cross product between *X* and *Y*. (*B*) Definition of the angles, γ_1 and γ_2, ισ γιτπεν. γ_1 is the angle between the projection of the vector v_N defined along the helical axis of the L-DNA on the *XZ* plane and the *Z* axis. γ_2 is the angle formed between the projection of v_N on the *XY* plane and the *Y* axis. An increase of γ_1 indicates opening at the 3′ L-DNA but closing at the 5′ L-DNA, whereas an increase of γ_2 indicates closing at the 3′ L-DNA and opening at the 5′ L-DNA. (*C*) Two-dimensional histograms showing the sampling of the γ_1/γ_2 conformational space for both 5′ L-DNA and 3′ L-DNA. The colors (*cyan*, *magenta*, and *gold*) indicate the sampling covered by each of the three 1 μs simulations. The arrows in the square inserts indicate the direction of the nucleosome opening. To see this figure in color, go online.

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

gold, and *cyan* in Fig. 3 *C*). This indicates that although full convergence was not achieved in 1 μs, the most relevant parts of the conformational space were sampled. Further, we confirmed the asymmetry in the nucleosome flexibility, the 5′ L-DNA sampling a larger area in the two-dimensional histogram than the 3′ L-DNA, for all three nucleosomes. At last, we confirmed the higher flexibility of Lin28b, especially at the 5′ end (Fig. 3 *C*, *upper right*), and the lower flexibility of Esrrb (Fig. 3 *C*, *middle column*).

At the 5′ L-DNA of the nucleosome, Widom and Esrrb adopted more closed conformations along γ_1 than Lin28b (Fig. 3 *C*, *upper row*). Although in all three simulations, Lin28b sampled more open conformations with lower γ_1, in one particular simulation (*gold* in Fig. 3 *C*), this nucleosome adopted wide, open conformations characterized by low γ_1 and high γ_2 simultaneously. The largest opening occurred shortly after the start of that simulation and lasted for ~400 ns (Fig. S5 *A*). In general, we observed reversible transitions between closed and open conformations (Fig. S5 *A*). At the 3′ end, the three nucleosomes sampled more similar regions of the conformational space. Lin28b sampled a larger area of the two-dimensional histogram compared to Esrrb but somewhat lower compared to Widom (Fig. 3 *C*, *lower column*). Interestingly, the Widom nucleosome displayed transitions to conformations with high γ_1 in two simulations (Fig. S5 *B*). However, during these transitions, there was no large opening because the opening along γ_1 was balanced by closing along γ_2 (Fig. S5 *B*). Similarly, Lin28b displayed more open conformations along γ_2 at the 3′ L-DNA arm in the simulation in which the 5′ L-DNA adopted the most closed conformations (*third simulation* in Fig. S5, *A* and *B*). However, the maximum Pearson correlation coefficient R^2 between two different angles during a simulation was roughly 0.6, which indicated no correlation between the motions of the two L-DNA arms.

To extract the slow low frequency motions from the simulations, we calculated the essential dynamics using PCA and monitored the evolution of the angles γ_1 and γ_2 along the trajectories projected on the first five principal components (PCs) (see Materials and Methods). This allowed us to characterize the correlated components of the different types of motions and to separate the uncorrelated motions. PC1 and PC2 described the largest breathing motions in the nucleosomes (Fig. 4). For Widom, both PCs revealed coupled breathing motions along γ_1 and γ_2 at the 5′ end, whereas for both native nucleosomes, these motions were uncoupled (Figs. 4 and S6). Lin28b displayed the largest amplitude breathing at the 5′ end. Esrrb displayed a larger amplitude breathing than Widom, despite its overall lower flexibility. At the 3′ end, the motion with the largest amplitude was in Widom along PC1, whereas the breathing of Esrrb had the lowest amplitude, and the breathing of Lin28b was restricted along γ_1. These findings suggest that the pattern of the breathing motions depends on the DNA sequence: in nucleosomes with strong positioning sequences such as Widom, the breathing along γ_1 and γ_2 are coupled and of lower amplitude, whereas natural nucleosomes have higher uncoupled motions.

Twisting motions in the DNA propagate into the nucleosome core

In addition to breathing, the PCA revealed overtwisting motions of the nucleosomal DNA (Video S1), which are

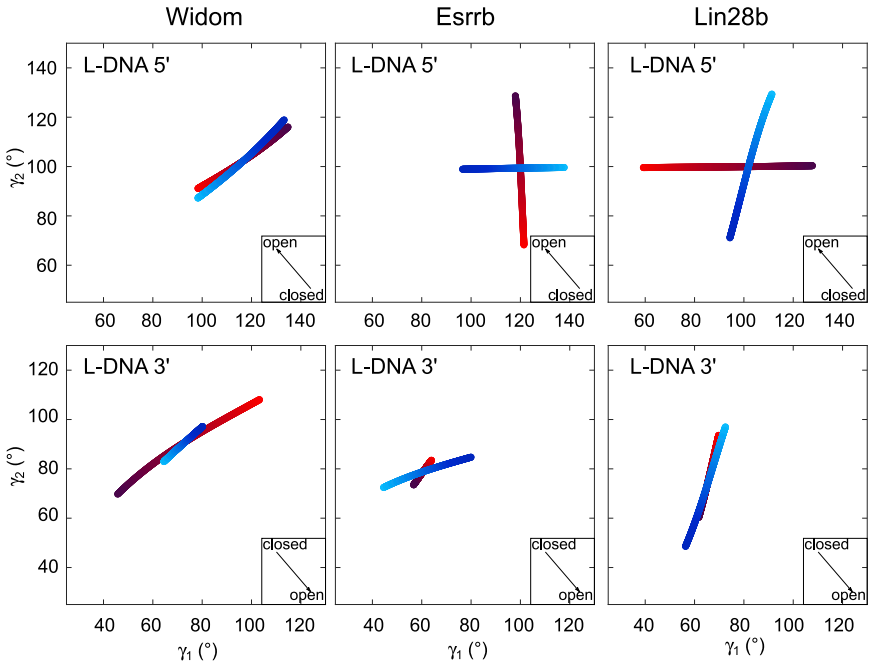


FIGURE 4 Correlations of nucleosome opening-closing motions. Evolution of the angles γ_1 and γ_2 in the two first PCs. After PCA of the ensemble of simulations for each nucleosome, pseudotrajectories along the two first eigenvectors were generated. Then, the motion of the L-DNA arms was analyzed as described in Fig. 3. The motions corresponding to the first and second eigenvectors are in red and blue, respectively. Light and dark colors indicate the low and high amplitudes of the motions, respectively. The arrows in the square inserts indicate the direction of the nucleosome opening. To see this figure in color, go online.

notoriously difficult to quantify. Previously, these were characterized by monitoring the changes in the contacts of selected arginine residues (30) or the change of the bp orientation relative to the nucleosome superhelical axis (15). Both approaches have important limitations. The first is only suitable for low number of contacts between amino acid side chains and DNA bases. For example, in coarse-grained simulations, often a single contact is established between one amino acid bead and one bp bead. On the other hand, in atomistic simulations, arginines make multiple simultaneous contacts with more than one bp. The second approach does not provide a clear separation between twisting and other motions. The change in bp orientation may be due to pure twisting motions or to a bp rotating during nucleosome breathing. To track the changes in both bp orientation and position, we developed a new, to our knowledge, approach by establishing a moving coordinate system for every bp with the *Y* axis perpendicular to the helical axis of DNA (Figs. 5, *A* and *B* and S7 *A*). We then calculated the angle α between *Y* at a given time and *Y* at initial time. A concerted, large-amplitude motion such as breathing or twisting is characterized by a wide range of values for α. To discriminate between breathing and twisting, we calculated the displacement *d* of the center of the coordinate system. Breathing motions involve a large *d*, twisting a moderate *d*, and no motion a low *d*. Therefore, a combination of wide α range and moderate *d* is a footprint for twisting. We performed the analysis on the pseudotrajectories along the first five PCs because these displayed the motion as unidirectional from low to high amplitudes (Fig. S7 *B*).

We focused on the motions along PC1 and PC5 because they described the highest-amplitude breathing and twisting. PC5 was the first PC in which twisting was uncoupled from breathing. We observed overtwisting of the DNA in the simulations of Widom and Lin28b (Figs. 5 *C* and S7 *B*), but not in those of Esrrb. For Widom, this

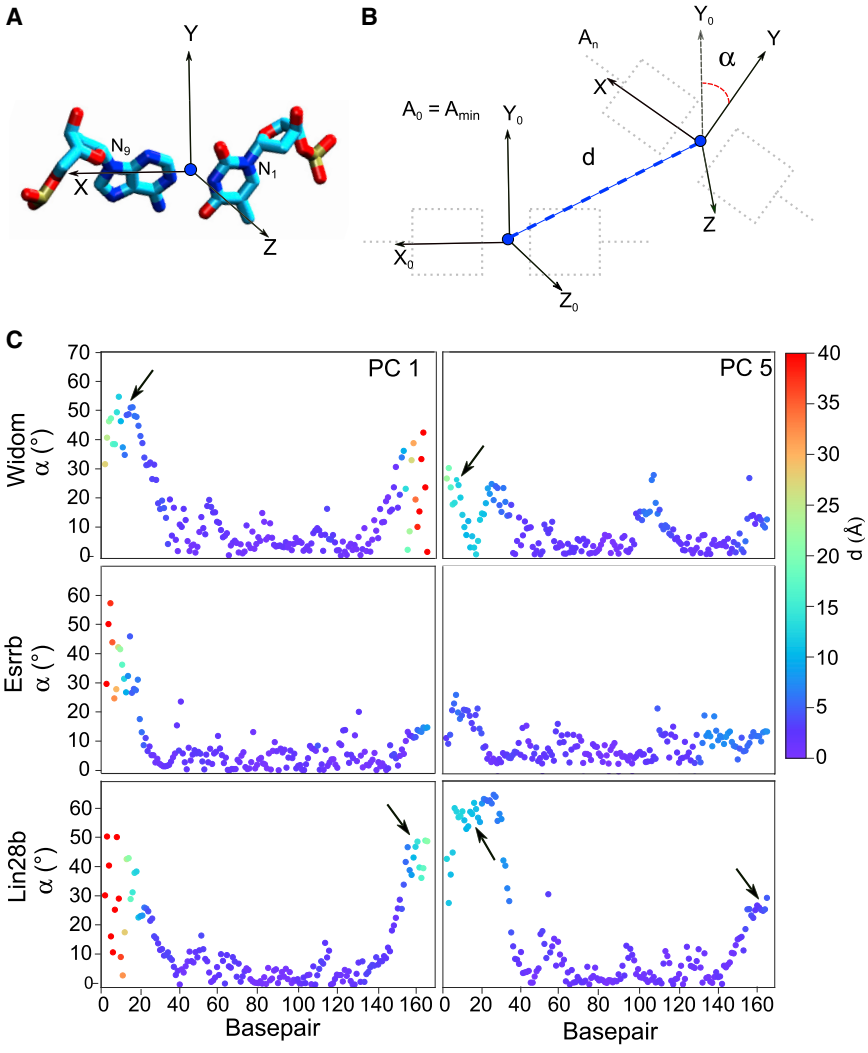


FIGURE 5 Twisting motions in the nucleosomes. (*A*) The coordinate system definition is given. The *XYZ* reference system was defined using the following approach (see also Materials and Methods): *X* is along a vector defined to connect the N1 atom of the pyrimidine with the N9 atom of the purine. *Z* is along the cross product between *X* and a vector defined from atom positions to be approximately orthogonal to *X*. The intersection of this vector with *X* is the origin of the coordinate system and is near the bp center. *Y* is the cross product between *X* and *Z* (see also Fig. S7 *A*). (*B*) A schematic representation of the motion of a bp during nucleosome twisting is shown. In each snapshot along the PC trajectory, the coordinate system is redefined. Then, the displacement *d* of the current position of the origin from the position at the minimum amplitude and the angle α between the current *Y* axis and the *Y* axis at minimum amplitude (Y_0) were calculated. (*C*) The range of α is plotted for each bp along the first five PC trajectories. Each point is colored by the value of *d*. Twisting motions are characterized by high values of α and low displacement values and are indicated with black arrows. To see this figure in color, go online.

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

motion is revealed mainly by PC1, occurs at the 5′ end, is coupled with breathing, and propagates ~30 bp into the core nucleosomal DNA (Fig. 5 *C*, *upper left*). For Lin28b, the highest amplitude twisting motion was at the 5′ end (Fig. 5 *C*). Along PC5, this motion was uncoupled from breathing and propagated almost 40 bps into the core nucleosomal DNA (Fig. 5 *C*, *lower right*; Video S1). A similar twisting at the 3′ end occurred, but it propagated less (Fig. 5 *C*). This large-scale twisting originated from a unidirectional overtwisting that occurred in the first simulation of the Lin28b nucleosome, reaching the maximum amplitude after ~750 ns (Fig. S8). The twisting motions observed in the other simulations (e.g., of Widom) were multistep bidirectional motions involving partial over- and undertwisting (Fig. S8). These findings indicate that the twisting motions also display a higher amplitude in Lin28b compared to Widom and Esrrb.

The pattern of local DNA flexibility varies among nucleosomes

Next, we aimed to characterize the local DNA flexibility and study whether the Oct4 binding sites are in regions of increased structural flexibility. All Oct4 binding sites proposed by Soufi et al. (two within Lin28b and one within Esrrb (4,5)) (Fig. 6 *A*) are exposed and accessible for Oct4 binding in our nucleosome models, despite the low resolution of the MNase data used to build the models (see Materials and Methods). In Lin28b, one partial binding site for the POU_{HD} is present and exposed close to the 5′ L-DNA, whereas closer to the dyad, there is an octamer motif, with only the partial

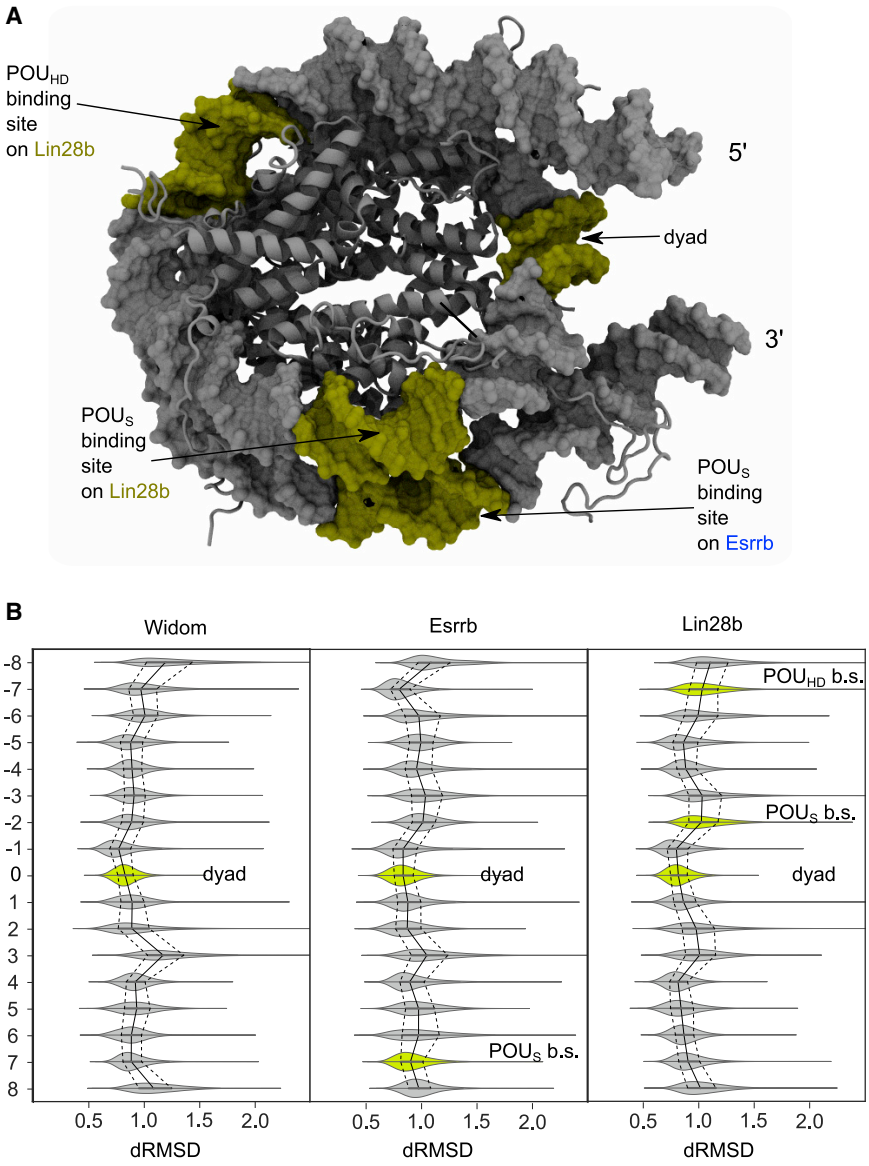


FIGURE 6 Local flexibility in the nucleosomes. Distributions of dRMSD for different DNA segments are shown. The first segment was the 8 bp region centered on the dyad. The other segments were 8 bp long, extending from the dyad segment to the ends of the 146 bp nucleosome core particle. (*A*) The position in the nucleosome of the dyad and the regions containing the Oct4 binding sites (in *lime*) are shown. (*B*) dRMSD distributions for all segments of each nucleosome are shown. The dyad and the Oct4 binding sites are in lime. Lines indicate the median (*solid*) and the first and third quartiles (*dashed*). To see this figure in color, go online.

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

motif recognized by the POU_S accessible. From here on, we refer to these sites as the POU_{HD} and POU_S binding sites, respectively. For Esrrb, only one partial binding site for the POU_S is present and exposed close to the 3′ end.

To investigate the local flexibility in the nucleosome (Fig. 1), we divided the DNA into eight bp fragments starting from the dyad. For each fragment, we calculated the dRMSD using the average structure of the fragment DNA backbone from all simulations as reference. We visualized the results as violin plots that revealed both the statistical median and quartiles and the histograms of the distribution of dRMSDs. As expected, the dyad fragment was the region with the lowest flexibility (Fig. 6 *B*). Interestingly, in both Esrrb and Lin28b, both regions next to the dyad showed increased local flexibility, whereas in Widom, only the region at the 3′ side of the dyad was more flexible. In addition, the oscillatory pattern of the local flexibility in which more flexible regions alternate with less flexible regions was more pronounced in the native nucleosomes (Fig. 6 *B*). In Lin28b, the Oct4 binding sites previously identified experimentally (4,5) were located either in or next to regions of increased local flexibility, suggesting that the accessibility to these binding sites may be facilitated by increased local nucleosome flexibility. On the other hand, in the Esrrb nucleosome, although the Oct4 binding site is located toward the 3′ end of the nucleosome, it is not in a fragment with increased flexibility.

Alternative configurations of Oct4 form stable complexes with nucleosomes

Finally, to reveal the structural basis for Oct4-nucleosome recognition, we modeled the structure of Oct4 bound to Lin28b using the experimental data available from Soufi et al. (5). The two binding sites for Oct4 on Lin28b were proposed to be recognized by the POU_S and POU_{HD}, respectively, from biochemical, footprinting, and ChIP-Seq experiments.

We confirmed that the recognition of the entire octamer motif by Oct4 with both domains bound in the configuration known from the crystal structure of Oct4 on free DNA (Fig. 7 *A*, *left structure*) was incompatible with nucleosome binding on all proposed binding sites because of steric clashes between one domain and the histones (Fig. 7, *B–D*, *left structures*). For this reason, we modeled alternative Oct4 configurations on the nucleosome. First, we used a configuration found in the structure of an Oct4 homodimer bound to the palindromic MORE motif (MORE configuration) (10). In this configuration, both the POU_S and POU_{HD} are bound on the same side of the DNA (Fig. 7 *A*, *middle structure*). Second, we used different Oct4 configurations generated from MD simulations of free Oct4 (Fig. 7 *A*, *right structure*) (MD configurations). We simulated the sequence-specific binding of one Oct4 domain in multiple-step restrained MD equilibrations by applying restraints to enforce the formation of the interactions with DNA bases known from the Oct4-DNA structure (PDB: 3L1P). In the last step of the equilibrations, we released these restraints. Then, we monitored the stability of the Oct4-DNA interactions in unrestrained, 100-ns-long MD simulations started from those equilibrations in which the interactions were maintained (Table 2) (see Materials and Methods for details). We evaluated the stability of the Oct4-nucleosome configurations with the following criteria: the number of stable contacts with the DNA and the presence of specific hydrogen bonds between specific amino acids and DNA bases (Q44-A in the POU_S and N143-A in the POU_{HD} (6,49)). We expected the protein-DNA interactions to break quickly at the beginning of the simulations for any invalid Oct4 binding mode.

Binding in the POU_S binding site was stable in two of the three equilibrations with the Oct4 MORE configuration and in one of the three equilibrations with the MD configuration. The interactions remained stable during the 100 ns simulations (Fig. 7 *B*, *middle* and *right structures*; Table 2; Video S2; Data S1 (*middle*) and S2 (*right*)). In these complexes, the POU_{HD} hovered nonspecifically on the same gyre of the nucleosome. For the POU_{HD} to reach the second gyre, a major conformational change of the linker between the two domains must occur. Interestingly, some models suggest that the two domains may recognize their sequence-specific binding sites simultaneously if the linker was in an extended conformation (Fig. 7 *B*, *right structure*).

Binding in the POU_{HD} binding site may occur in two different orientations. For a complete recognition of the typical "TAAT" site, the POU_{HD} has to be bound in the 3′-5′ orientation (Fig. 7 *C*, *left drawing*). The MORE configuration was not compatible with such a binding mode because of steric clashes between the POU_S and the histones (Fig. 7 *C*, *middle structure*). Moreover, in all simulations started with MD configurations, the interactions of the POU_{HD} residues with DNA bases were not stable (Fig. 7 *C*, *right structure*; Table 2; Video S3). Therefore, we concluded that this binding mode is not valid. In an alternative binding mode, the POU_{HD} recognizes sequence specifically only an "AT" dinucleotide on the 5′-3′ strand (Fig. 7 *D*, *left drawing*). Interestingly, both the MORE and MD configurations of Oct4 were compatible with this binding mode, forming stable complexes with typical POU_{HD}-DNA interactions maintained during the simulations (Fig. 7 *D*, *middle* and *right structures*; Table 2; Video S4; Data S3 (*middle*) and S4 (*right*)). This suggests that the POU_{HD} is able to bind to the nucleosomal DNA in different orientations comparing to free DNA. In these models, the POU_S remained nonspecifically close to DNA on the same gyre as the POU_{HD}. Again, the binding of the two domains to different gyres may occur only upon a large conformational transition in the linker region.

These findings show that alternative, but not all, configurations of Oct4 with either the POU_S or the POU_{HD} bound to

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

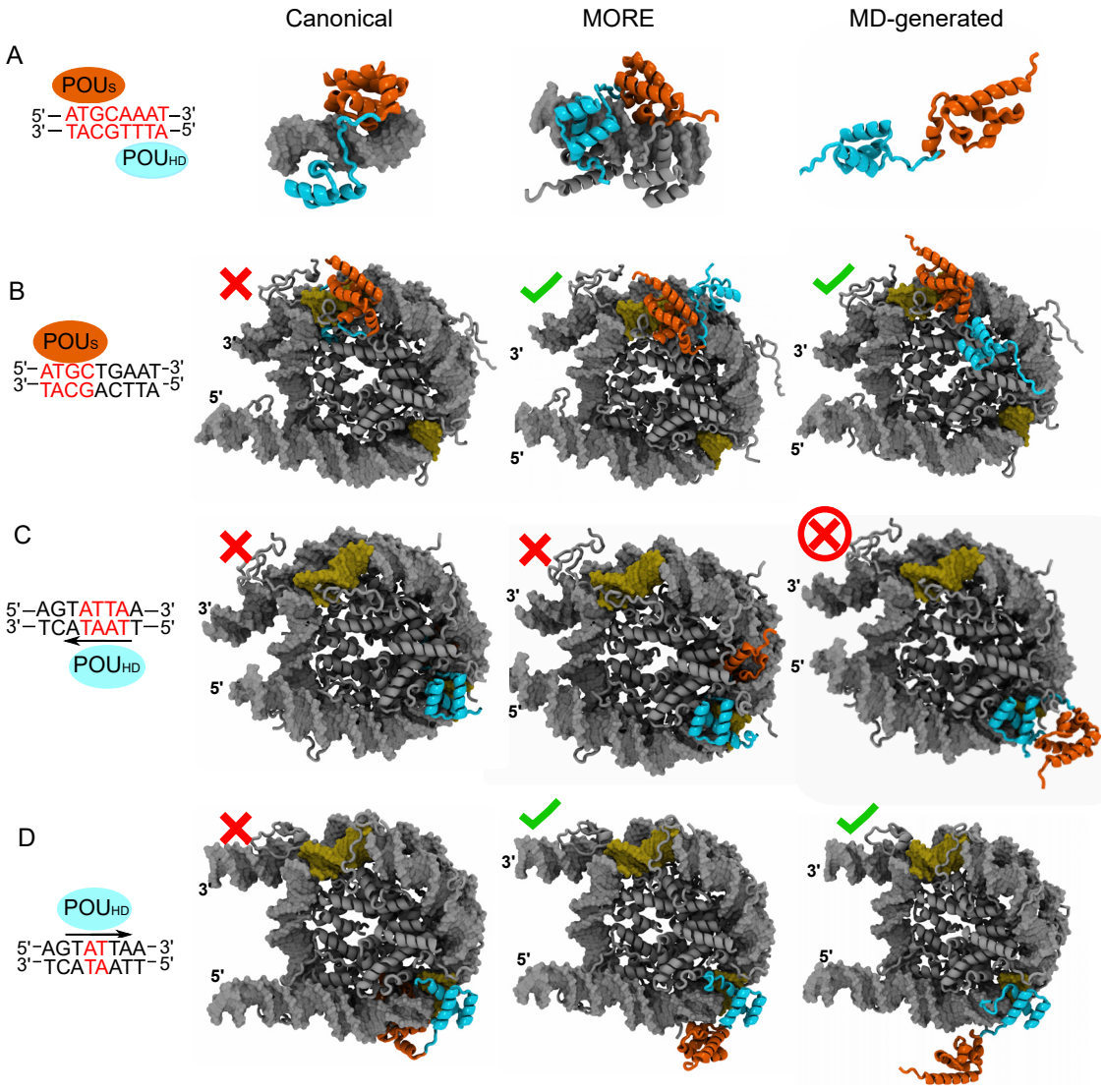


FIGURE 7 Structural basis for Oct4-nucleosome recognition. (*A*) Structures of Oct4 used to build Oct4-nucleosome complexes are shown. From left to right, a schematic representation of the canonical configuration found in the crystal structure of an Oct4 dimer bound to palindromic DNA (6) is followed by structural views of the canonical configuration, MORE homodimer configuration (originally found in the crystal structure of Oct1-Oct1-DNA complex and modeled in our previous work (10)), and an example of a configuration obtained from MD simulations of free Oct4. The recognized bases are highlighted in red. (*B–D*) Oct4-nucleosome complexes are shown. From left to right, a schematic representation indicating which domain is bound in a sequence-specific manner is followed by structural models of Oct4-nucleosome complexes built using the canonical, MORE, and MD-generated configurations. (*B*) POU_S is bound sequence specifically. (*C* and *D*) POU_{HD} is bound sequence specifically either in an orientation as observed in structures of Oct4-DNA complexes (*C*) or in an orientation in which an AT bp step on the opposite strand is recognized (*D*). The black arrows indicate the orientation of the POU_{HD} binding. The POU_S and POU_{HD} are in orange and cyan cartoons, and the nucleosome core is in gray cartoons. To see this figure in color, go online.

sequence-specific sites formed stable interactions with the nucleosomes on the simulated timescale. This demonstrates that Oct4 binding is compatible with the nucleosome curvature and the histone-DNA interactions and that MD simulations are powerful in identifying configurations and orientations of multidomain TFs bound to nucleosomes.

DISCUSSION

Here, we revealed the nucleosome dynamics that mediate the binding of the pioneer TF Oct4 to DNA wrapped in nucleosomes and probed for the structural basis of Oct4-nucleosome recognition using microsecond timescale MD simulations. We first studied the dynamics of three nucleosomes, two with natural DNA sequences proposed to be bound by Oct4 during cellular reprogramming to pluripotency (4) and one with a strong positioning sequence and no TF binding site. Whereas the dynamics of artificial nucleosomes have been previously studied with atomistic simulations (14–23), the sequence dependence of these dynamics has not been explored. We found significant differences between the dynamics of the three nucleosomes,

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

TABLE 2 Oct4-DNA Contacts in Oct4-Nucleosome Complexes

Binding Site	Oct4 Configuration	Simulation Step	# Stable Contacts[b] POU_S-DNA Bases	# Stable Contacts[b] POU_S-DNA Backbone	# Stable Contacts[b] POU_{HD}-DNA Bases	# Stable Contacts[b] POU_{HD}-DNA Backbone	% Time with Selected Contact[c] Present
POU_S	MD	Eq01[a]	67	38	0	0	92.40%
		Eq02	14	43	0	0	0%
		Eq03	30	51	0	0	0.52%
		Prod	61	50	0	0	95.32%
	MORE	Eq01	69	47	4	23	85.80%
		Eq02	47	71	0	6	6.56%
		Eq03[a]	69	77	0	1	99.64%
		Prod	65	54	2	7	98.73%
POU_{HD} 1 (3′-5′)	MD1	Eq01	0	0	31	8	0%
		Eq02	0	2	3	13	0%
		Eq03	0	0	21	15	0%
		Prod	–	–	–	–	–
	MD2	Eq01[a]	1	5	41	44	65.52%
		Eq02	1	5	17	20	0%
		Eq03	0	0	12	26	0%
		Prod	0	0	5	11	19.84%
POU_{HD} 2 (5′-3′)	MD	Eq01	0	0	46	47	69.80%
		Eq02	0	0	34	37	87.08%
		Eq03[a]	0	0	42	16	88.16%
		Prod	0	0	34	17	63.98%
	MORE	Eq01[a]	18	47	48	37	97.68%
		Eq02	23	51	50	50	88.68%
		Eq03	16	50	42	51	95.84%
		Prod	2	35	31	27	85.43%

Eq, Equilibration (last 5 ns); Prod, Production (100 ns).
[a]Equilibration chosen for production.
[b]Stable contact: atoms at a distance <4.5 Å for at least 75% of the simulation.
[c]Selected contacts: Q44-A for the POU_S and N143-A for the POU_{HD}. Contact was considered present if donor-acceptor distance was smaller than 4 Å.

with the nucleosome with the most TF binding sites showing the largest amplitude of breathing and twisting motions. This is in agreement with previous studies that revealed the more dynamic nature of natural nucleosomes (13,29,37). Based on our findings, we propose that nucleosomes with multiple TF binding sites are more mobile and less compact. However, we also found that one of the natural nucleosomes we studied was at least as rigid as the artificial nucleosome, suggesting that the flexibility of natural nucleosomes varies significantly.

Our findings were in agreement with in vitro experiments in which temperature-induced nucleosome disassembly was performed. The most mobile nucleosome in the simulations also disassembled at the lowest temperature. Moreover, the natural nucleosome that was more rigid in the simulations had different disassembly profiles in the experiments (Fig. S4). In some experiments, the disassembly temperature of this nucleosome was higher, in agreement with the rigidity observed in the simulations. This suggests that natural nucleosomes may exist in different states, some more dynamic than others. In such cases, the simulations are unlikely to describe the transition between different states because of insufficient sampling and the large energy barriers involved.

Nucleosome motions are particularly difficult to study in atomistic simulations because of the long timescales involved (52,53). Despite the lack of full sampling convergence, we observed a good overlap between independent simulations for nucleosome breathing (Fig. 3), and we achieved a wider conformational sampling by combining independent simulations of the same species into ensembles. This enabled us to study both breathing and twisting motions in the nucleosomes and to draw conclusions about the sequence dependence of these motions at a level of detail not feasible in coarse-grained simulations, which were previously used to study nucleosome-dependent dynamics (29,37). The atomic resolution provides the means to study the determinant modulators of nucleosome dynamics such as the role of anchoring histone residues (15) and the motion of histone tails. The phase space sampling of the histone tails we report here is more exhaustive than in other published studies to date (17,18,54), but it is still insufficient for a converged exploration of their dynamics. Because of their disordered nature, the dynamics observed here may be affected both by insufficient sampling and force field induced overstabilization of their contacts with DNA (55). The conformational ensemble explored by the histone tails may be increased using enhanced sampling techniques (20), whereas larger-scale nucleosome motions like unwrapping and their mechanisms can be studied by forcing such events (56–58). Ultimately, a combined technique that

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

would allow on-the-fly conversion between different representations during the simulation may be the most powerful in revealing dynamics of large structures such as the nucleosome or larger chromatin fibers.

Our findings suggest that a nucleosome with multiple TF binding sites is more dynamic and that for that nucleosome, TF binding sites are located in or near regions of increased local flexibility. This enhanced flexibility might facilitate the recognition and binding of nucleosomes by TFs, with both the global and local structural flexibility of the DNA wrapped around the core histones acting as mediators for binding of proteins to nucleosomes. We have previously shown that nucleosome dynamics play a role in the binding of the linker histone to nucleosomes (25). Based on these findings, we propose that alongside DNA sequence and post-translational modifications of histone tails, nucleosome dynamics provide an additional platform for protein-nucleosome recognition.

Ultimately, we provide here a structural basis for Oct4-nucleosome recognition. This is particularly important because to date, to our knowledge, no experimental structure of a TF bound to nucleosome has been resolved. Understanding this interaction at atomic resolution is necessary to provide novel approaches to modify the functional outcome of TF binding to DNA. For example, enhancing nucleosome binding of selected pioneer factors may lead to faster and more efficient chromatin decompaction, which in turn may have profound effects on gene regulation programs. In the case of Oct4, we first confirmed the partial binding site recognition model proposed by Soufi et al. (5). Then, we revealed alternative configurations of Oct4-nucleosome complexes that maintained the Oct4-DNA interactions during 100-ns-long MD simulations. Because all our simulations were started with intact nucleosomes, we cannot exclude a mechanism for TF-nucleosome binding that involves an initial partial decompaction and unwrapping of the nucleosome followed by recognition of the binding sites as in free DNA.

CONCLUSIONS

In this study, we present how structural dynamics mediate the binding of the pioneer TF Oct4 to nucleosomes. From a series of atomistic MD simulations of nucleosomes, we reveal that a nucleosome with multiple TF binding sites displays higher structural flexibility compared to a nucleosome with a single TF binding site and another nucleosome with an artificial sequence engineered for strong positioning. We experimentally validated these findings. The increased structural flexibility lead to higher amplitude breathing and twisting motions. Moreover, in this nucleosome, the Oct4 binding sites are located in regions of increased local flexibility. Based on these findings, we propose that nucleosome dynamics mediate the binding of pioneer TFs. However, we also observed that one nucleosome with a single Oct4 binding site is not more flexible than the reference artificial nucleosome. This suggests that Oct4 and other pioneer factors may use different structural features to bind to different nucleosomes.

In addition, we reveal alternative structures of Oct4-nucleosome complexes that are stable on a timescale of 100 ns. Therefore, we validate the proposed model of the partial binding of Oct4 to nucleosomes, according to which only one of the DNA binding domains establishes sequence-specific interactions with the DNA. These findings demonstrated that the nucleosome curvature and DNA-histone interactions do not interfere with Oct4 binding.

Lastly, we show that MD simulations represent a powerful tool to identify configurations that are incompatible with nucleosome binding. Longer classical MD simulations of TF-nucleosome complexes and comparing with similar-length simulations of free nucleosomes will provide a framework for studying the role of TF binding to nucleosome dynamics. For this, enhanced sampling techniques may be required to capture larger-scale motions and longer timescales (59).

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2019.12.038.

AUTHOR CONTRIBUTIONS

J.H. designed, performed, and analyzed the simulations and wrote the manuscript. C.M.M. designed, performed, and analyzed the experiments and contributed discussion and writing. H.R.S. contributed to study design, discussion, and writing. V.C. designed the simulations, contributed analysis and discussion, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank Jürgen Walther for providing the covariance matrices and the code for the elastic deformation energy calculations. We thank Abdenour Soufi for providing the nucleosome sequences and for discussions. We also thank Peter Becker and Catherine Regnard at the Ludwig-Maximilians University in Münich for nucleosome assembly training and reagents.

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.). Funding for open access charge: The Royal Netherlands Academy of Arts and Sciences.

REFERENCES

1. Luger, K., A. W. Mäder, …, T. J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*. 389:251–260.

2. Zaret, K. S., and S. E. Mango. 2016. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* 37:76–81.

Please cite this article in press as: Huertas et al., Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2019.12.038

3. Takahashi, K., and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 126:663–676.
4. Soufi, A., G. Donahue, and K. S. Zaret. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell.* 151:994–1004.
5. Soufi, A., M. F. Garcia, …, K. S. Zaret. 2015. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell.* 161:555–568.
6. Esch, D., J. Vahokoski, …, H. R. Schöler. 2013. A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat. Cell Biol.* 15:295–301.
7. Chen, X., H. Xu, …, H. H. Ng. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell.* 133:1106–1117.
8. Merino, F., C. K. L. Ng, …, V. Cojocaru. 2014. Structural basis for the SOX-dependent genomic redistribution of OCT4 in stem cell differentiation. *Structure.* 22:1274–1286.
9. Merino, F., B. Bouvier, and V. Cojocaru. 2015. Cooperative DNA recognition modulated by an interplay between protein-protein interactions and DNA-mediated allostery. *PLoS Comput. Biol.* 11:e1004287.
10. Jerabek, S., C. K. Ng, …, R. Jauch. 2017. Changing POU dimerization preferences converts Oct6 into a pluripotency inducer. *EMBO Rep.* 18:319–333.
11. Zhu, F., L. Farnung, …, J. Taipale. 2018. The interaction landscape between transcription factors and the nucleosome. *Nature.* 562:76–81.
12. Radman-Livaja, M., and O. J. Rando. 2010. Nucleosome positioning: how is it established, and why does it matter? *Dev. Biol.* 339:258–266.
13. Eslami-Mossallam, B., H. Schiessel, and J. van Noort. 2016. Nucleosome dynamics: sequence matters. *Adv. Colloid Interface Sci.* 232:101–113.
14. Bishop, T. C. 2005. Molecular dynamics simulations of a nucleosome and free DNA. *J. Biomol. Struct. Dyn.* 22:673–686.
15. Shaytan, A. K., G. A. Armeev, …, A. R. Panchenko. 2016. Coupling between histone conformations and DNA geometry in nucleosomes on a microsecond timescale: atomistic insights into nucleosome functions. *J. Mol. Biol.* 428:221–237.
16. Lehmann, K., R. Zhang, …, K. Toth. 2017. Effects of charge-modifying mutations in histone H2A α3-domain on nucleosome stability assessed by single-pair FRET and MD simulations. *Sci. Rep.* 7:13303.
17. Erler, J., R. Zhang, …, J. Langowski. 2014. The role of histone tails in the nucleosome: a computational study. *Biophys. J.* 107:2911–2922.
18. Biswas, M., K. Voltz, …, J. Langowski. 2011. Role of histone tails in structural stability of the nucleosome. *PLoS Comput. Biol.* 7:e1002279.
19. Chakraborty, K., M. Kang, and S. M. Loverde. 2018. Molecular mechanism for the role of the H2A and H2B histone tails in nucleosome repositioning. *J. Phys. Chem. B.* 122:11827–11840.
20. Ikebe, J., S. Sakuraba, and H. Kono. 2016. H3 histone tail conformation within the nucleosome and the impact of K14 acetylation studied using enhanced sampling simulation. *PLoS Comput. Biol.* 12:e1004788.
21. Collepardo-Guevara, R., G. Portella, …, M. Orozco. 2015. Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: a multiscale computational study. *J. Am. Chem. Soc.* 137:10205–10215.
22. Pasi, M., and R. Lavery. 2016. Structure and dynamics of DNA loops on nucleosomes studied with atomistic, microsecond-scale molecular dynamics. *Nucleic Acids Res.* 44:5450–5456.
23. Sun, R., Z. Li, and T. C. Bishop. 2019. TMB library of nucleosome simulations. *J. Chem. Inf. Model.* 59:4289–4299.
24. Winogradoff, D., and A. Aksimentiev. 2019. Molecular mechanism of spontaneous nucleosome unraveling. *J. Mol. Biol.* 431:323–335.
25. Öztürk, M. A., G. V. Pachov, …, V. Cojocaru. 2016. Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome. *Nucleic Acids Res.* 44:6599–6613.
26. Biswas, M., J. Langowski, and T. C. Bishop. 2013. Atomistic simulations of nucleosomes. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 3:378–392.
27. Davey, C. A., D. F. Sargent, …, T. J. Richmond. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J. Mol. Biol.* 319:1097–1113.
28. Freeman, G. S., J. P. Lequieu, …, J. J. de Pablo. 2014. DNA shape dominates sequence affinity in nucleosome formation. *Phys. Rev. Lett.* 113:168101.
29. Niina, T., G. B. Brandani, …, S. Takada. 2017. Sequence-dependent nucleosome sliding in rotation-coupled and uncoupled modes revealed by molecular simulations. *PLoS Comput. Biol.* 13:e1005880.
30. Brandani, G. B., T. Niina, …, S. Takada. 2018. DNA sliding in nucleosomes via twist defect propagation revealed by molecular simulations. *Nucleic Acids Res.* 46:2788–2801.
31. Lequieu, J., D. C. Schwartz, and J. J. de Pablo. 2017. In silico evidence for sequence-dependent nucleosome sliding. *Proc. Natl. Acad. Sci. USA.* 114:E9197–E9205.
32. Kenzaki, H., and S. Takada. 2015. Partial unwrapping and histone tail dynamics in nucleosome revealed by coarse-grained molecular simulations. *PLoS Comput. Biol.* 11:e1004443.
33. de Bruin, L., M. Tompitak, …, H. Schiessel. 2016. Why do nucleosomes unwrap asymmetrically? *J. Phys. Chem. B.* 120:5855–5863.
34. Lowary, P. T., and J. Widom. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* 276:19–42.
35. Vasudevan, D., E. Y. D. Chua, and C. A. Davey. 2010. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol.* 403:1–10.
36. Chua, E. Y. D., D. Vasudevan, …, C. A. Davey. 2012. The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res.* 40:6338–6352.
37. Culkin, J., L. de Bruin, …, H. Schiessel. 2017. The role of DNA sequence in nucleosome breathing. *Eur. Phys. J. E Soft Matter.* 40:106.
38. Brandani, G. B., and S. Takada. 2018. Chromatin remodelers couple inchworm motion with twist-defect formation to slide nucleosomal DNA. *PLoS Comput. Biol.* 14:e1006512.
39. Adachi, K., W. Kopp, …, H. R. Schöler. 2018. Esrrb unlocks silenced enhancers for reprogramming to naive pluripotency. *Cell Stem Cell.* 23:266–275.e6.
40. Zhang, J., S. Ratanasirintrawoot, …, G. Q. Daley. 2016. LIN28 regulates stem cell metabolism and conversion to primed pluripotency. *Cell Stem Cell.* 19:66–80.
41. Case, D. A., T. E. Cheatham, III, …, R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.
42. Kelly, T. K., Y. Liu, …, P. A. Jones. 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* 22:2497–2506.
43. Phillips, J. C., R. Braun, …, K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
44. Maier, J. A., C. Martinez, …, C. Simmerling. 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11:3696–3713.
45. Ivani, I., P. D. Dans, …, M. Orozco. 2016. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods.* 13:55–58.
46. Roe, D. R., and T. E. Cheatham, III. 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9:3084–3095.
47. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28..
48. Pasi, M., J. H. Maddocks, …, R. Lavery. 2014. μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.* 42:12272–12283.

49. Jerabek, S., F. Merino, …, V. Cojocaru. 2014. OCT4: dynamic DNA binding pioneers stem cell pluripotency. *Biochim. Biophys. Acta.* 1839:138–154.

50. Luger, K., T. J. Rechsteiner, and T. J. Richmond. 1999. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol. Biol.* 119:1–16.

51. Taguchi, H., N. Horikoshi, …, H. Kurumizaka. 2014. A method for evaluating nucleosome stability with a protein-binding fluorescent dye. *Methods.* 70:119–126.

52. Tomschik, M., K. van Holde, and J. Zlatanova. 2009. Nucleosome dynamics as studied by single-pair fluorescence resonance energy transfer: a reevaluation. *J. Fluoresc.* 19:53–62.

53. Gansen, A., A. Valeri, …, C. A. M. Seidel. 2009. Nucleosome disassembly intermediates characterized by single-molecule FRET. *Proc. Natl. Acad. Sci. USA.* 106:15308–15313.

54. Li, Z., and H. Kono. 2016. Distinct roles of histone H3 and H2A tails in nucleosome stability. *Sci. Rep.* 6:31437.

55. Duong, V. T., Z. Chen, …, R. Luo. 2018. Computational studies of intrinsically disordered proteins. *J. Phys. Chem. B.* 122:10455–10469.

56. Ettig, R., N. Kepper, …, K. Rippe. 2011. Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping. *Biophys. J.* 101:1999–2008.

57. Dobrovolskaia, I. V., and G. Arya. 2012. Dynamics of forced nucleosome unraveling and role of nonuniform histone-DNA interactions. *Biophys. J.* 103:989–998.

58. Zhang, B., W. Zheng, …, P. G. Wolynes. 2016. Exploring the free energy landscape of nucleosomes. *J. Am. Chem. Soc.* 138:8126–8133.

59. Miao, Y., and J. A. McCammon. 2016. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Mol. Simul.* 42:1046–1055.

Biophysical Journal, Volume *118*

Supplemental Information

Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding

Jan Huertas, Caitlin M. MacCarthy, Hans R. Schöler, and Vlad Cojocaru

Supplementary Document S1:

Sequences of Widom, Esrrb and Lin28B used in the simulations.

Widom

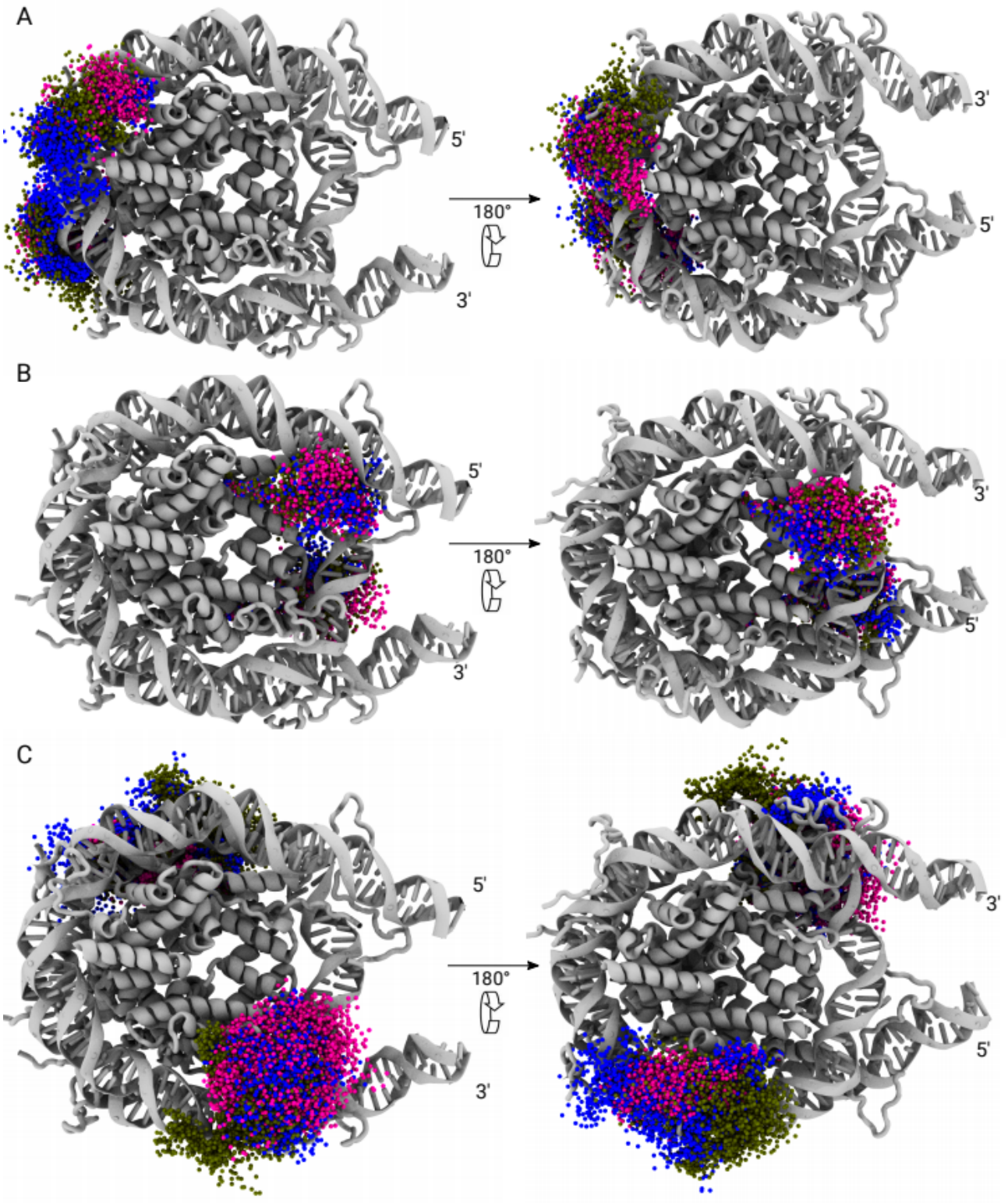
ACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCT
CTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGA
TTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGTGCATGTATTG

ESRRB

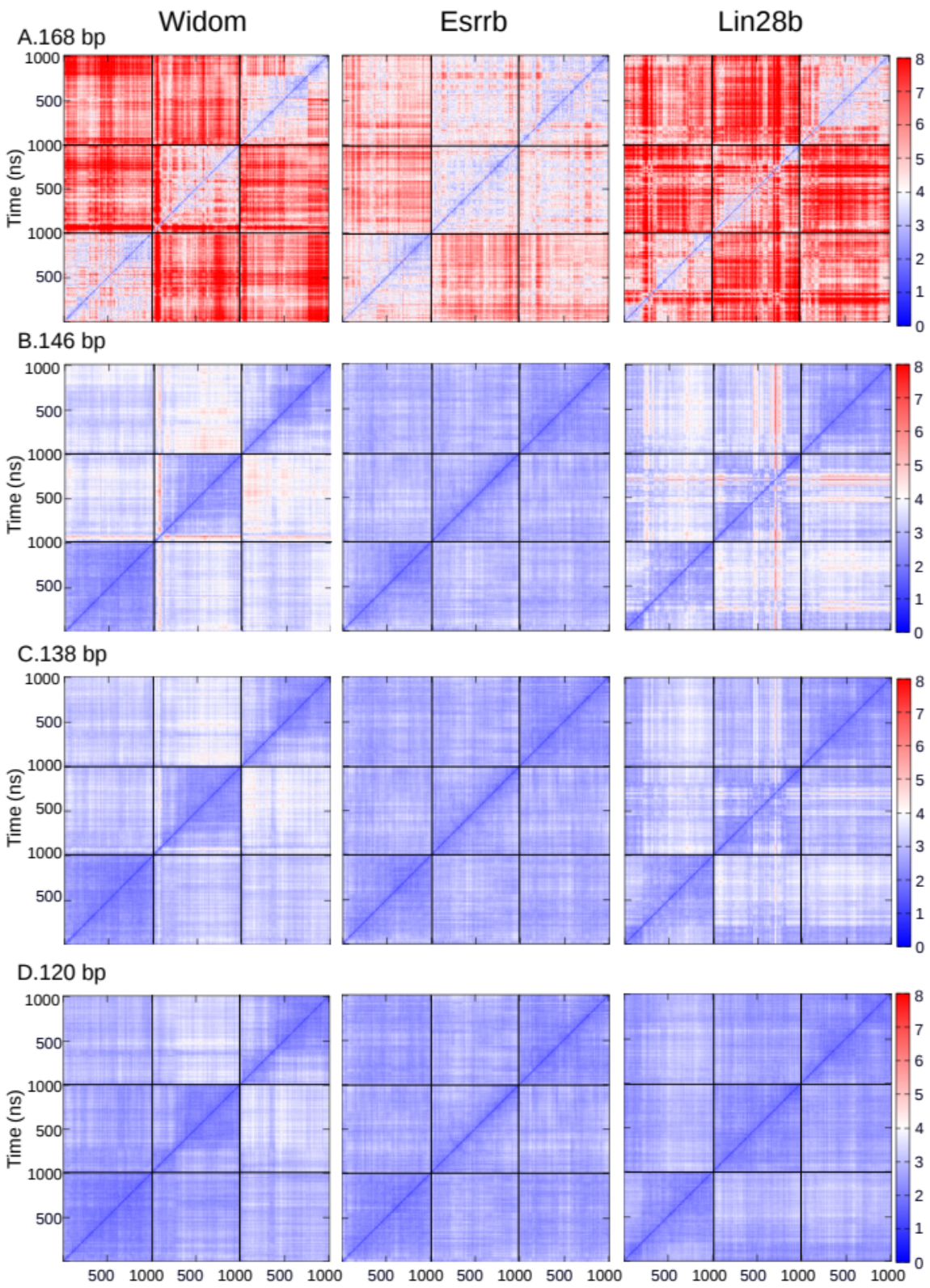
ATCAGCAGGGAGAAGGAGCGCCTCCCCATGTGGGACCTGGAGAAACAGAGGGTGGAG
GGAGCATAGAGAGTCTGTTCTAAGCTGCAAAGCAAAGGCCTGGCGACCTAGGAGACC
ATGGAGTTCCAGAAAGTGATAGTTATGCAGAGCGAATGGAGGGAATCAGCACGC

Lin28B

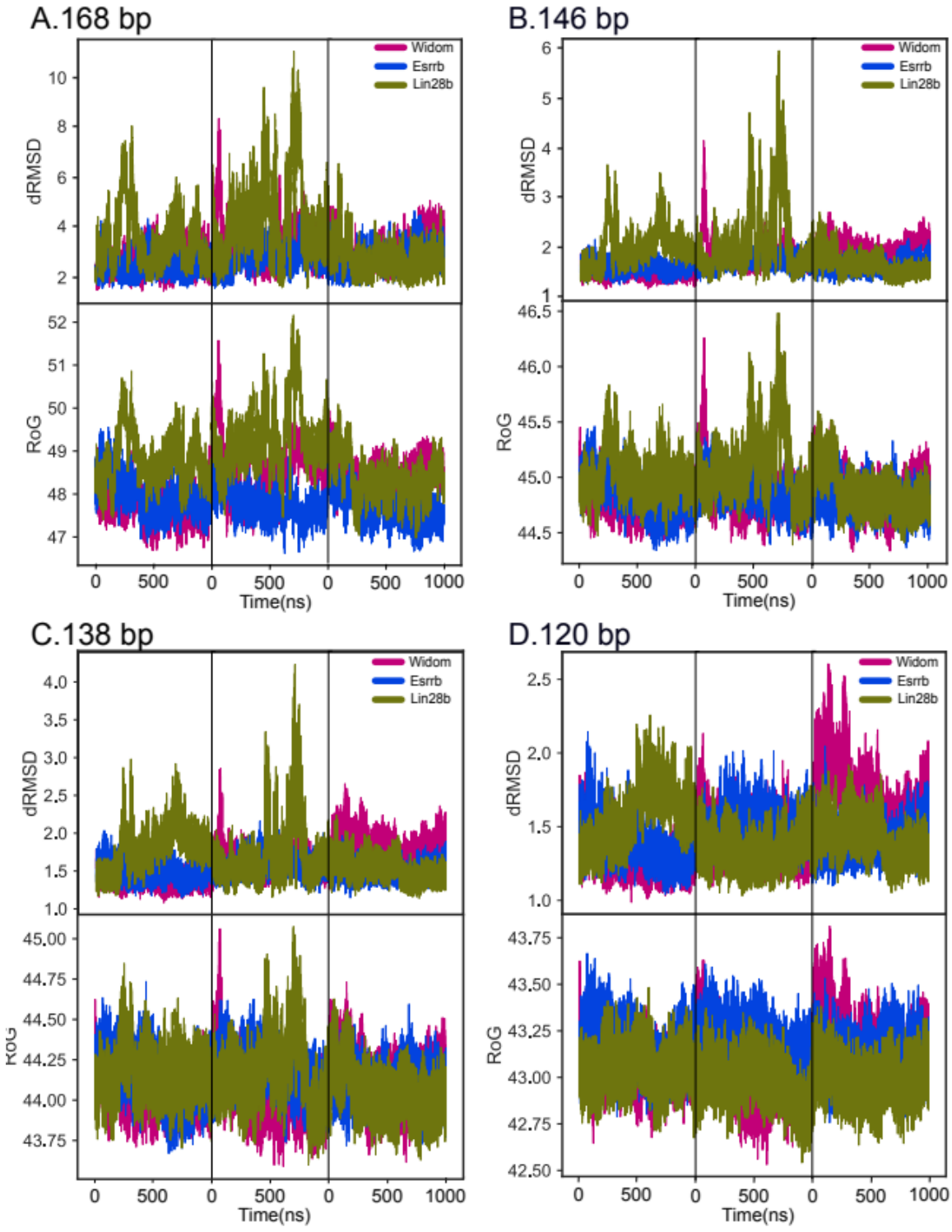
AGTTAAGTGGTATTAACATATCCTCAGTGGTGAGTATTAACATGGAACTTACTCCAA
CAATACAGATGCTGAATAAATGTAGTCTAAGTGAAGGAAGAAGGAAAGGTGGGAGCT
GCCATCACTCAGAATTGTCCAGCAGGGATTGTGCAAGCTTGTGAATAAAGACAC



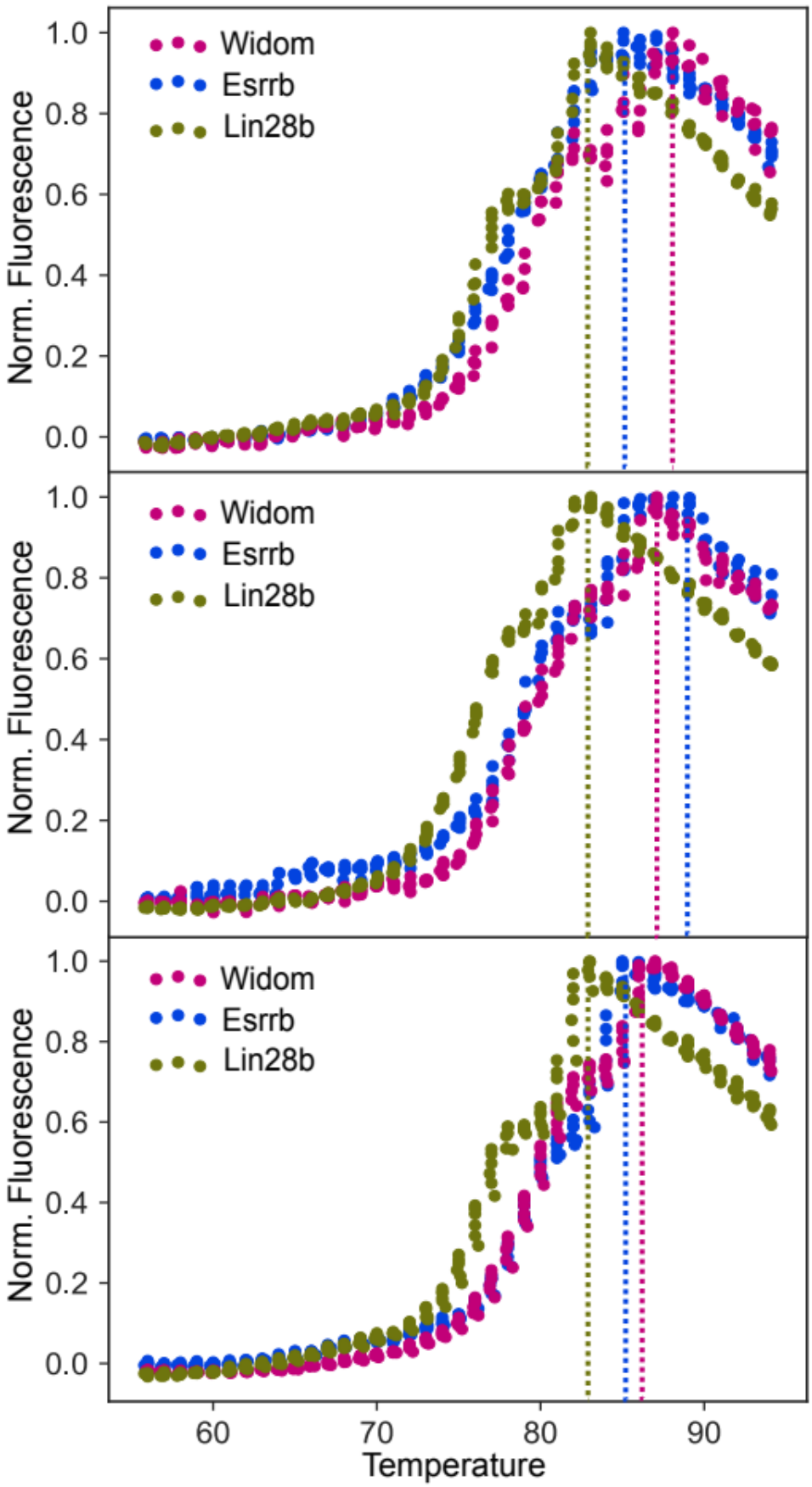
Supplementary Figure 1 (see also Figure 1 D,E). Dynamics of the core histone tails dynamics. Phase space sampling of the core histone tails in the simulation ensembles. The positions of all C_α atoms of each tail every 20 ns are shown. (A) H2A N-terminal tails. (B) H2A C-terminal tails. (C) H4 N-terminal tails. The sampling in the simulation ensembles of the 3 nucleosomes are shown in magenta (Widom), blue (Esrrb) and green (Lin28b). This coloring is maintained throughout this article.



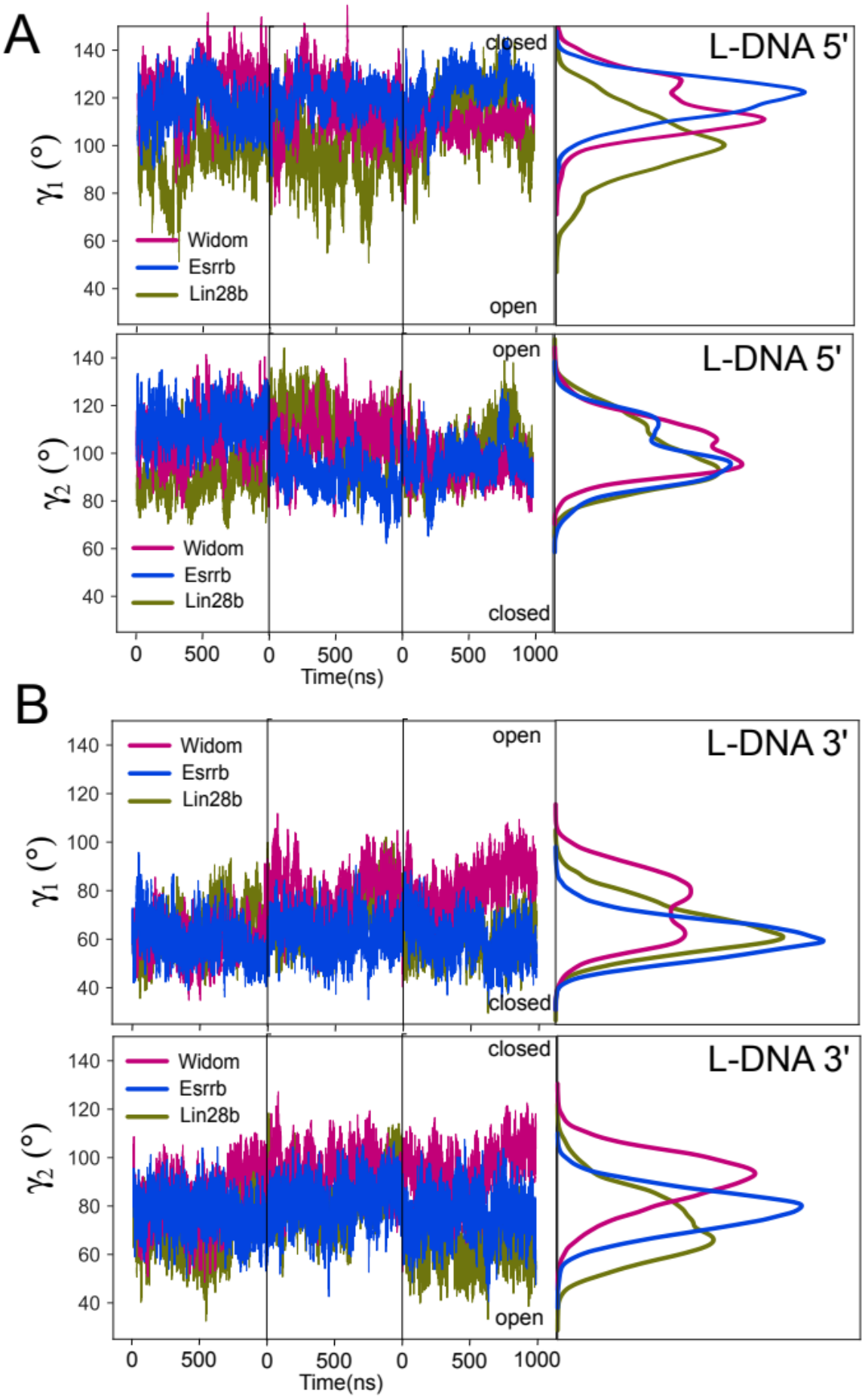
Supplementary Figure 2. RMSD matrices for different DNA regions (see also Figure 1 A- C). From the 3 μs ensemble of simulations for every nucleosome, snapshots were saved every 5 ns. After fitting the snapshots to the heavy atoms of the structured part of the histones (excluding the histone tails), the RMSD of the heavy atoms of the DNA between each 2 snaphots was calculated for the following DNA regions: (A) The entire simulated DNA (168 bp). (B) The core nucleosomal DNA (146 bp) obtained by excluding 11 bp of linker DNA at each end. (C) Nucleosomal DNA obtained by excluding 15 bp of DNA at each end (138 bp). (D) Nucleosomal DNA corresponding to the complete inner DNA turn and half of the outer DNA turn obtained by excluding 23 bp of DNA at each end (120 bp). The individual simulations are separated by black lines.



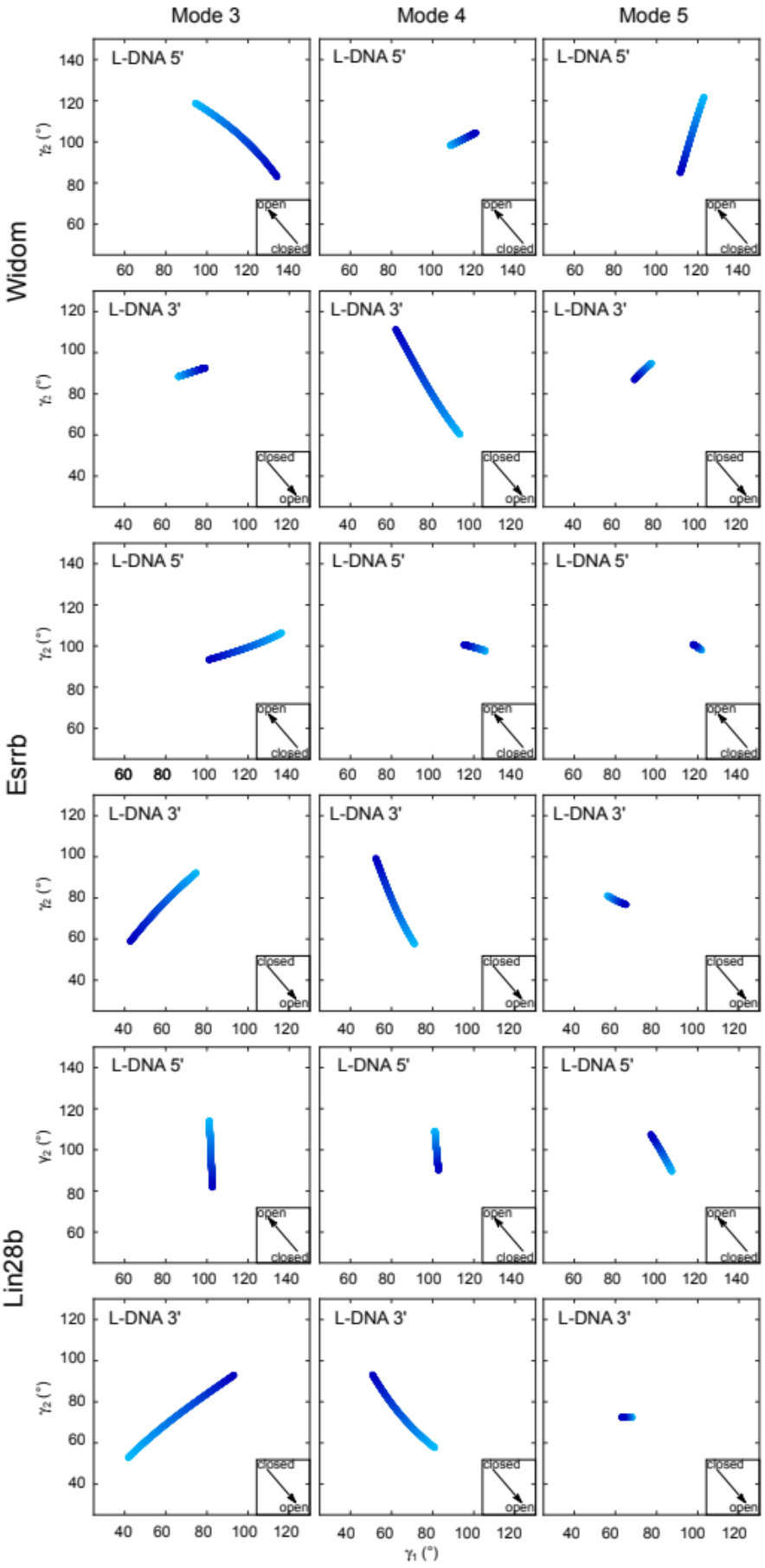
Supplementary Figure 3. Nucleosome flexibility and compaction (see also Figure 2). The evolution of the dRMSD (top) and RoG (bottom) of the DNA over time, for the three simulations of the three nulceosomes, is shown for the following DNA regions: (A) The entire simulated DNA (168 bp). (B) The core nucleosomal DNA (146 bp) obtained by excluding 11 bp of linker DNA at each end. (C) Nucleosomal DNA obtained by excluding 15 bp of DNA at each end (138 bp). (D) Nucleosomal DNA corresponding to the complete inner DNA turn and half of the outer DNA turn obtained by excluding 23 bp of DNA at each end (120 bp). The individual simulations are separated by vertical black lines.



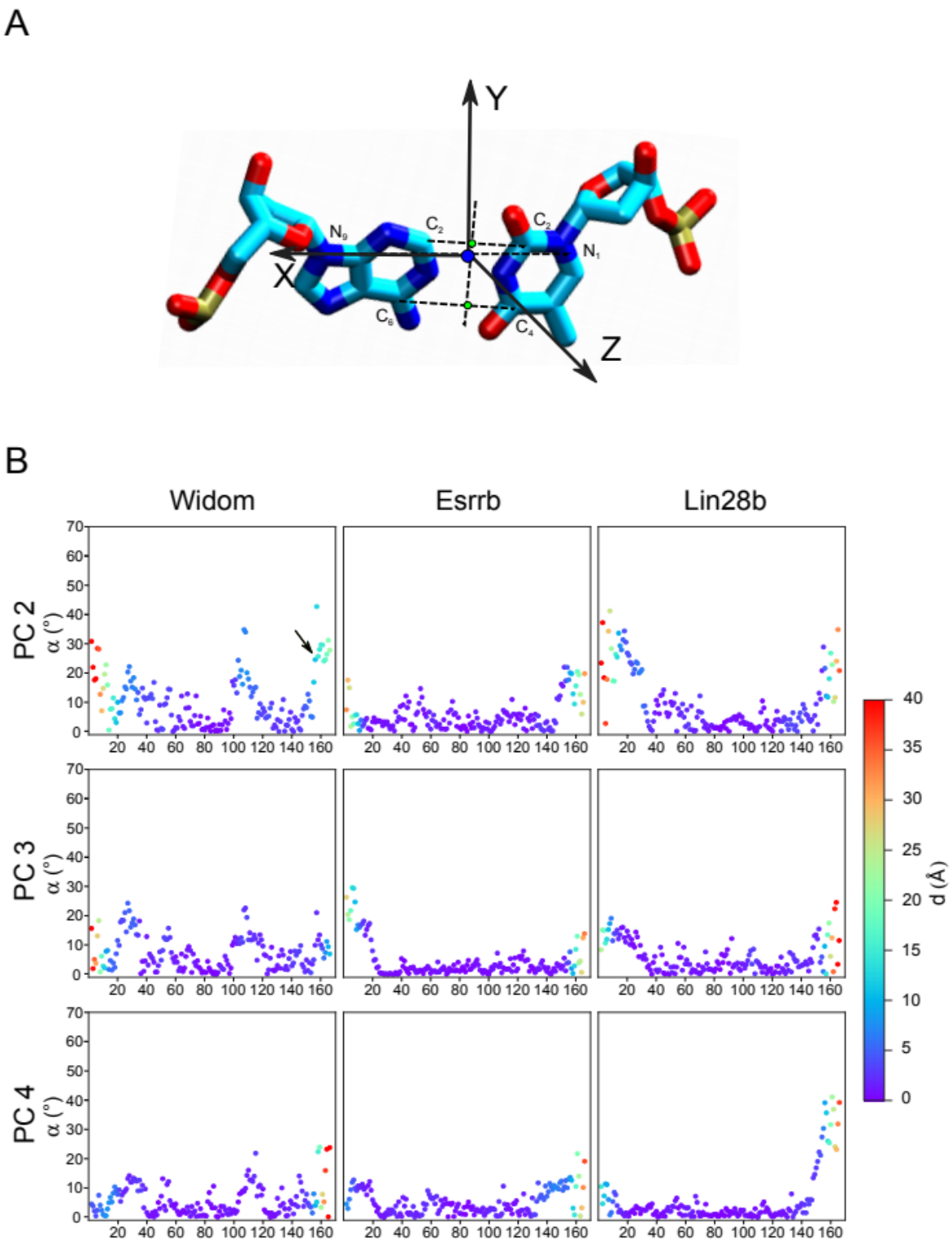
Supplementary Figure 4: Spyro Orange Thermal Stability Assay (see also Table 1). Fluorescence readings of 3 of the replicas of the Spyro Orange Thermal Stability Assay. For each sequence, the fluorescence is normalized, and the temperature at which there is a peak of fluorescence is marked with a vertical dotted line as the nucleosome dissociation temperature.



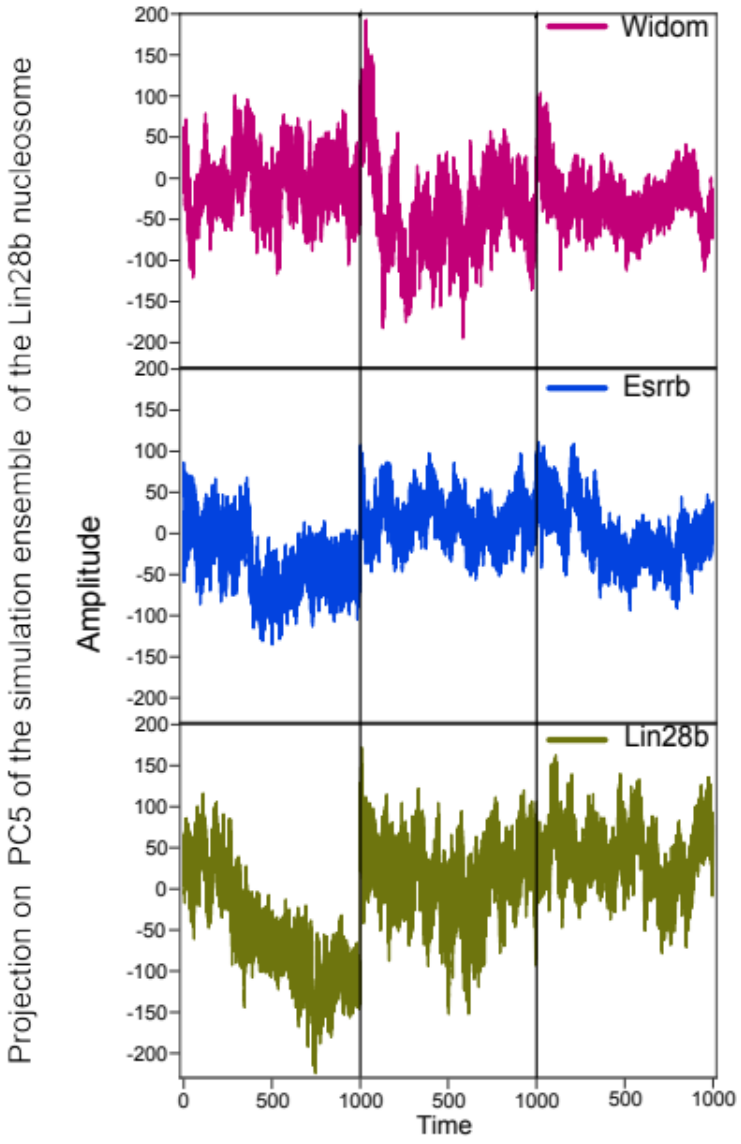
Supplementary Figure 5. Time evolution of nucleosome opening-closing motions (see also Figure 3). Time series for the γ angles described in Figure 3. Next to every plot, a histogram indicating the total sampling of the angle values during the simulations is shown (A) γ_1 of LDNA-5' (B) γ_2 of LDNA-5' (C) γ_1 of LDNA-3' (D) γ_2 of LDNA-3'. The individual simulations are separated by vertical black lines.



Supplementary Figure 6. Opening-closing motions along the first five principal components (see also Figure 4). The motion of the linker DNA arms was analyzed using the coordinate system established in Figure 3 on the pseudo-trajectories generated along modes 3 to 5. A lighter tone indicates the lowest amplitude of the mode, with darker tones corresponding to the higher amplitudes. The arrows in the square inserts indicate the direction of the nucleosome opening



Supplementary Figure 7 (see also Figure 5). Twisting motions in the nucleosomes . (A) Definition of the basepair-centered coordinate system. X was defined along the vector connecting the N_1 atom of the pyrimidine with the N_9 atom of the purine. Then, a vector connecting the geometrical center of the C_2 atoms of both bases and the geometrical center of the C_4 atoms of both bases was defined. Z was set along the cross product between this vector and X. The origin of the coordinate system is at the intersection between this vector and X which is near the basepair center. Y was set as the cross product between X and Z. (B) Twisting motions along principal components 2 to 4. The plots show the standard deviation of range the angle α, colored by the mean displacement d, calculated as defined in Figure 5 on the pseudo-trajectories generated along the first five eigenvectors of the PCA analysis for each nucleosome. The occurrence of twisting motions is indicated by black arrows.



Supplementary Figure 8 (see also Figure 5). Projection of trajectories on the fifth eigenvector of the ensemble of simulations of the Li2n8b nucleosome. The individual simulations are separated by vertical black lines.