OCT4 interprets and enhances nucleosome flexibility

Caitlin M. MacCarthy[$,1] Jan Huertas[$,1,2,3] Claudia Ortmeier[1]
Hermann vom Bruch[1] Deike Reinke[4] Astrid Sander[4] Tim Bergbrede[5]
Hans R. Schöler*[1,6] Vlad Cojocaru*[,1,2,3]

[$] These authors contributed equally

[1]Department of Cellular and Developmental Biology, Max Planck Institute for Molecular Biomedicine, Münster, Germany

[2]In Silico Biomolecular Structure and Dynamics Group, Hubrecht Institute, Utrecht, The Netherlands

[3]Center for Multiscale Theory and Computation, Westfälische Wilhelms University, Münster, Germany

[4] Max Planck Institute of Molecular Physiology, Dortmund, Germany

[5] Lead Discovery Center GmbH, Dortmund, Germany

[6]Medical Faculty, Westfälische Wilhelms University, Münster, Germany

∗ To whom correspondence should be addressed:

h.schoeler@mpi-muenster.mpg.de, v.cojocaru@hubrecht.eu

Pioneer transcription factors induce cellular identity transitions by binding to sites on inaccessible DNA wrapped in nuclear chromatin. They contribute to chromatin opening and recruit other factors to regulatory DNA sites. The structural basis of their interaction with the chromatin structural unit, the nucleosome, is still unresolved. From a combination of experiments and molecular simulations we reveal here how Oct4, the master regulator and inducer of cellular pluripotency, interprets and enhances nucleosome structural flexibility. The magnitude of Oct4 impact on nucleosome dynamics depends on the binding site position and the mobility of unstructured tails of nucleosomal histone proteins. Oct4 propagates and stabilizes open nucleosome conformations by specific sequence recognition and unspecific DNA exploration. Our findings provide a structural basis for the versatility of transcription factors in engaging with nucleosomes and have implications for understanding how pioneer factors induce chromatin dynamics.

Introduction

In eukaryotic cells, the DNA containing all the information about the cell's fate and function is packed inside the nucleus into chromatin. The nucleosome, the basic unit of chromatin, is a nucleoprotein complex formed by $\approx$ 146 base pairs of DNA wrapped around an octamer of four histones proteins (H3, H4, H2A and H2B), each present twice. Histones have a globular domain that forms the core of the nucleosome and a disordered, charged region at the N-terminus (and at the C-terminus of H2A), known as the histone tail (*1*).

Transcription factors (TFs) are site-specific DNA binding proteins that control transcription (*2*). Nevertheless, the extent of genome packing leaves a sizeable fraction of the DNA inaccessible for the majority of TF binding due to occlusion of binding sites by the histone core. Interestingly, a subgroup of TFs known as pioneer TFs (pTFs) can recognize and bind their motifs in nucleosome occupied regions of the genome (*3–5*). A recent systematic study revealed that several families of transcription factors can have pioneer activity, binding with different affinities and mechanisms to nucleosomes (*6*).

One particularly interesting pTF is Oct4, a master regulator of pluripotency (*7*). Oct4 cooperates with pTFs Sox2 and Klf4 in conversion of somatic cells to pluripotent cells during reprogramming (*8*). Structurally, Oct4 contains a DNA binding domain divided into two subdomains, the POU specific domain (POU_S) and the POU homeodomain (POU_{HD}), connected by a linker, that together recognize an octamer sequence in free DNA (*9*), and partial motifs in nucleosomes (*10, 11*). Oct4 is known to bind cooperatively with Sox2 in the genome (*12*). Recent studies suggest they may also work together as pioneers on nucleosomal DNA (*13*), where one TF binding event alters DNA-histone contacts and DNA positioning, facilitating binding of the second TF (*14–17*). However, the molecular mechanism that controls the binding of pTFs to the nucleosome, or the effect this binding has on nucleosome dynamics (and chromatin structure) is still poorly understood.

Structural studies of pTFs often use synthetic sequences, which are proven to tightly bind the histone core in one location making them well-suited for structural techniques. Research on Oct4

pTF activity using either synthetic or native sequences all reported Oct4 utilizing only half of its eight base pair binding site due to the twist of DNA and a sterical clash with the histone core, which inhibits the canonical binding of the second subdomain (*11, 15, 16, 18*).

Recently, we used molecular modeling and short molecular dynamics simulations to build models of Oct4 bound to nucleosomes on the binding sites proposed by Soufi et al. (*19*). We confirmed that the canonical configuration of Oct4 with the subdomain bound on opposite faces of DNA is incompatible with nucleosome binding and we proposed alternative binding modes. In 2020, two papers were published describing the first experimentally resolved structures of pTFs bound to a nucleosome. One suggested a mechanism by which Sox2 binds and unravels nucleosomes (*20*) and then second showed Oct4 together with Sox2 bound to nucleosome (*15*). However, in the first structure the entire outer gyres of the nucleosomal DNA were not resolved and in the second structure only the Oct4 POU_S was resolved. The histone tails were not resolved in these structures. Recently, Echigoya et al resolved the structure of a genomic nucleosome bound by Oct4 in the absence of Oct4 and revealed an Oct4 binding site different that those described by Soufi et al. (*18*). In addition, they showed that Oct4 interacts with the histone H3 and competes with the linker histone for binding. Although, they suggested that the new binding site is recognized by the POU_S subdomain, they did not reveal the structure of Oct4 bound to the new binding site. Interestingly, these studies highlighted the importance of the POU_S but not the POU_{HD} subdomain in Oct4's nucleosome binding. From molecular simulations using a reduced (coarse-grained) representation of the molecular species, Tan et al. also observed binding of Oct4 to the site proposed by Echigoya et al, however with the POU_{HD} domain (*16*).

Here we report the structural features and dynamics involved in the Oct4-nucleosome interaction at atomic resolution. We used a combination of *in vitro* experiments and atomistic molecular dynamics simulations to understand how Oct4 binds to its motif in two genomic nucleosomes from the LIN28B and the ESRRB genes. These sequences were chosen from the first data set demonstrating Oct4's pioneer function (*10*) for their importance in defining stem cell pluripo-

tency (*21–24*) and their distinct Oct4 binding sites. We demonstrate Oct4's sequence specific binding to these two native nucleosomes. In agreement with previous studies, we observed that Oct4 favors binding sites positioned near the DNA ends. Nevertheless, we show that Oct4 recognizes its binding sites using either DNA binding subdomain, rather than the POU_S alone as previously suggested. Furthermore, we establish that Oct4 requires some nucleosome structural flexibility for efficient binding. Finally, with our simulations, we were able to describe how Oct4 binding impacts the structural flexibility of the nucleosomal DNA at atomistic resolution. We found that the magnitude of this impact depends on the position of the binding site, the mobility of histone tails, and the motion of the DNA binding subdomains.

Results

Oct4 binds to specific sites on nucleosomes with either DNA binding subdomain

The native sequences used in this study reflect the diversity of Oct4 genomic binding in that one sequence contains multiple sites for Oct4 as well as other factors (Figure 1A, LIN28B. (*11*)) and the other a solitary Oct4 binding site (Figure 1B, ESRRB). Here we focused on Oct4's relative affinities between wild-type (WT) and mutated DNAs, however, we observed the affinity for the Lin28B WT nucleosome is lower than the values reported by Soufi et al. (*11*). This is likely due to the different experimental setup and the source of histones and Oct4 protein (see Methods). We first confirmed Oct4 binding by mutating key DNA bases within expected binding sites and then tested Oct4 binding in electrophoretic mobility shift assays (EMSAs) with reconstituted nucleosomes. Oct4 binding to LIN28B was still visible even after mutating previously published Oct4 sites and the proposed Sox site (*11*), (Figure 1C, "m3x"). Competition assays using a molar excess of specific or non-specific competitor showed that the interaction between Oct4 and the m3x nucleosome is sequence specific (Figure 1C, 3rd panel). This opened the possibility that Oct4 binds to additional sites on this nucleosome. A visual scan of the LIN28B sequence revealed two additional potential

homeodomain binding sites at superhelical locations (SHL) -7 and -1 (Figure 1A, HD^{-7} and HD^{-1}). Mutating all four Oct4 sites and the Sox site (m5x) reduced Oct4 nucleosome binding dramatically (Figure S1).

Furthermore, mutating each site individually revealed that each site contributes to Oct4's affinity for the LIN28B nucleosome except the Sox site and HD^{-1} sites (Figure 1C, right panel; Figure S1). It also showed that HD^{-7} is the primary Oct4 binding site in agreement with two recent studies that also reported Oct4 binding to the HD^{-7} site (*16, 18*). The sites previously proposed by Soufi et al (*10, 11*) $HD^{-4.5}$ and $S^{-1.5}$ are secondary binding sites. The HD^{-7} contains a generic homeodomain binding site with an adjacent POU_S-like half-site (ATGA, not the canonical ATGC), however, mutation of only the HD half-site reduced Oct4 binding was drastically, demonstrating Oct4 POU_{HD}-driven nucleosome binding (Figure 1C). Moreover, the sequence of the HD^{-7} and $HD^{-4.5}$ are identical, indicating Oct4's preference for the HD^{-7} position and not the DNA sequence alone (Figure 1A). Therefore, Oct4 binds to LIN28B mainly through sequence specific binding of the POU_{HD} to the HD^{-7}.

In contrast to LIN28B, ESRRB contains one clear Oct4 POU_S binding site followed by a potential POU_{HD} binding site (Figure 1B). To confirm binding, we first mutated the POU_S site, $S^{+5.5}$, and observed a complete loss of Oct4 binding (Figure 1D, second panel). The integration of a canonical homeodomain site directly 3' to the POU_S site did not restore binding, indicating that $S^{+5.5}$ is likely facing away from the histone core but the following bases may not be sufficiently exposed for POU_{HD} binding. The generation of atomistic models confirmed the exposed position of the POU_S half-site, which will be discussed in later sections. To check the potential binding of the POU_{HD} on an exposed site at SHL +5.5, we introduced a canonical homeodomain binding site in the endogenous POU_S half-site location (Figure 1D, $mS^{+5.5}$+$mHD^{+5.5}$). However, this mutation also inhibited Oct4 binding. Together these results suggest the POU_S domain drives ESRRB nucleosome binding at $S^{+5.5}$.

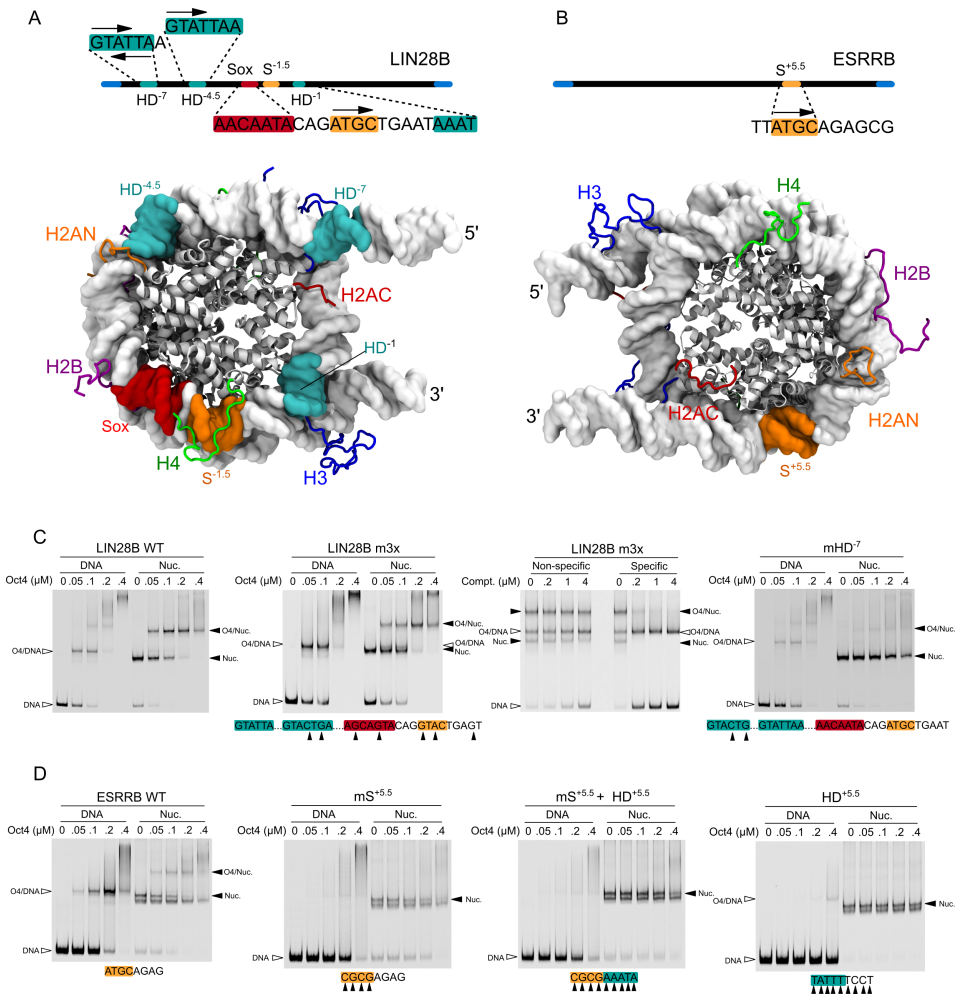

Figure 1: **Oct4 binds to native nucleosomes at different positions with either DNA binding subdomain A-B)** Schematics of the sequence, organization, and 3D positioning of the published and potential Oct4 and Sox2 binding sites on genomic DNAs LIN28B (A) and ESRRB (B). Teal and orange denote POU_{HD} and POU_S domain binding sites, respectively. Red denotes the Sox2 binding site. Arrows indicate the binding orientation. The POU_S binds in a 5'-ATGC-3' orientation to free DNA, whereas the POU_{HD} in a 5'-T(A)AAT-3' orientation with the N-terminal tail recognizing the first half and the globular part the second half. The structural views show the histone core in gray cartoons, the DNA in gray surface, the binding sites in the corresponding colors, and the histone tails in blue (H3), green (H4), orange (H2AN), red (H2AC), and purple (H2B). These representations and coloring are kept throughout the manuscript **B-C)** EMSAs of purified Oct4 with free DNA (left) and reconstituted nucleosomes (right). The third gel in 1B shows a competition assay using excess specific and non-specific competitor. Filled horizontal arrowheads indicate nucleosomes or nucleosome-protein complexes and unfilled arrowheads, free DNA or DNA-protein complexes. WT and mutant sequences are displayed below the corresponding EMSA. Mutated bases are highlighted by vertical arrowheads.

Oct4 requires nucleosome structural flexibility to bind

In a recent publication and a preprint, we reported differences in local nucleosome dynamics that extended beyond the linker DNA arms into the body of the nucleosome (*19,25*). Notably, dynamics were sequence dependent and greater in native nucleosomes than in the synthetic ones with the Widom 601 positioning sequence (*26*). We wanted to know how restricting these nucleosome dynamics would influence Oct4 binding.

To test this, we crosslinked nucleosomes, confirmed crosslinking (Figure S2A), and then incubated them with purified Oct4. For the LIN28B nucleosome, crosslinking slightly enhanced binding at low Oct4 concentrations but reduced binding at high concentrations (Figure 2A). This can be explained by the free LIN28B nucleosome adopting more open conformations (*19, 25*). Upon crosslinking, the higher affinity Oct4 binding site HD^{-7} may be fixed in a more exposed configuration, while the other sites, located on the nucleosome core, would only be occupied at higher Oct4 concentrations. Crosslinking ESRRB resulted in overall reduced Oct4 binding (Figure 2B). This demonstrates that the dynamics between the ESRRB DNA and the histone core facilitate Oct4 nucleosome binding, while binding to the LIN28B appears to be more complex due to the multiple Oct4 binding sites present.

We then wanted to know how stable the Oct4-nucleosome complexes are between different sequences or mutants. After protein-nucleosome complex formation, we added increasing amounts of unlabeled specific competitor and monitored Oct4's dissociation from the nucleosome. Notably, Oct4 detachment from the LIN28B nucleosome required 20-40 fold higher amounts of competitor and six times longer incubation than from the ESRRB nucleosome (Figure 2C, Figure S2B). These data correlate with the number of Oct4 binding sites for each sequence, the higher affinity of the POU_{HD} to free DNA comparing to POU_S (*27*), and the established dynamics of the free nucleosomes alone ((*19*))

From the data in Figure 1, we know that not all of the Oct4 binding sites contribute equally to overall affinity and HD^{-7} is the predominant binding site for Oct4 on the LIN28B nucleosome.

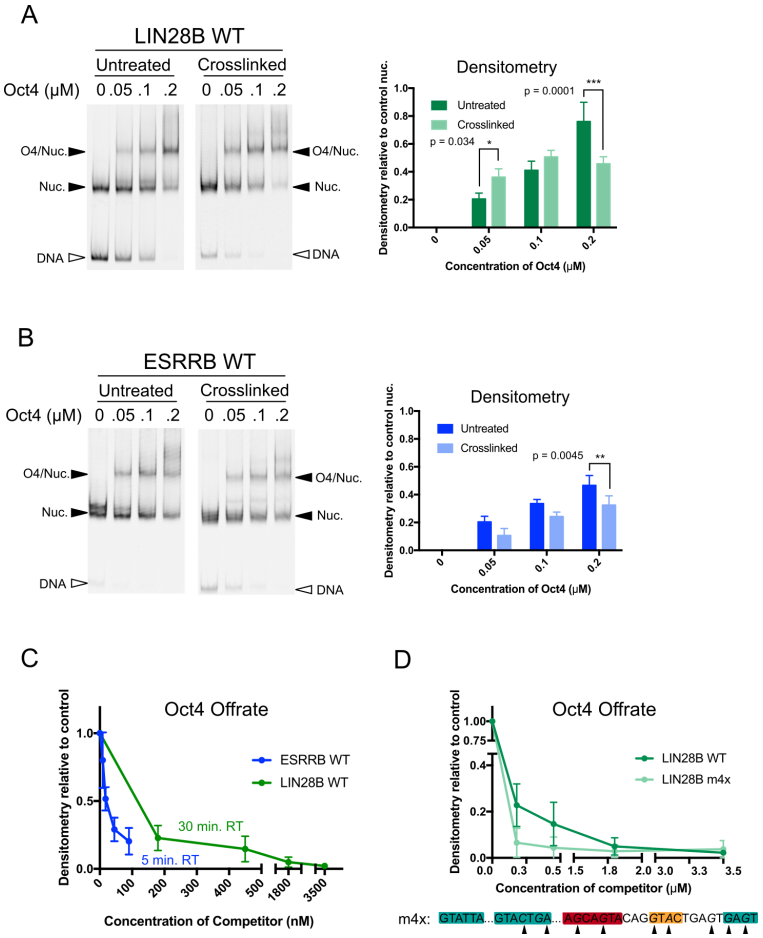

Figure 2: **Nucleosome dynamics influence Oct4 binding. A-B)** EMSAs of purified Oct4 with untreated (left) or crosslinked (right) LIN28B (A) or ESRRB (B) nucleosomes. Filled arrowheads signal nucleosomes or nucleosome-protein complexes and unfilled arrowheads point to free DNA. Bar graphs display the mean and standard deviation of densitometry relative to starting nucleosome amounts. n=3 per condition, * p=0.0344,** p=0.0045, *** p=0.0001. **C)** Densitometry of Oct4 off-rate EMSAs using LIN28B (green) or ESRRB (blue) WT nucleosomes. Values represent the mean and the error bars reflect the standard of deviation. ESRRB n=4 and LIN28B n=5. **D)** Same experiment as panel C but comparing LIN28B WT and m4x nucleosomes. m4x n=6. The scheme below shows the m4x sequence similar to those in Figure 1. Mutated bases are highlighted by arrowheads.

However, disruption of all Oct4 and Oct4-like binding sites except HD^{-7} ($HD^{-4.5}$, Sox, $S^{-1.5}$, and HD^{-1} in m4x), showed that although these sites are not the primary binding site for Oct4, they do significantly influence the affinity of Oct4 for the LIN28B nucleosome (Figure 2D, Figure S2C). In summary, we observed that Oct4 has a higher affinity for the LIN28B nucleosome in part due to the multiple binding sites and the dynamics between the DNA and the histone core. Oct4's affinity for the lone ESRRB nucleosome binding site is partially dependent on nucleosome dynamics and substantially lower than that for the more dynamic LIN28B nucleosome.

Oct4 modifies nucleosome structural flexibility

To evaluate the effects of the binding of Oct4 on nucleosome dynamics, we generated structural models of Oct4 bound to the different binding sites (Figure 3, Figure S3), performed MD simulations with the different models (Table 1), and compared the simulations with Oct4 bound to simulations of the nucleosomes alone already published (*19, 25*). The general compactness of the nucleosome was analyzed using the radius of gyration, whereas the breathing motions in two orthogonal planes were characterized using the angles γ_1/γ_2, which describe the motions in the planes perpendicular to the superhelical axis and the dyad axis, respectively (*19, 25, 28*).

The initial modelling of Oct4 bound to the sites proposed by Soufi et al. (*11*) on the LIN28B nucleosome was presented in a previous report (*19*). We now extended the simulations of the viable configurations to 1μs. The analysis of the Oct4-DNA contacts confirmed that Oct4 was bound stably on this time-scale (Table 1). When Oct4 was bound to the Soufi et al. binding sites, $HD^{-4.5}$ and $S^{-1.5}$, nucleosome breathing was not altered compared to that observed in the simulations of the nucleosome alone (Table 1, Figure S3).

When modelling Oct4 bound to the HD^{-7} site, we used conformations of the LIN28B nucleosome partially open on the 5' L-DNA from the simulations of the free nucleosome, to avoid clashes between the POU_{HD} domain of Oct4 and the other DNA gyre. Given that the sequence was identical to the one in the $HD^{-4.5}$ binding site and our previous report revealed a preference

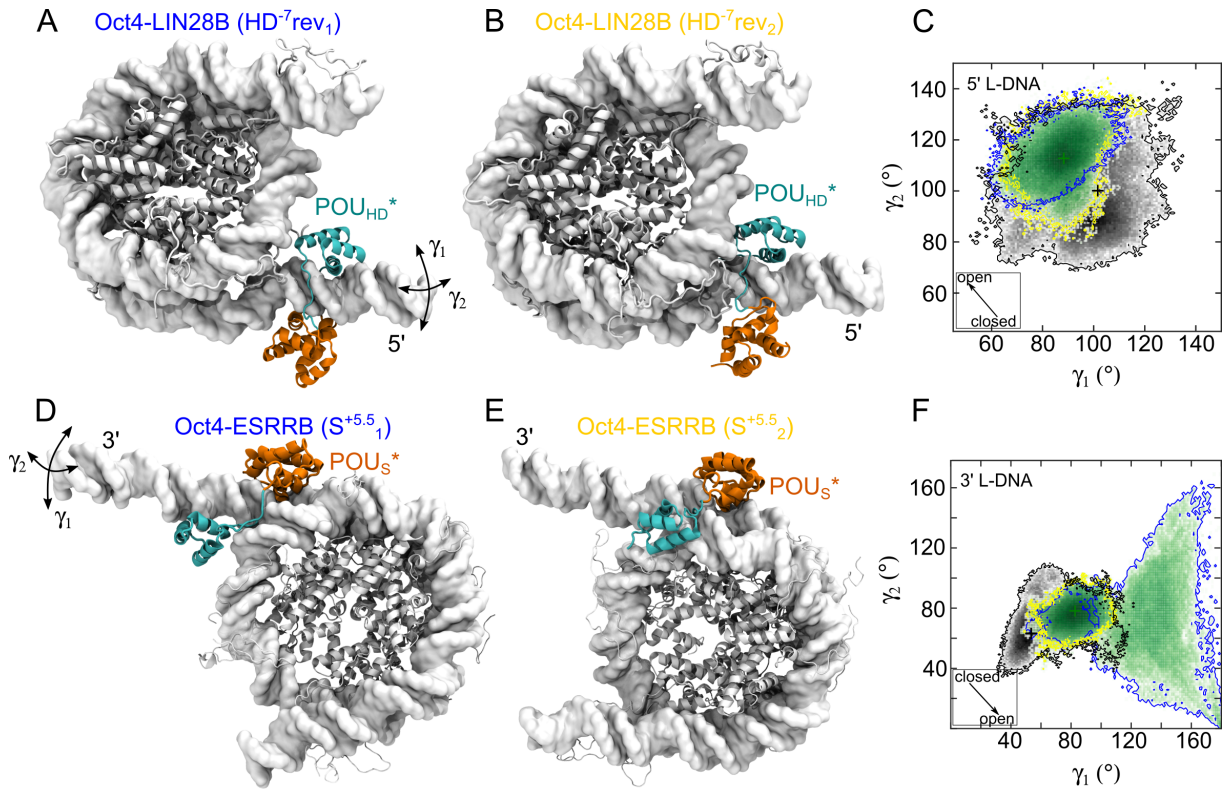

Figure 3: **Oct4 binding modifies nucleosome breathing A-B)** Representative structures from the simulations of the Oct4-LIN28B complex with Oct4 bound to the HD^{-7} binding site, in reverse orientation. The starting configurations of Oct4 were the canonical configuration in which the two subdomains bind to opposite faces of the DNA (A) and a configuration generated from a simulation of apo Oct4 (see Methods) (B). **C)** Two-dimensional histogram depicting the sampling of the γ_1/γ_2 conformational space in the simulations of the complexes from A and B. **D-E)** Representative structures from the simulations of the Oct4-ESRRB complex with Oct4 bound to the S$^{+5.5}$ binding. Two starting configurations of Oct4 were taken from the simulation of apo Oct4. **F)** Two-dimensional histogram depicting the sampling of the γ_1/γ_2 conformational space in the simulations of the complexes from A and B. The motions described by γ_1 and γ_2 are indicated with double headed arrows. In the histograms, in black is the sampling from 2 μs of simulations of free nucleosome, whereas green marks the sampling from 2μs of simulations of Oct4-Nucleoosme complexes. The sampling from each individual 1μs simulation of the complexes is in yellow and blue, respectively. The arrows in the square insets indicate the direction of the nucleosome opening. The "*" labels the sequence specific bound subdomain (also in subsequent figures)

Table 1: **Overview of the simulations performed**

Simulation name	DNA	Binding site	Oct4 configuration	Time	RoG*	Number of Oct4 - DNA contacts+			
						POU_S		POU_{HD}	
						Bases	Backbone	Bases	Backbone
$LIN28B_1$	LIN28B	-	-	1 μs	48.34-50.90	-	-	-	-
$LIN28B_2$	LIN28B	-	-	1 μs	48.00-49,91	-	-	-	-
$ESRRB_1$	ESRRB	-	-	1 μs	47.83-51.32	-	-	-	-
$ESRRB_2$	ESRRB	-	-	1 μs	47.14-48.18	-	-	-	-
$HD^{-7}rev_1$	LIN28B	HD^{-7}	Canonical	1 μs	49.30-50.35	5-28	27-99	86-126	172-250
$HD^{-7}rev_2$	LIN28B	HD^{-7}	MD	1 μs	49.58-50.60	0-10	0-73	36-81	90-172
$HD^{-7}rev_3$	LIN28B	HD^{-7}	Canonical	1 μs	49.64-50.89	33-109	63-155	83-131	147-254
$HD^{-7}rev_4$	LIN28B	HD^{-7}	MD	1 μs	48.57-49.84	2-14	23-104	40-87	132-212
$HD^{-7}for_1$	LIN28B	HD^{-7}	Canonical	1 μs	49.16-51.51	25-56	69-117	92-131	177-253
$HD^{-7}for_2$	LIN28B	HD^{-7}	MD	1 μs	49.11-50.08	0-26	10-68	60-117	116-188
$HD^{-7}for_3$	LIN28B	HD^{-7}	Canonical	1 μs	49.17-51.11	35-59	74-143	62-121	150-220
$HD^{-7}for_4$	LIN28B	HD^{-7}	MD	1 μs	49.24-51.85	3-56	32-127	42-86	57-139
$S^{-1.5}{}_1$	LIN28B	$S^{-1.5}$	MORE	1 μs	48.38-49.54	69-102	64-97	6-39	50-125
$S^{-1.5}{}_2$	LIN28B	$S^{-1.5}$	MD	1 μs	47.81-49.72	52-99	63-102	0-5	0-55
$HD^{-4.5}for_1$	LIN28B	$HD^{-4.5}$	MORE	1 μs	48.97-50.83	12-72	72-127	35-61	68-105
$HD^{-4.5}for_2$	LIN28B	$HD^{-4.5}$	MD	1 μs	47.88-49.28	0-8	3-48	71-104	117-192
$S^{+5.5}{}_1$	ESRRB	$S^{+5.5}$	MD1	2 μs	52.04-56.64	101-135	132-236	10-41	62-179
$S^{+5.5}{}_2$	ESRRB	$S^{+5.5}$	MD2	1 μs	48.28-49.75	89-121	91-155	16-60	20-88
$S^{+5.5}{}_3$	ESRRB	$S^{+5.5}$	MD1	1 μs	47.82-48.97	93-141	78-190	6-34	54-136
$S^{+5.5}{}_4$	ESRRB	$S^{+5.5}$	MD2	1 μs	48.15-49.25	55-112	104-170	3-86	40-112
$S^{+5.5}{}_5$	ESRRB	$S^{+5.5}$	MD1	2 μs	50.00-56.82	77-142	105-187	0-24	27-76
$S^{+5.5}{}_6$	ESRRB	$S^{+5.5}$	MORE	1 μs	47.60-48.87	82-111	81-146	12-64	57-190
$ESRRB_NoOct_1$	ESRRB	-	-	2 μs	55.60-58.98	-	-	-	-
$ESRRB_NoOct_2$	ESRRB	-	-	1 μs	47.54-48.66	-	-	-	-

* RoG and the number of contacts are shown as the percentiles 5 to 95.
+ A contact was defined as a non-hydrogen atom closer than 4.5 Å to another non-hydrogen atom.

for an atypical binding orientation (*19*), we modelled Oct4 in two binding orientations. In the first orientation, the POU_{HD} domain was bound to the TAAT(AC) motif, a typical homeodomain site, referred to as the reverse orientation because it's on the DNA strand oriented 3'-5' in the genome. The second orientation had the POU_{HD} binding to the GTAT(TA) motif on the genomic 5'-3' DNA strand and is referred to as the forward binding orientation. On both motifs the globular region of the POU_{HD} recognizes specifically the central AT bases. The forward orientation was the only orientation in which Oct4 was stably bound to the $HD^{-4.5}$ site in our simulations (*19*). In the case of the HD^{-7} binding site, Oct4 was stably bound in both orientations (Table 1, Videos S1, S2). We only present the data from the simulations with the reverse orientation below because it is the more typical binding orientation of homeodomains. The data regarding for the forward Oct4 binding

orientation can be found in the Supplementary Material.

Due to the location of the HD^{-7} site in the L-DNA (Figure 1), Oct4 binding in the canonical configuration with the two subdomains embracing the DNA was also possible (Figure 3A, Figure S3). To enhance the sampling of the POU_S domain non-specifically bound to this binding site, we generated models of OCT4 bound in configurations selected from MD simulations of free Oct4 (Figure 3B, Figure S3). The binding of Oct4 was stable on the 1μs timescale, regardless of the modelled configuration (Table 1).

From the simulations, we found that when bound to the HD^{-7} binding site, Oct4 stabilized the nucleosome in partly open conformations, sampled more rarely in the simulations of the free nucleosomes. Moreover, the closed conformations of this nucleosome were not sampled with Oct4 bound (Figure 3C, Figures S3 and S4) because the POU_{HD} domain, bound specifically, blocked the closing.

On the ESRRB nucleosome, the binding of Oct4 in the canonical configuration was not possible because of the binding site location in the core nucleosomal DNA. Instead, the Oct4-nucleosome interaction was modelled using two different MD-generated configurations (Figure 3D,E), as well as the MORE configuration, a configuration found in Oct4 homodimers bound to the palindromic ”MORE” sequence, that has both subdomains bound to the same face of the DNA (*29*). Regardless of the Oct4 configuration used, the binding was stable in all the simulations (Table 1).

One of the simulations of the Oct4-ESRRB complex ($S^{+5.5}{}_1$) revealed a major nucleosome opening motion, more extensive than any of the observed natural breathing of free nucleosomes (Table 1, Figure 3F, Video S3). In this simulation, the 3’ L-DNA, next to the $S^{+5.5}$ binding site, displayed large amplitude motions in both directions defined by the γ angles. On the other hand, the 3’ L-DNA adopted conformations similar to those observed in the simulations of the free ESRRB nucleosome in the other simulations (Video S4).

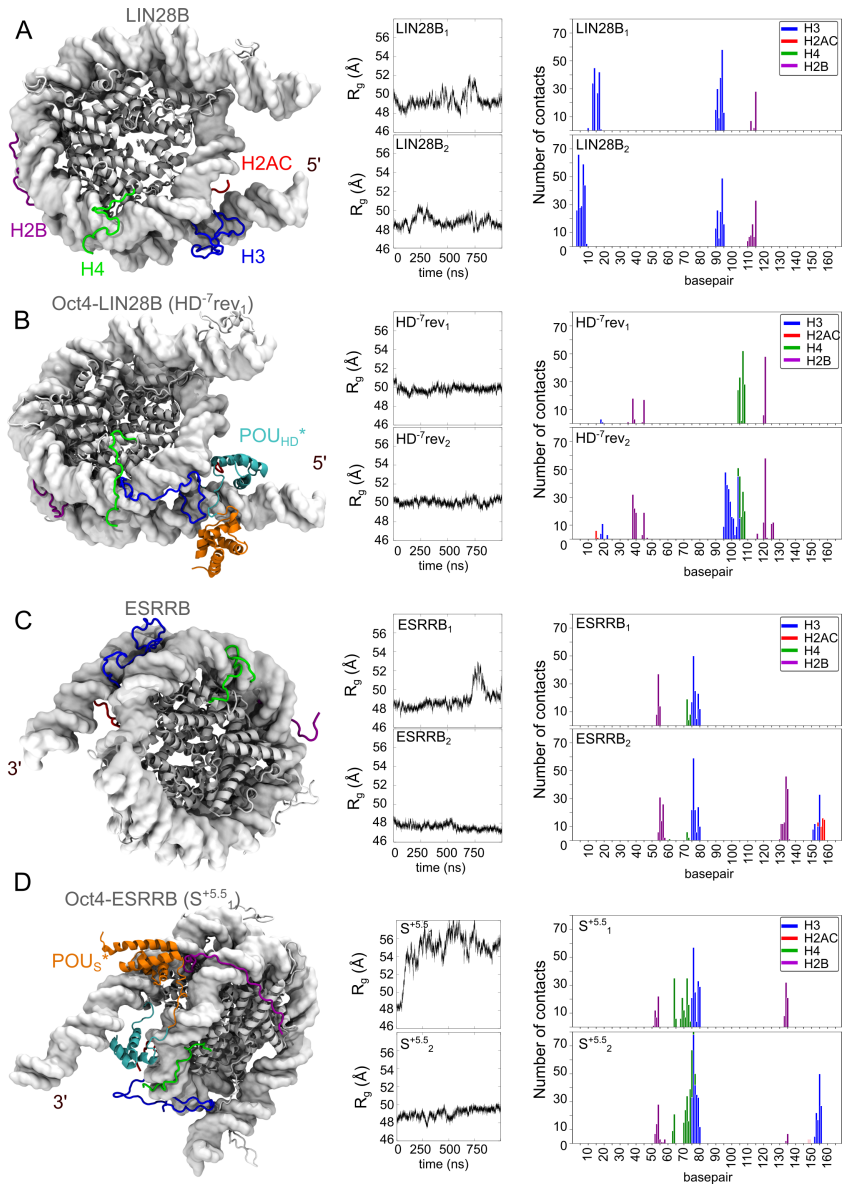

Figure 4: **Oct4 modifies the histone tails-DNA interactions**. Each panel depicts a representative snapshot of the tails on the left, a plot of the time series of the radius of gyration (R_g) in the middle, and a graph with the number of stable contacts between histone tails and the DNA on the right. A contact was defined as stable if it was present in more than the 75% of the 1μs aggregate simulation time. **A)** Simulations of the LIN28B nucleosome alone. **B)** Simulations of Oct4-LIN28B complexes with Oct4 bound in the HD^{-7} binding site in reverse orientation. **C)** Simulations of the ESRRB nucleosome alone. **D)** Simulations of Oct4-ESRRB complexes with Oct4 bound in the S$^{+5.5}$ binding site.

Oct4's impact on nucleosome dynamics depends on histone tails

Histone tails regulate inter- and intra- nucleosome dynamics (*30*). We recently reported in a manuscript currently under peer review that the positions and conformations of histone tails modulate nucleosome breathing (*25*). Therefore, we characterized how Oct4 binding modifies the interaction of histone tails with the DNA. For this, we calculated the number of stable contacts (defined as contacts present in at least 75% of the simulation) of each tail with the DNA. We focused on those tails in proximity to the HD^{-7} and $S^{+5.5}$ binding sites on LIN28B and ESRRB, respectively, thus excluding one monomer of each histone and both N-terminal tails of the H2A histones. The histone tail interaction profiles of the Oct4 bound nucleosomes were compared to those obtained from free nucleosome simulations.

In the simulations of the free LIN28B nucleosome, we observed that in the simulation in which the nucleosome opened, the H3 tail not only established fewer stable contacts with the 5' L-DNA, but the stable interactions were shifted towards the core DNA, which is formally defined as starting at base pair 12 (Figure 4A, Figure S5). As we previously demonstrated, this together with the motions of the H2AC tail, were the cause of the nucleosome opening (*25*).

When Oct4 was bound to the HD^{-7} in the reverse orientation (Figure 4B, Figure S5) almost all the H3 contacts to the 5' L-DNA were lost. This behavior is to be expected, given that the Oct4 POU_{HD} domain bound at the edge between the L-DNA and the nucleosome core, blocks the access of the H3 tail to the L-DNA. The conformation of the Oct4 linker in the canonical configuration ($HD^{-7}rev_1$ simulation) also blocked H3 tail binding near the dyad region (base pairs 90-95) (Figure 4B, right). Moreover, when Oct4 was bound to the LIN28B nucleosome, an increase in the number of stable contacts between the H4 tail and the DNA was observed.

In the simulations of the free ESRRB nucleosome (Figure 4C, Figure S5), all contacts on the 3' end were lost when the nucleosome opened (Figure 4C, right), not only for the H3 and H2AC tails but also for the H2B tail. Similarly, in the simulations of the Oct4 bound ESRRB nucleosome (Figure 4D, Figure S5), the H3 and H2AC tails established stable contacts with the DNA only

when the nucleosome remained closed and the contacts were lost when the nucleosome opened (Figure 4D, right). As in the case of the LIN28B nucleosome, Oct4 binding to the nucleosome was also accompanied by an increase of H4 tail-DNA contacts (Figure 4D, Figure S5).

In the simulations of Oct4-ESRRB complex in which the nucleosome remained closed, the H3 and H2AC tails interact both the inner and outer gyres of the DNA (Figure 5A, B). Based on these findings, we hypothesized that the large amplitude opening was observed in a single simulation of the Oct4-nucleosome complex (Figure 5C, D) because the conformational transitions of the histone tails required for the opening of the nucleosome are rare on a 1μs timescale. To test this, we performed a simulation of Oct4 bound to the $S^{+5.5}$ binding site starting from a closed nucleosome but with the H3 and H2AC tails in configurations compatible with nucleosome opening (simulation $S^{+5.5}{}_5$, see Table1). In this simulation we observed a large opening, similar to the one occurring in the original simulation (Figure 5E, F, Video S5). However, after 1μs, the nucleosome started closing (Figure 5F). When we extended the simulation for an additional 1μs, the H3 tail - DNA contacts reformed and the nucleosome closed, albeit not as tightly as in another simulation started after removing Oct4 after 1μs of the $S^{+5.5}{}_5$ simulation (Figure 5F). This demonstrates that the configurations of the H3 and H2AC tails are crucial for the initial nucleosome opening in the presence of Oct4. We recently showed that these tails modulate the intrinsic breathing of free nucleosomes (*25*). Taken together, our findings indicate that Oct4's ability to stabilize open nucleosome conformations depends on the nucleosome structural flexibility induced by histone tail mobility.

Oct4 stabilizes open nucleosomes by unspecific DNA exploration

When we visually compared the two simulations with extensive opening of the ESRRB nucleosome (simulations $S^{+5.5}{}_1$ and $S^{+5.5}{}_5$), we observed that in the original simulation ($S^{+5.5}{}_1$), the unspecifically bound POU_{HD}, relocated between both DNA gyres, preventing the nucleosome from closing (Figure 5C). This did not occur in the simulation started with the histone tails repositioned

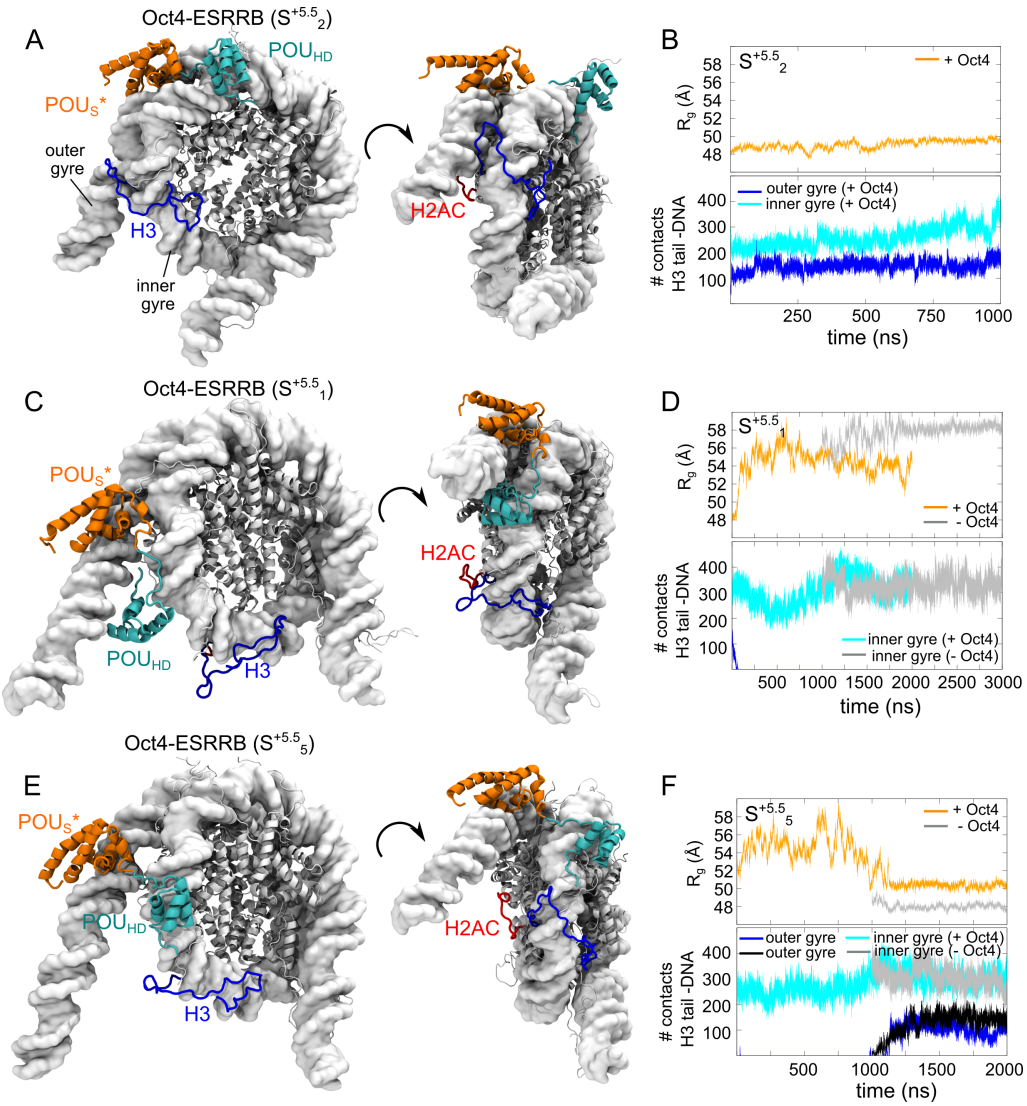

Figure 5: **Oct4 stabilizes and amplifies histone tail induced nucleosome opening** **A)** Representative structure of the Oct4-ESRRB complex in two orientations from the $S^{5.5}{}_2$ simulation in which the nucleosome remained closed. **B)** Time series of $_g$ and the number of contacts between the H3 tails and the inner (middle 128 base pairs) and outer (terminal 40 base pairs) gyres of the DNA (in blue and cyan respectively) in the $S^{5.5}{}_2$ simulation. **C)** Representative structure of the Oct4-Nucleosome complex from the $S^{5.5}{}_1$ in which the nucleosome opened. **D)** Same as B for the $S^{5.5}{}_1$ simulation. In gray are the time series from a 2μs simulation of the nucleosome after removing Oct4 started after 1μs of the $S^{5.5}{}_1$ simulation. **E)** Representative structure of the Oct4-Nucleosome complex from the $S^{5.5}{}_5$ simulation in which the nucleosome opened. This simulation was started with Oct4 bound to a closed nucleosome conformation with the H3 and H2AC tails remodelled with the configurations found in the open nucleosome conformation from the $S^{5.5}{}_1$ simulation. **F)** Same as B for the $S^{5.5}{}_5$ simulation. In gray and black are time series from a 1μs simulation of the nucleosome after removing Oct4 started at 1μs of the $S^{5.5}{}_5$ simulation.

(Figure 5E), explaining why the nucleosome eventually closed in the latter. The ample motion of the POU_{HD} stabilized a nucleosome conformation that was significantly more open that any conformation observed in the nucleosome alone simulations. The nucleosome opened even further after removing Oct4 in a 2 μs simulation started after 1 μs of the $S^{+5.5}{}_1$ simulation (Figure 5F). In contrast, in free nucleosomes any large opening was short lived. When the POU_{HD} domain didn't occupy the space between gyres ($S^{+5.5}{}_5$ simulation), the opening was still extensive but less stable. Nevertheless, the nucleosome closed faster and tighter when Oct4 was removed (Figure 5F), indicating that Oct4 has a moderate contribution to nucleosome opening independent of the histone tails.

Next, we visualized the motion of the Oct4 domains and the nature of their binding to DNA. For this, we used 4μs of aggregate time of the Oct4-LIN28B complex with Oct4 bound to the HD^{-7} binding site in reverse orientation (Figure 6A) and 5μs of aggregate simulation time of the Oct4-ESRRB complex with Oct4 bound to the $S^{+5.5}$ binding site (Figure 6B). We plotted the number of sequence specific and unspecific contacts of each domains to each gyre of the DNA.

We observed that the subdomain bound sequence specifically (POU_{HD} on LIN28B and POU_S on ESRRB) had a limited range of motion, as expected. On both nucleosomes, this subdomain also formed few contacts with the other gyre, wich had no binding sites (Figure 6C and D). On the other hand, the other subdomain explored an extensive region of the nucleosomal DNA in an unspecific manner. On LIN28B, the POU_S domain hovered only on the outer gyre (Figure 6C). On ESRRB, the POU_{HD} explored mainly the inner gyre (Figure 6C). In the simulation in which the ESRRB nucleosome opened, the POU_{HD} established additional contacts with the outer gyre after 600 ns of simulations (Figure 6C). Following this motion of the POU_{HD} domain, an increase of the number of contacts was observed. Finally this subdomain bridged the two DNA gyres, stabilizing the extensively open nucleosome conformation (Figure 6C and D, Figure S6).

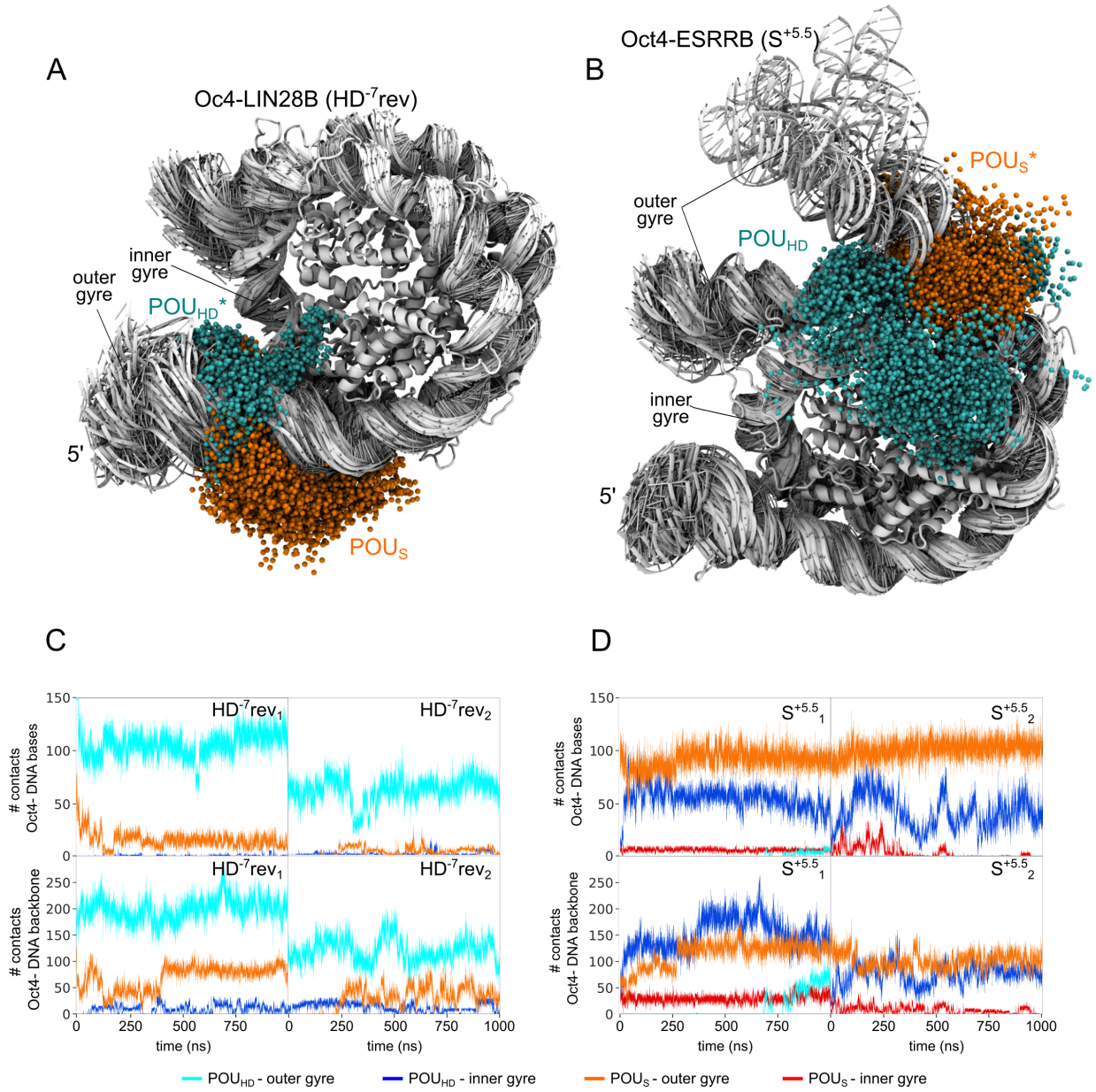

Figure 6: **Oct4 modifies nucleosome dynamics by sequence specific recognition and unspecific DNA exploration**. **A-B)** Structures showing the motions of the two Oct4 domains in the Oct4-Nucleosome simulations. The positions of the C_α atoms of the Oct4 domains are shown every 100ns and the DNA is shown every 50ns. A) 4μs of aggregate simulation time of the Oct4-LIN28B complex with Oct4 bound to the HD^{-7} binding site in reverse orientation. B) 5μs of aggregate simulation time of the Oct4-ESRRB complex with Oct4 bound to the $S^{+5.5}$ binding site . **C-D** Number of contacts of both Oct4 domains the DNA bases (sequence specific) and backbone (sequence unspecific), split between the inner and the outer gyre, in the selected two 1 μs simulations of Oct4-LIN28B (C) and Oct4-ESRRB (D) complexes. For the $S^{+5.5}{}_1$ only the first 1μs is shown

Discussion

Pioneer transcription factors are essential regulators of cellular transitions (*4, 5*). In genomic context, they recognize their binding sites in closed chromatin states, on DNA that is wrapped into nucleosomes. How they transition closed, silent chromatin to transcriptionally active DNA, sometimes using only partial binding motifs, remains poorly understood. Zhu et al published a systematic screen of TFs on nucleosomal DNA revealing a range of translational and orientation binding preferences, which strongly suggests an equally diverse range of potential pTF mechanisms (*6*). One popular mechanism proposes that for Oct4-Sox2 composite motifs, Sox2 binding to the minor groove deforms the DNA and mechanically loosens the grip of histones, freeing up buried binding sites and facilitating Oct4 binding (*15, 20*). This mechanism involves Sox2-induced DNA opening and the use of only one of Oct4's DNA binding subdomains, the POU_S domain. These structural studies provided the first evidence for a direct transcription factor-induced nucleosome opening. However, several questions remain unanswered: (i) Does the multi-domain pTF Oct4 alone have a similar impact on the nucleosome as Sox2?, (ii) How do pTFs bind native nucleosome sequences? and (iii) How do histone tails modulate or adapt to the binding?

Here we investigated the binding of Oct4 alone to two native nucleosomes: LIN28B, which contains multiple Oct4 binding sites, and ESRRB that has a single DNA binding site located in the core nucleosomal DNA. We found that Oct4 binds to LIN28B primarily on a POU_{HD} binding site at the 5' end of the nucleosomal DNA (the HD^{-7} site), but also to the POU_S and POU_{HD} sites located in the core nucleosomal DNA (*11*). We confirmed that Oct4 can use both domains to recognize its binding sites on nucleosomes, as originally proposed (*11*). Mutating only the HD^{-7} site reduced Oct4 binding substantially, whereas mutating the other binding sites had only a moderate effect on binding kinetics. These findings are consistent with previously published studies of the LIN28B nucleosome (*11, 16, 18*). Overall, these data point to Oct4's adaptability as a pTF as it is able to recognize multiple binding sites on the LIN28B nucleosome using either the POU_S or POU_{HD}.

On ESRRB, Oct4 binds to the POU_S binding site. The dissociation of Oct4 from this nucleo-

some is significantly faster than from the LIN28B nucleosome. This may be explained both by the difference in the number of binding sites between the two nucleosomes and by the higher affinity of the POU_{HD} domain to the DNA (*27*). Additionally, we showed that the binding of Oct4 to crosslinked nucleosomes is reduced at least at high concentrations of Oct4, suggesting that structural flexibility in the nucleosomes is required for optimal Oct4 binding. Genomic nucleosomes are generally believed to be more flexible than those assembled on engineered sequences (*31–33*). We also showed that both genomic nucleosomes open transiently in the μs long simulations of free nucleosomes (*19, 25*). The opening was dependent on histone tail conformations and was not observed in simulations of an engineered nucleosome sequence (Widom 601). Taken together, these findings indicate that Oct4 may use the intrinsic structural flexibility of the nucleosome to access its binding sites.

Our approach has few limitations. For structural models of Oct4-nucleosome complexes, we first modelled free nucleosomes by positioning the dyad at the center of the reported micrococcal nuclease (MNase) peak. The MNase digestion does not reveal the nucleosome position at base pair resolution, so this approximation should be considered when interpreting the models. All validated Oct4 binding sites were exposed in our models and accessible for the binding of the corresponding Oct4 subdomain. This suggests that, although we cannot exclude mispositioning of the nucleosome by a few base pairs, the approximation is in good agreement with experimental data regarding the positions of the binding sites. We forced Oct4 subdomains to form the contacts with the DNA bases observed in the crystal structure of Oct4 bound to free DNA (*9*) and first monitored the stability of the complexes in short free MD simulations. This assumes the same contacts are formed with the nucleosomes, and there is no evidence against this assumption. For the $HD^{-4.5}$ binding site, we tested two different binding orientations and reported that in one of them the binding of Oct4 was unstable (*19*). Since the sequence in the $HD^{-4.5}$ and HD^{-7} sites are the same, we also generated models for Oct4 bound in two orientations to the HD^{-7} site. However, for this binding site, the complex was stable in both orientations. Therefore, the question whether the binding orientation

of the POU_{HD} is important for nucleosome recognition remains unanswered.

We showed that Oct4 can bind to the primary binding site on LIN28B only if the nucleosome is in a partially open conformation. Oct4 stabilized such conformations, hindering nucleosome closing by the sequence specific binding of the POU_{HD} to the HD^{-7} site in the space between the core and linker DNA (Figure 7). The POU_S domain remained near the LIN28B DNA, sampling a range of positions without affecting nucleosome breathing. This suggests that one mechanism by which Oct4 impacts chromatin dynamics is by restricting the breathing of the nucleosome towards partly open nucleosome conformations. Although we did not observe an Oct4 induced opening of the LIN28B nucleosomes, we cannot exclude that this was due to the limited sampling achieved in our simulations. When binding to the binding sites located in the core nucleosomal DNA, Oct4 did not affect the dynamics of the LIN28B nucleosome. This suggests that the impact of Oct4 binding on nucleosome dynamics depends on the position of the binding site. In contrast, we observed a major Oct4-mediated opening of the ESRRB nucleosome. The open conformation was stable even after Oct4 was removed. These findings suggest that under certain conditions, the binding of Oct4 may force nucleosomes to adopt meta stable open conformations that will impact the dynamics of chromatin fibers. Interestingly, the free ESRRB nucleosome displayed the highest flexibility at the 3' end of the nucleosomal DNA (*25*), precisely in the region where the Oct4 binding site is located. This further suggests that the intrinsic nucleosome flexibility is important for Oct4 binding and its effect on nucleosome dynamics.

The major opening of the ESRRB nucleosome depended on two factors: (i) the position and conformation of the histone H3 and H2AC tails, and (ii) the motion of the POU_{HD} domain of Oct4, which explores the nucleosomal DNA in a sequence unspecific manner. The position and conformation of the H3 and H2AC tails are critical for nucleosome breathing (*25, 34, 35*). We proposed that substantial, short lived opening of free nucleosomes is only possible when these tails establish a minimum number of interactions with the linker DNAs (*25*). Here we also observed that the long lived Oct4 mediated nucleosome opening depends on the mobility of histone tails. As long as the

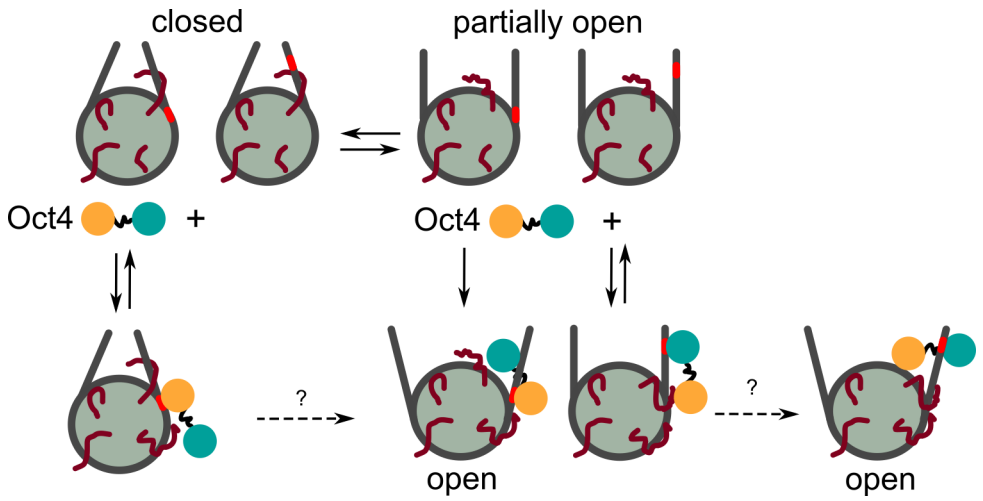

Figure 7: **Summary of Oct4 mediated nucleosome dynamics** The Oct4 POU_S and POU_{HD} subdomains are shown as orange and teal circles, respectively. Larger circles with straight lines represent nucleosomes and their linker DNA as viewed from the top in closed, partially open, or fully open configurations. The bright red points on linker DNA indicate Oct4's sequence specific binding half-sites. Dark red coiled lines represent histone tails. Dotted arrows indicate hypothesized configurations.

tails embraced the L-DNAs arms and bridged them to the core nucleosomal DNA, such an opening did not occur. The H3 and H2AC tails also influenced Oct4 binding to the LIN28B nucleosome. When Oct4 was bound to the HD^{-7} site, both tails were displaced into a configuration with fewer contacts with the L-DNA arms. Taken together, these findings suggest that any transition to a pioneer factor induced metable open nucleosome conformation relies on the mobility of histone tails regulating intrinsic nucleosome breathing. Finally, in all our simulations, the binding of Oct4 to the nucleosome was accompanied by a substantial increase of the H4-nucleosome contacts. Because the H4 tail is essential to establish nucleosome-nucleosome interactions (*36–38*), the binding of Oct4 might also alter the higher order chromatin structures. The interplay between TF binding and histone tails was also observed in the Sox2 bound nucleosome structure in which the H4 tail was displaced compared to the free nucleosome (*20*).

In addition to the role of histone tails in any TF mediated nucleosome opening, our simulations revealed distinct roles for the Oct4 domains in nucleosome binding and dynamics. One domain

is recognizing specific DNA sequences on nucleosomes while the other explores both DNA gyres in a sequence unspecific manner (Figure 7), lower half). While exploring, the domains may get trapped in positions that prevent an open nucleosome from closing (Figure 7, lower right quadrant). How often this happens in the genomic context remains unclear. We observed one such stable position of the POU_{HD} between the core and the linker DNA of the ESRRB nucleosome, which stabilized an extensively open nucleosome conformation that remained stable even when Oct4 was removed. This suggests that the pioneer factors with multiple DNA binding domains might use their domains not only for sequence specific recognition of the binding site but also to establish barriers to nucleosome closing or inter-nucleosome stacking by binding sequentially distant pieces of DNA. The relative motion of the domains in one TF is restricted by the linker between them. This may further explain why a major effect on nucleosome dynamics could be observed only for certain binding site positions. For example, on the LIN28B nucleosome, the binding sites located in the nucleosome core are too far from the linker DNA arm and thus the POU_{HD} could not move into a position to stabilize a largely open nucleosome conformation.

Finally, our study is a plea for the strength of molecular simulations in revealing the structural features and the dynamic mechanisms involved in pioneer factor-nucleosome binding. We previously showed that short MD simulations are sufficient to exclude Oct4 binding modes that are incompatible with nucleosome structure (*19*). Here we show that MD simulations not only revealed TF-induced conformational transitions in nucleosomes but also enabled us to discuss their mechanisms in detail including motions of histone tails and the Oct4 DNA binding subdomains. However, limitations remain and are mainly in the amount of sampling achieved in these simulations (*39*). Despite reporting here extensive aggregate simulation times, the sampling achieved is still not sufficient to observed multiple opening events in a single long simulation. In addition, the accuracy of the force fields used for simulating the dynamics of unstructured regions such as the histone tails is still under scrutiny (*40, 41*). Nevertheless, this approach has been successfully employed to investigate nucleosome dynamics and can be used to study the structure and dynamics

of nucleosome-protein complexes for which no experimental structure is available (*42*).

Materials and Methods

Full-length Oct4 expression and purification

M. musculus full-length Oct4 was purified. Briefly, Oct4 was cloned into the pOPIN expression vector using the SLIC method and Phusion Flash High-Fidelity PCR Master Mix (Finnzymes/New England Biolabs). SLIC reactions were then transformed into One Shot™ OmniMAC™ 2 T1® Chemically Competent *E. coli* (ThermoFisher Scientific; Waltham, MA). After sequencing, the pOPIN-cHis-Oct4 construct was co-transfected with flashBACULTRA™ bacmid DNA (Oxford Expression Technologies; Oxford, UK) into Sf9 cells (ThermoFisher Scientific) using Cellfectin II® (ThermoFisher Scientific) to generate recombinant baculovirus. Mid-log phase Sf9 cells were used to amplify the virus. Suspension High Five™ cells were infected with P3 virus for two days at 27 °C and 120 rpm shaking. After expression, crude lysates were purified on a HiTrap TALON column (GE Healthcare; Chicago, IL), cleaved on the column with 3C protease and followed by size exclusion chromatography (HiLoad Superdex 200, GE Healthcare). The final product was collected in 25 mM HEPES pH 7.8, 150 mM NaCl, 1 mM TCEP, and 5% glycerol with around 90% purity.

Nucleosome reconstitution

The LIN28B and ESRRB nucleosome sequence was taken from Soufi et al (*10*) (Hg18 chr6:105, 638,004-105,638,165 and Hg18 chr14:75,995,474-75,995,636, respectively, see also Supplementary Methods). Both genes are important for pluripotency (*22, 23*). WT and mutant sequences were purchased from IDT gBlocks with flanking AvaI restriction sites, sequenced, amplified in E. coli, digested, and finally purified by native PAGE. After Cy5 labelling, DNAs were reconstituted at DNA:octamer ratios ranging from 1:1.2-1:1.6 with purified full-length *D. melanogaster* histone octamer using the salt gradient dialysis method previously described (*43*), final buffer composi-

tion: 10 mM HEPES pH 7.6, 50 mM NaCl, 1 mM EDTA, and 0.5 mM DTT. Following dialysis, nucleosomes were heat shifted at 37 °C for 2 hours and then checked for quality and concentration by native PAGE.

Electrophoretic mobility shifts and competition assays

For binding reactions, 21 nM nucleosomes were incubated with 0.05-0.4 μM of purified Oct4 in binding buffer (25 mM HEPES pH 7.6, 50 mM NaCl, 0.5 mM EDTA, 1 μg/μL BSA, 0.8 mM DTT, and 10% glycerol) for 1 hour at 25 °C. After incubation, reactions were run directly on a 6% native polyacrylamide gel (acrylamide/bis-, 37.5/1) containing 27 mM Tris-borate, 0.6 mM EDTA, and 5% glycerol and run in the same buffer. All Cy5-labeled DNA was detected using Fujifilm FLA-9000 (GE Healthcare). Competition assays using specific and non-specific oligos were performed as previously described (*11*) using 0.2-4 μM of competitor, 21 nM nucleosome, and 105 nM Oct4. Oct4 off-rates from nucleosomes were determined by incubating specific competitor with the pre-formed Oct4/nucleosome complex that occurs after 1 hour incubation at 25 °C. Off-rate conditions were empirically determined for LIN28B and ESRRB nucleosomes due to the substantial differences in complex stability: LIN28B – 9 nM nucleosome, 45 nM Oct4, and 0.2-3.5 μM unlabeled competitor at 25 °C for 30 minutes; and ESRRB – 9 nM nucleosome, 67.5 nM Oct4, and 9-90 nM unlabeled competitor at 25 °C for 5 minutes. A short oligo containing the ESRRB Oct4 binding sequence was used as competitor for the ESRRB nucleosomes: AAGTGATAGTTATGCAGAGC-GAATGGAGGG. For LIN28B, the specific competitor sequence published in Soufi et al. was used (*11*). Experiments were performed in triplicate and densitometry was carried out using a DNA standard curve and Quantity One® software (Bio-Rad, Hercules, CA). The values reported in Figure 2 were calculated by dividing the portion of Oct4-nucleosome complex by the value of starting free nucleosome (control). Two-way ANOVA using Sidak's multiple comparison test was performed on relevant mean and standard deviation values. Statistics were done using Prism 7.0a for Mac.

Crosslinking experiments

Assemblies were generated as described in a previous section. Half of the assembly preparation was incubated with a final concentration of 1% formaldehyde on ice for 15 minutes. Crosslinking reactions were quenched by adding glycine to 250 mM final concentration. Crosslinking efficiency was checked by incubating an aliquot of crosslinked and uncrosslinked nucleosome in 15 mM $MgCl_2$ and 300 mM NaCl at 60 ° C for 17 minutes and then running the samples on a 6% native PA gel. Samples were purified on a 10–30% sucrose gradient spun at 30,000 x g and 4 ° C in a Beckman Coulter Optima L-100 XP swing bucket rotor (SW-41; Brea, CA) for 18 hours. 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). Off-rates with crosslinked nucleosomes were performed as described in the previous section.

Modelling Oct4-nucleosome complexes

For building the starting structural models of the LIN28B and ESRRB nucleosomes, we first selected 168 base pair sequences from the data reported by Soufi et al. (*10,11*), including the flanking bases of the sequences used in the *in vitro* nucleosome assembly assay. The selection was based on a comparison of ChIP-Seq data (accession code GSE36570) for Oct4 binding during reprogramming of fibroblasts to pluripotency (48 h after induction of Oct4 and the other three transcription factors required) with MNAse-seq data revealing nucleosome positioning in human fibroblasts (accession code GSM543311) (*44*). We considered the highest region of the MNAse peak to roughly correspond to the dyad and optimized the accessibility of the Oct4 binding sites proposed by Soufi et al. ($S^{-1.5}$ and $HD^{-4.5}$ on LIN28B and $S^{+5.5}$ on ESRRB) (*10, 11*) (see Supplementary Methods). Then we built the structural models by threading these selected sequences on the 168 base pairs nucleosomes with the original Widom 601 DNA sequence (*26*) and with *Drosophila* or human histones using the "swapna" function in Chimera (*45*) to swap each base according to the new sequence. The complete 168 base pair Widom nucleosomes were built using the DNA from the 3LZ0

structure with the histones (including the histone tails) modeled by homology modeling using the 2PYO and 1KX5 structures as templates (see Supplementary Methods). We performed extensive MD simulations with these nucleosomes (*25*).

Next, we modeled Oct4-nucleosome complexes using the *Drosophila* LIN28B nucleosome and human ESRRB nucleosome models built as described above (see Supplementary Methods). The initial configuration of Oct4 we took from the following structures: (i) Oct4 bound to DNA in the canonical configuration with the two subdomains bound on opposite sides of DNA obtained from the published models of Oct4-DNA complexes (*27*) that were built based on the structure of Oct4 bound as a homodimer to the PORE motif (*9*). (ii) Oct4 bound with both subdomains on the same side of DNA (MORE configuration) obtained from the structure of Oct4 bound as the homodimer to the MORE motif (*29*) by stripping one monomer. (iii) Oct4 configurations obtained from a 100 ns MD simulation of apo Oct4 (*19*) performed with the same protocol as described below.

To model the binding of Oct4 to the nucleosomes, the binding sites from the structures of Oct4 bound to free DNA were superposed to the binding sites on the nucleosomes, using only the base pairs that form sequence-specific interactions with Oct4. Then the free DNA was removed from the resulting models. The Oct4 configurations selected from the MD simulation of apo Oct4 were superposed on models of Oct4 bound to that site in canonical configuration (in which the subdomain not bound specifically clashes with the histone core (*19*)). Then we removed the canonical configuration of Oct4. All models were validated in initial 100 ns classical MD simulations (protocol below). The validating short simulations of Oct4 bound to the LIN28B nucleosome on the binding sites proposed by Soufi et al. ($HD^{-4.5}$ and $S^{-1.5}$) were presented in our previous study (*19*).

Molecular dynamics simulations

Molecular dynamics simulations were performed as previously described (*19,25*). Every simulated species was first solvated in a truncated octahedron box of SPC/E water molecules, with a layer of at least 12 Å of water around the solute. Na^+ ions were added to counter the negative charges of

the system. K^+ and Cl^- ions were added, up to a concentration of 150 mM. The systems were then optimized with an energy minimization, performed with the AMBER software (*46*). Then, the systems were equilibrated for 13.5 ns, using NAMD (*47*). The equilibration protocol was adapted from Jerabek et al. (*29*) and is described in detail in the Supplementary Methods. Harmonic distance restraints were applied to maintain DNA base pairing and Oct4-DNA base interactions. The force constant for these was gradually decreased. At the latest stages, the equilibration was restraint free. Then, we performed production simulations in NAMD, in the isobaric-isothermic (NPT, p = 1 atm, T = 300K) ensemble, with Langevin dynamics for temperature control and a Nosé-Hoover and Langevin piston for pressure control. The Li-Merz ion parameters (*48*), the ff14SB (*49*) and the parmbsc1 force fields (*50*) were used for ions, protein, and DNA, respectively. Each individual simulation was 1 μs long and multiple replicas were performed (Table 1).

Analysis of the MD simulations

All simulations were fitted to the histone core (excluding histone tails). Histone tails were defined as residues 1-45 for H3 (Human and *Drosophila*), 1-32 for H4(Human and *Drosophila*), 1-18 and 119-129 for Human H2A, 1-17 and 116-124 for *Drosophila* H2A, 1-33 for Human H2B and 1-31 for *Drosophila* H2B. For the DNA, the 146 basepairs centered on the dyad were defined as the nucleosome core, whereas the remaining 11 basepair on each side were named as linker DNAs (L-DNAs)

The characterization of the breathing motions was done using the coordinate system originally described by Öztürk et al. (*28*), also used in our previous work (*19, 25*). First, a coordinate system XYZ was defined, with the origin on the dyad. X was described as the vector along the dyad axis, Y as the cross product between X and a vector perpendicular to X intersecting it approximately at the center of the nucleosome, and Z as the cross product between X and Y. Two vectors, v_3 and v_5 were defined along the 3' and 5' L-DNas. Then , the angle γ_1 was defined as the angle between the projection of these vectors in the XZ plane and the Z axis, and γ_2, as the angle between the

projection of the vectors on the XY plane and the Y axis.

The mass-weighted radius of gyration of the DNA (RoG) was calculated using the *cpptraj* software (*51*). The same software was used for calculating the histone tail-DNA contacts and Oct4-DNA contacts. A contact was defined as a protein and a DNA non-hydrogen atoms closer than 4.5 Å. Contacts were considered stable if they were present for more than 75% of the simulation.

Acknowledgments

We would like to thank Peter Becker and Catherine Regnard at the LMU in Munich for their support by providing us with plasmids, reagents, technical assistance, and, most importantly, for their advice and active discussions. We would also like to thank Ken Zaret and Abdenour Soufi for providing us with alignments and native nucleosome sequences. 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. The MD simulations were supported 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).

References

1. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **389**, 251 (1997).

2. P. Cramer, *Nature* **573**, 45 (2019).

3. L. A. Cirillo, *et al.*, *Molecular Cell* **9**, 279 (2002).

4. K. S. Zaret, S. E. Mango, *Current opinion in genetics & development* **37**, 76 (2016).

5. K. S. Zaret, Pioneer Transcription Factors Initiating Gene Network Changes (2020).

6. F. Zhu, *et al.*, *Nature* **562**, 76 (2018).

7. S. Jerabek, F. Merino, H. R. Schöler, V. Cojocaru, *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1839**, 138 (2014).

8. K. Takahashi, S. Yamanaka, *Cell* **126**, 663 (2006).

9. D. Esch, *et al.*, *Nat Cell Biol* **15**, 295 (2013).

10. A. Soufi, G. Donahue, K. S. Zaret, *Cell* **151**, 994 (2012).

11. A. Soufi, *et al.*, *Cell* **161**, 555 (2015).

12. F. Merino, *et al.*, *Structure* **22**, 1274 (2014).

13. V. B. Teif, *et al.*, *Nature Structural and Molecular Biology* **19**, 1185 (2012).

14. V. Malik, *et al.*, *Nature Communications* **10**, 1 (2019).

15. A. K. Michael, *et al.*, *Science* **368**, 1460 (2020).

16. C. Tan, S. Takada, *Proceedings of the National Academy of Sciences of the United States of America* **117**, 20586 (2020).

17. S. Li, E. B. Zheng, L. Zhao, S. Liu, *Cell Reports* **28**, 2689 (2019).

18. K. Echigoya, *et al.*, *Scientific Reports* **10**, 11832 (2020).

19. J. Huertas, C. M. MacCarthy, H. R. Schöler, V. Cojocaru, *Biophysical Journal* **118**, 2280 (2020).

20. S. O. Dodonova, F. Zhu, C. Dienemann, J. Taipale, P. Cramer, *Nature* **580**, 669 (2020).

21. K. Adachi, *et al.*, *Cell Stem Cell* **23**, 266 (2018).

22. J. Zhang, *et al.*, *Cell Stem Cell* **19**, 66 (2016).

23. N. Festuccia, N. Owens, P. Navarro, *FEBS Letters* **592**, 852 (2018).

24. F. Kisa, *et al.*, *Stem Cell Reports* **9**, 1825 (2017).

25. J. Huertas, H. R. Schöler, V. Cojocaru, *bioRxiv* pp. 1–42 (2020).

26. P. T. Lowary, J. Widom, *Journal of Molecular Biology* **276**, 19 (1998).

27. F. Merino, B. Bouvier, V. Cojocaru, *PLoS Computational Biology* **11**, 1 (2015).

28. M. A. Öztürk, G. V. Pachov, R. C. Wade, V. Cojocaru, *Nucleic Acids Research* **44**, 6599 (2016).

29. S. Jerabek, *et al.*, *EMBO reports* **18**, e201642958 (2017).

30. K. Zhou, G. Gaullier, K. Luger, *Nature Structural and Molecular Biology* **26**, 3 (2019).

31. B. Eslami-Mossallam, H. Schiessel, J. van Noort, *Advances in Colloid and Interface Science* **232**, 101 (2016).

32. J. Culkin, L. de Bruin, M. Tompitak, R. Phillips, H. Schiessel, *European Physical Journal E* **40** (2017).

33. Y. Takizawa, *et al.*, *Open Biology* (2018).

34. Z. Li, H. Kono, *Scientific Reports* **6**, 31437 (2016).

35. K. Lehmann, *et al.*, *Nucleic Acids Research* **48**, 1551 (2020).

36. R. Collepardo-Guevara, *et al.*, *Journal of the American Chemical Society* **137**, 10205 (2015).

37. S. Saurabh, M. A. Glaser, Y. Lansac, P. K. Maiti, *Journal of Physical Chemistry B* **120**, 3048 (2016).

38. H. Ishida, H. Kono, *Biophysical Journal* **113**, 978 (2017).

39. A. Grossfield, D. M. Zuckerman, *Annual Reports in Computational Chemistry* (2009).

40. S. Rauscher, *et al.*, *Journal of Chemical Theory and Computation* **11**, 5513 (2015).

41. V. T. Duong, Z. Chen, M. T. Thapa, R. Luo, *Journal of Physical Chemistry B* **122**, 10455 (2018).

42. J. Huertas, V. Cojocaru, *Journal of Molecular Biology* p. 166744 (2021).

43. K. Luger, T. J. Rechsteiner, T. J. Richmond, *Methods in molecular biology (Clifton, N.J.)* **119**, 1 (1999).

44. T. K. Kelly, *et al.*, *Genome Research* pp. 2497–2506 (2012).

45. E. F. Pettersen, *et al.*, *Journal of Computational Chemistry* (2004).

46. D. A. Case, *et al.*, *Journal of computational chemistry* **26**, 1668 (2005).

47. J. C. Phillips, *et al.*, *Journal of computational chemistry* **26**, 1781 (2005).

48. P. Li, L. F. Song, K. M. Merz, *Journal of Chemical Theory and Computation* **11**, 1645 (2015).

49. J. A. Maier, *et al.*, *Journal of Chemical Theory and Computation* **11**, 3696 (2015).

50. I. Ivani, *et al.*, *Nature Methods* **13**, 55 (2015).

51. D. R. Roe, T. E. Cheatham, *Journal of Chemical Theory and Computation* **9**, 3084 (2013).