PNAS
www.pnas.org

Supplementary Information for

Nucleosome Allostery in Pioneer Transcription Factor Binding

Cheng Tan[†] and Shoji Takada*[,†]

[†] Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan

Corresponding Author

*Shoji Takada

takada@biophys.kyoto-u.ac.jp

Department of Biophysics, Graduate School of Science

Kyoto University, Sakyo Kyoto 606-8502 Japan

(phone) +81-75-753-4220

This PDF file includes:

Supplementary text

Figures S1 to S12

Legends for Movies S1 to S8

SI References

Other supplementary materials for this manuscript include the following:

Movies S1 to S8

Supplementary Methods

Protein AICG2+ model details. As an enhancement of the original Gō-model[1], the AICG2+ integrated the information from atomic force field into the coarse-grained protein model[2]. The potentials of the AICG2+ are defined as[2]:

$$V_{AICG2+}(R|R_0) = V_{local} + V_{\text{Gō}} + V_{exv}$$

$$= \sum_{i}^{bonds} k_{b,i}\left(b_i - b_{i,0}\right)^2$$

$$+ \sum_{i}^{1-3\ pairs} \epsilon_{loc2,i} \exp\left(-\frac{\left(d_i - d_{i,0}\right)^2}{2W_{d,i}^2}\right) + \sum_{i}^{dihedrals} \varepsilon_{loc3,i} \exp\left(-\frac{\left(\phi_i - \phi_{i,0}\right)^2}{2W_{\phi,i}^2}\right)$$

$$+ V_{loc}^{flp}(R)$$

$$+ \sum_{i}^{native\ contacts} \varepsilon_{\text{Gō},i,j}\left[5\left(\frac{r_{ij,0}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij,0}}{r_{ij}}\right)^{10}\right]$$

$$+ \sum_{i}^{nonnative\ contacts} \varepsilon_{exv}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij,cutoff}}\right)^{12}\right]. \quad (S1)$$

In the above equation the first three terms are for the bonds, the next neighboring particles, and the dihedral angles, respectively. The fourth term is a statistical flexible local potential:

$$V_{loc}^{flp}(R) = -k_B T\left[\sum_i \ln\frac{P_\theta(\theta_i|i)}{\sin\theta_i} - \sum_i \ln P_\varphi(\varphi_i|i)\right], \quad (S2)$$

where $P_\theta(\theta_i|i)$ and $P_\varphi(\varphi_i|i)$ are the residue-type dependent probability distributions of angles and dihedral angles, respectively. And k_B is the Boltzmann constant, and T is temperature. For all the parameters we used the default values in CafeMol.[3]

Electrostatic interactions. We used the Debye-Hückel model for the electrostatic interactions between charged particles. The potential is given by:

$$V_{ele} = \sum_{i<j} \frac{1}{4\pi\epsilon(T,C)} \frac{q_i q_j e^{-\frac{r_{ij}}{\lambda_D}}}{r_{ij}}, \quad (S3)$$

where λ_D is the Debye length, and $\epsilon(T,C)$ is the temperature (T) and ionic strength (C) dependent dielectric constant.

We employed the RESPAC method[4] to determine the partial charge distribution on the surface of the folded domains of proteins. Whereas for the flexible histone tails, we used ordinary integer number charge values.

Calibration of the PWMcos parameters. We followed our previously proposed scheme[5] to calibrate the two PWMcos parameters γ and ϵ' for Sox2, the POU_S, and the POU_{HD} domains, respectively. Practically, we computed the simulated dissociation constant K_d for Sox2 binding on DNA and compared it with experimental K_d.[6] By calibration, we determined $\gamma = 2.8$ and $\epsilon' = -0.01\ kcal/mol$ for Sox2 and used these values for all the following simulations. Whereas for Oct4, we separately calibrated the parameters for the two POU domains, based on experimental K_d results.[7] The determined values were $\gamma = 2.5$ and $\epsilon' = 0.1\ kcal/mol$ for the POU_S, and $\gamma = 4.0$ and $\epsilon' = 0.2\ kcal/mol$ for the POU_{HD}.

Scaling of inter-protein Gō-type interactions. We used the default value of $\varepsilon_{G\bar{o},i,j}$ in equation (S1) for all the intra- and inter-protein native contacts, except for the interaction between Sox2 and Oct4, and the interactions among histones. For the $G\bar{o}$-interactions between Sox2 and Oct4, we calibrated the energy coefficient by matching the simulated DNA-binding cooperativity to the experimental results.[8] To investigate possible disassembly of nucleosome, we assigned weakened energy coefficients for the $G\bar{o}$-type interactions between the (H2A-H2B)-dimer-(H2A-H2B)-dimer and the (H2A-H2B)-dimer-$(H3\text{-}H4)_2$-tetramer interfaces. Practically, we used $1.5\varepsilon_{G\bar{o},i,j}$ as the force coefficient for the Sox2-Oct4 native contacts, which were modeled from the PDB 1O4X (a ternary complex structure for Oct1-Sox2-DNA). Whereas for the contacts between the H3-H4 tetramer and the H2A-H2B dimers, and the contacts between the two H2A-H2B dimers, we used $0.3\varepsilon_{G\bar{o},i,j}$.

Definition of bending angle of DNA in the simulations of Sox2-naked DNA binding. The simulation of Sox2 binding on naked DNA was performed on a 29bp dsDNA (see Table I in the main text for DNA sequence). We first chose three points on DNA, A (center of mas (COM) of DNA base pair at index 1), B (COM of DNA base pair at index 10, the center of Sox2 consensus sequence), and C (COM of DNA base pair at index 29). The bending angle η of DNA was then defined as the angle between two vectors $\overrightarrow{BA}$ and $\overrightarrow{BC}$.

Sequence of the 601-nucleosome. We used the 147-bp 601-nucleosome as template for Sox2 and Oct4 binding. The sequence of the original 601 and the ones with insertion of Sox2 motif "CTTTGTT" can be found in Fig. S1. For Oct4, we inserted the Oct4 target motif "ATGCAAAT" using the same way as for Sox2.

Sequence of the *LIN28B*-nucleosome. We used the 177-bp *LIN28B*-nucleosome in simulations for Sox2 and Oct4 binding (Fig. 5 and 6 in the main text). The DNA sequence fragment corresponds to the genome region of hg18-chr6 105,637,996-105,638,172:

ATAAGTTAAGTGGTATTAACATATCCTCAGTGGTGAGTATTAACATGGAACTTACTCCAACAATACAGATGCT
GAATAAATGTAGTCTAAGTGAAGGAAGAAGGAAAGGTGGGAGCTGCCATCACTCAGAATTGTCCAGCAGGGAT
TGTGCAAGCTTGTGAATAAAGACACATACTT

In the above sequence, we highlighted the consensus sequence of Sox2 ("AACAATA" at 59—65) and the POU_S domain ("ATGC" at 69—72) with cyan and blue, respectively. In addition, a pseudo motif for Sox2, "GATTGTG" at 144—150, was shown in yellow. As for the POU_{HD} domain, we highlighted several motifs that are only 1-base different from the consensus motif "AAAT" (see Fig. 1). These POU_{HD} motifs were either shown in green or underlined, which were "ATTA" at 15—18, "ATTA" at 39—42, "CAAT" at 61—64, "GAAT" at 74—77, "ATTG" at 131—134, "ATTG" at 145—148, and "AAAG" at 164—167.

Definition of the exposure extent of nucleosomal DNA minor or major groove. We first determined all the "surface" particles in the nucleosomal DNA. We put the whole nucleosome into a 3-dimentional grid space, with 8.0Å padding on each direction $(\pm x, \pm y, \pm z)$. The grid had a gap length of 2.0Å. We first determined the "outer" grid points by removing those overlapped with nucleosomes. Then we put a small sphere with radius of 4.0Å on the outer grid points to "detect" the nucleosome particles facing outward. If a phosphate P_i is within 4.0Å of any outer grid point, we set a coordinate $u(P_i)$ to be 1 (facing outward), otherwise 0 (facing inward). We then used a coordinate ϕ_{minor} to describe the minor groove exposure extent of base-pair at index m, which was defined in the following:

$$\phi_{minor}(m) = u(P_{m+2}) \cdot u(P_{m'+2}),$$

where P_{m+2} is the $m+2$nd phosphate on the same DNA stand, and $P_{m'+2}$ is the $m'+2$nd phosphate on the complementary strand (see Fig. S2A and S2B). As an example, we showed an example of the 601-nucleosomal DNA colored by ϕ_{minor} (1 as white and 0 as gray) in Fig. S2C. The one-dimensional plots were used in main text Fig. 3 for the 601-nucleosome and Fig. 5 for the *LIN28B*-nucleosome.

Similarly, we used a coordinate ϕ_{major} to describe the major groove exposure, which was defined by taking the opposite of the minor groove exposure extent (ϕ_{minor}):

$$\phi_{major}(m) = \begin{cases} 1, & \phi_{minor}(m) = 0 \\ 0, & \phi_{minor}(m) = 1 \end{cases}$$

Definition of the binding energy score. We computed the sequence-dependent binding energy score γ_{PWM} of a given sequence s in the following way:

$$\gamma_{PWM}(s) = -k_B T \log\left(e^{-\frac{E_{PWM}(s)}{k_B T}} + e^{-\frac{E_{PWM}(s')}{k_B T}}\right),$$

where s' is the complementary sequence of s, and $E_{PWM}(s)$ is defined as:

$$E_{PWM}(s) = \sum_{n,b} e_n(b) S(n,b),$$

where $e_n(b)$ is the element of the position weight matrix (PWM) and $S(n,b)$ is the element of the sequence matrix. $S(n,b)$ is 1 if the base at position n is of type $b \in \{A, C, G, T\}$, and 0 otherwise.

Bootstrapping error estimation in motif binding probability analysis. For the Sox2-BSn-nucleosome ($n = 95, 96, \ldots, 104$) systems, we performed bootstrapping to estimate the accuracy of the binding probability analysis (Fig. 3A in the main text). As described in the Methods, we carried out 20 independent 3×10^7-step MD simulations for each Sox2-BS n-nucleosome ($n = 95, 96, \ldots, 104$). For each system, we then randomly chose 10 MD trajectories to calculate the Sox2-consensus motif binding probability (P'). We repeated this procedure for 100 times and computed the standard error σ_P for the 100 P' results. The σ_P was then used to plot the error bars in Fig. 3A.

***LIN28B* nucleosome spontaneous sliding simulations.** For the spontaneous sliding of the LIN28B nucleosome, we carried out 64 independent 4×10^8-step MD. These simulations were started from 16 different initial structures with rotational positioning from $u = 98$ to $u = 113$. For each initial structure we conducted 4 independent runs (in total $4 \times 16 = 64$). For these simulations, we abandoned the first 10^8 steps and only used the remaining data for analysis. We then divided all the remaining data into two parts, 1.0×10^8 to 2.5×10^8 and 2.5×10^8 to 4.0×10^8 steps (Fig. S6A). We computed the probability distributions of u in these two data sets (Fig. S6B and S6C). Finally, we used the data from 1.0×10^8 to 4.0×10^8 to plot the probability distribution ($P(u)$) shown in Fig. 5A in the main text. We also used bootstrapping to get the error estimation for $P(u)$. Practically, we randomly chose 32 trajectories, from which we calculated a temporary probability distribution P'. We repeated this procedure for 100 times and get the standard error σ_P and used it to plot the error bars in Fig. 5A.

We also used the web platform "nuMap"[9,10] to predict the positioning of the LIN28B nucleosome. We input the central part of the LIN28B nucleosome and got the prediction as a score function of dyad position. For convenience, we translated the results into a function of our sliding coordinate u and plotted it in Fig. S6D. As can be seen, the nuMap prediction is consistent with our MD simulation result, especially around the peak ($u = 104$).

All-atom simulation of the POU_S domain binding at the acidic patch of the H2A-H2B dimer. As described in the main text, we found that in the CG simulations of Oct4 binding on the nucleosome, the POU_S domain had a high probability to bind at the acidic patch on the H2A-H2B dimer (Fig. 4). To testify this finding, we conducted all-atom MD simulations. For computational efficiency, we only considered the POU_S-H2A-H2B ternary complex. We first superimpose the atomistic structures (PDB 3L1P for the POU_S domain of Oct4 and PDB 1KX5 for the histone H2A-H2B dimer) onto a representative snapshot taken from our CG simulations (see Fig. S11A). Then we solvated the POU_S-H2A-H2B complex into a cubic water box of size $126.6 \times 126.6 \times 126.6$ Å^3 (Fig. S11A). The system was then neutralized by adding Na^+ and Cl^- ions. Additional ions were added to set the ionic concentration to be same as the CG simulations (200mM). The whole system contained 199137 atoms. We used the AMBER99SB force field[11] and the TIP3P water model for MD simulation. The whole system was first energy minimized and equilibrated at 300K with 1ns MD in NVT ensemble and 200ps MD in NPT ensemble, with position restraint applied on all protein atoms. We then equilibrated the system for 2ns when restraining the H2A-H2B dimer. Finally, we carried out the production run at 300K using the NVT ensemble for 50ns (Fig. S11B, C, and D). For computational efficiency, we also applied position restraints on the H2A-H2B dimer during the production run, whereas the POU_S domain was free to move. Constraints were applied on all hydrogen-involved bonds using the LINCS algorithm. A distance of 1.0 nm was used as the short-range electrostatic cutoff and the Particle Mesh Ewald method was employed for long-range electrostatics. All the all-atom MD simulations and data analysis were performed with the MD package Gromacs (version 2020.2).[12]

Test simulations of Sox2-BS102-nucleosome at temperature 310K or ionic strength 410mM. To verify our models and methods at more physiological environments, we used the Sox2-BS102-nucleosome system as an example and performed simulations at the temperature of 310K. The initial structure was taken from the simulations at 300K, with Sox2 binding at its consensus motif (Fig. S12A). We calculated the radius-of-gyration of the nucleosomal DNA ($R_{g,DNA}$, in the unit of Å) and histones ($R_{g,histone}$, Å), the binding position of Sox2 on DNA (b_{Sox2}, DNA index), the energy of sequence-

specific interactions (E_{PWMcos}, kcal/mol) and nonspecific electrostatic interactions between Sox2 and DNA (E_{ele}, kcal/mol), and the bending angle of DNA at the Sox2 binding motif (γ, defined in Fig. S4A) (Fig. S12B). We compared the probability distribution of the R_g values with those from the simulations at 300K (Fig. S12C and D).

Some early experiments suggested that the salt concentration inside cell nuclei can be as large as ~410mM.[13,14] However, later some other experiments showed lower estimation of the cation content to be around 200-250mM.[15,16] In the main text we have shown the simulation results conducted at the condition of 200mM ionic strength. Here we also tested the 410mM concentration using the Sox2-BS102-nucleosome system as an example. Same as described above for the 310K simulations, we also monitored the R_g of histones and DNA, the interaction energies, and the bending of DNA (Fig. S12E-H). We found that at 410mM, nucleosome underwent more significant partial unwrapping than at 200mM, which might contribute to higher TF accessibility at the entry/exit region (Fig. S12F and H). However, the high salt concentration did not alter the sequence-specific interaction between Sox2 and DNA, as well as the bending of DNA (Fig. S12E).

Supplementary Figures

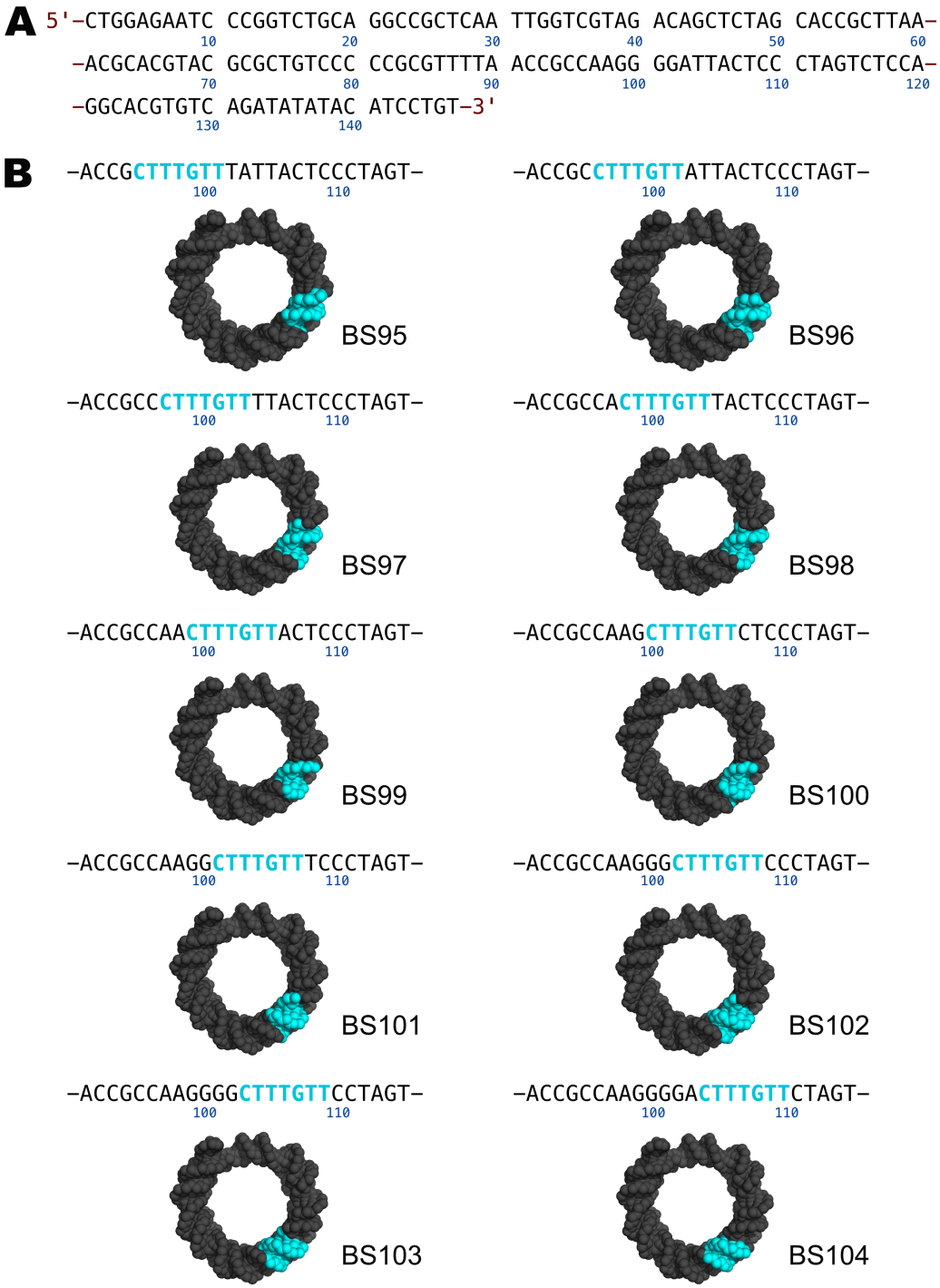


Figure S1. Sox2 consensus motifs inserted into the template of 601-nucleosome. (A) The 601-nucleosomal DNA sequence. (B) The consensus motif of Sox2 ("CTTTGTT") was inserted into different positions of the template nucleosome. For every system the DNA sequence from indices 91-115 is shown on the top, and the nucleosomal DNA structure is shown in the below with the consensus sequence colored in cyan. Histones are hidden for clarity. Each system is named as "BS*n*", where *n* is the index of the insertion starting point.

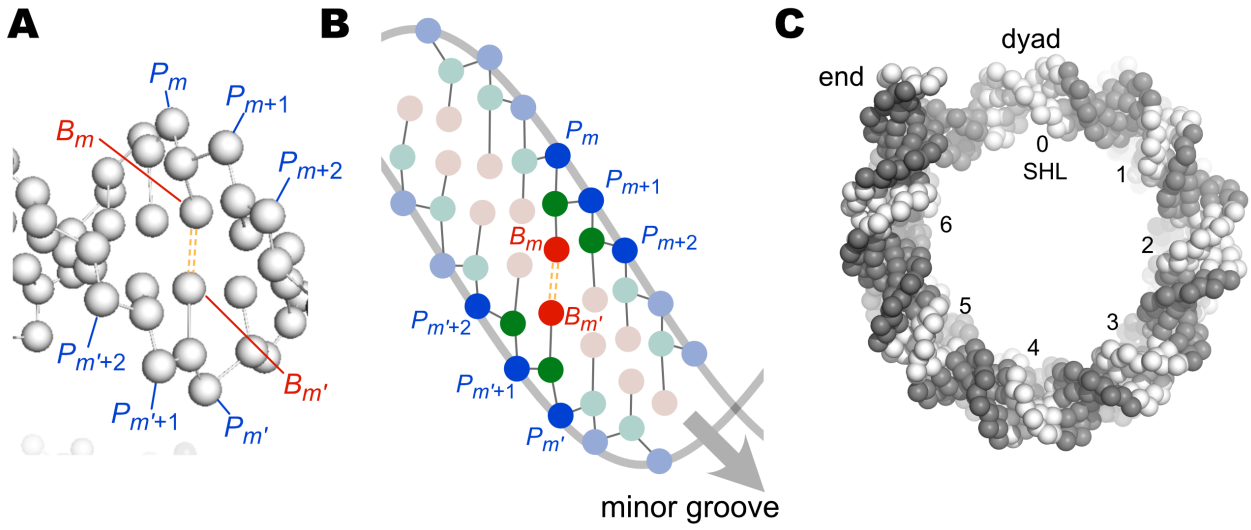


Figure S2. Determination of the exposed minor groove base pairs. (A) Local structure of B-type DNA around base pair $B_m - B_{m\prime}$. Phosphates are labeled P_m, P_{m+1}, and P_{m+2} for nucleotides in the same strand as base B_m, whereas those in the complementary strand are labeled as $P_{m\prime}$, $P_{m'+1}$, and $P_{m\prime+2}$. When both P_{m+2} and $P_{m'+2}$ are defined as surface particles, the $B_m - B_{m\prime}$ pair is defined as exposed. (B) Cartoon representation of (A). (C) Exposure extent of the nucleosomal DNA minor groove. The exposed minor groove regions are shown in white ($\phi_{minor} = 1$) and the buried minor groove regions are shown in dark gray ($\phi_{minor} = 0$). "SHL" is short for "super helical location". In Fig. 3 and 5 in the main text, we show the ϕ_{minor} values as a function of DNA indices, using the same color scheme as here.

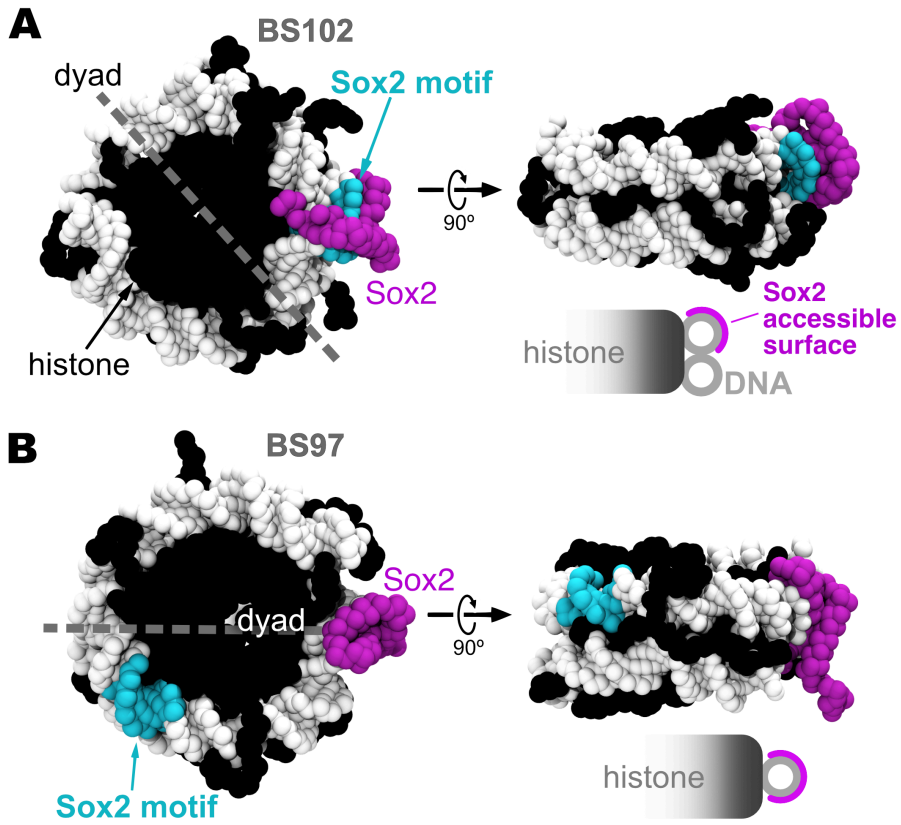


Figure S3. The sequence-specific and sequence-nonspecific binding of Sox2 on the BS102 (A) and BS97 (B) nucleosomes, respectively. Histones are shown as black balls. DNA is shown as white balls, except the Sox2 motif in cyan.

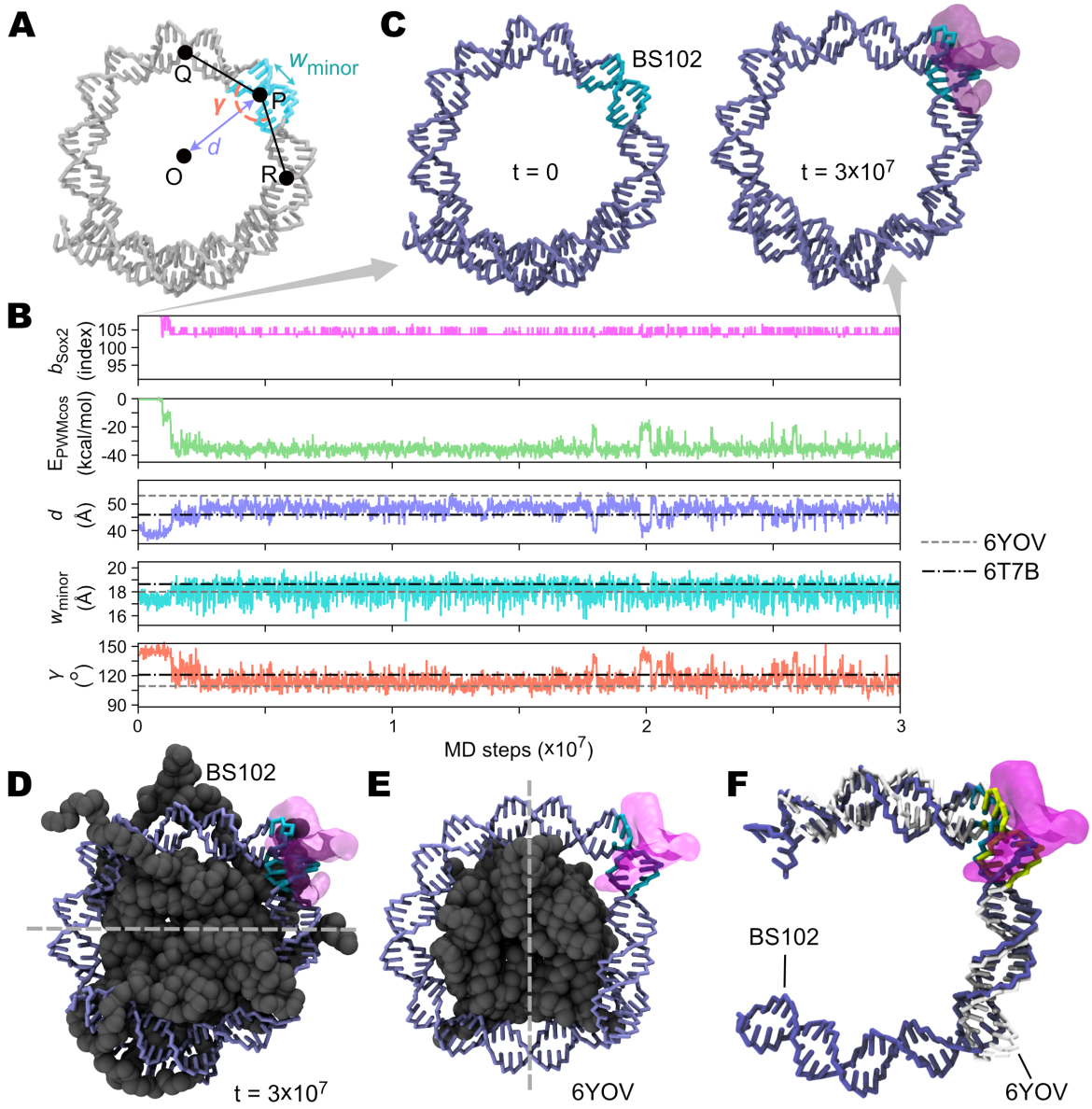


Figure S4. Sox2 binding to the consensus motif in BS102 caused local conformational changes of the nucleosomal DNA. (A) To quantitatively monitor the Sox2-binding induced DNA conformational change, we tracked three geometric quantities during MD simulations. We first define several key points on the nucleosome structure: O is the center-of-mass (COM) of all histones; P is the COM of the four bases in the TT dinucleotide of the "TTGT" (the central part of the Sox2 consensus motif); Q and R are the two COMs of the dinucleotides 10-bp away from P. We then calculated the distance from O to P (d) and the angle QPR (denoted as γ). We also calculated the minor groove width at the first A-T base pair of the "TTGT" (w_{minor}), which was defined as the distance between the CG phosphate particles in the two DNA strands. (B) A representative time series of Sox2-binding position (b_{Sox2}), the sequence-specific interaction energy (E_{PWMcos}), as well as d, w_{minor}, and γ (defined in A). We also calculated these three geometric quantities in the recently resolved cryo-EM structures (PDB entries 6YOV[17] and 6T7B[18]). As can be seen, after Sox2's recognition of its consensus motif (indicated by the binding energy E_{PWMcos}), the TT dinucleotide was pulled away from the histones (d increased from ~40Å to ~50Å), the minor groove was widened (w_{minor} increased from ~17Å to ~19Å), and the DNA was bent more sharply (γ changed from ~150° to ~110°). These results show that our simulated Sox2-BS102 structure are highly consistent with the latest cryo-EM structures.[17,18] (C) Two snapshots of the nucleosomal DNA structure at the beginning (left) and the end (right) of the simulation trajectory shown in B. Histones are hidden for clarity. Sox2 is shown as transparent surface. (D) Same structure as the right one in C, but with histones shown as black balls. (E) Structure of the Sox2-nucleosome complex (PDB entry 6YOV).[17] Here Oct4 is hidden for clarity. (F) Superimposition of the simulated DNA structure (dark blue and cyan (consensus motif)) and the 6YOV structure (white and yellow (consensus motif)). The consensus motif (CTTTGTT) and 20bp upstream/downstream nucleotides were used for the superimposition.

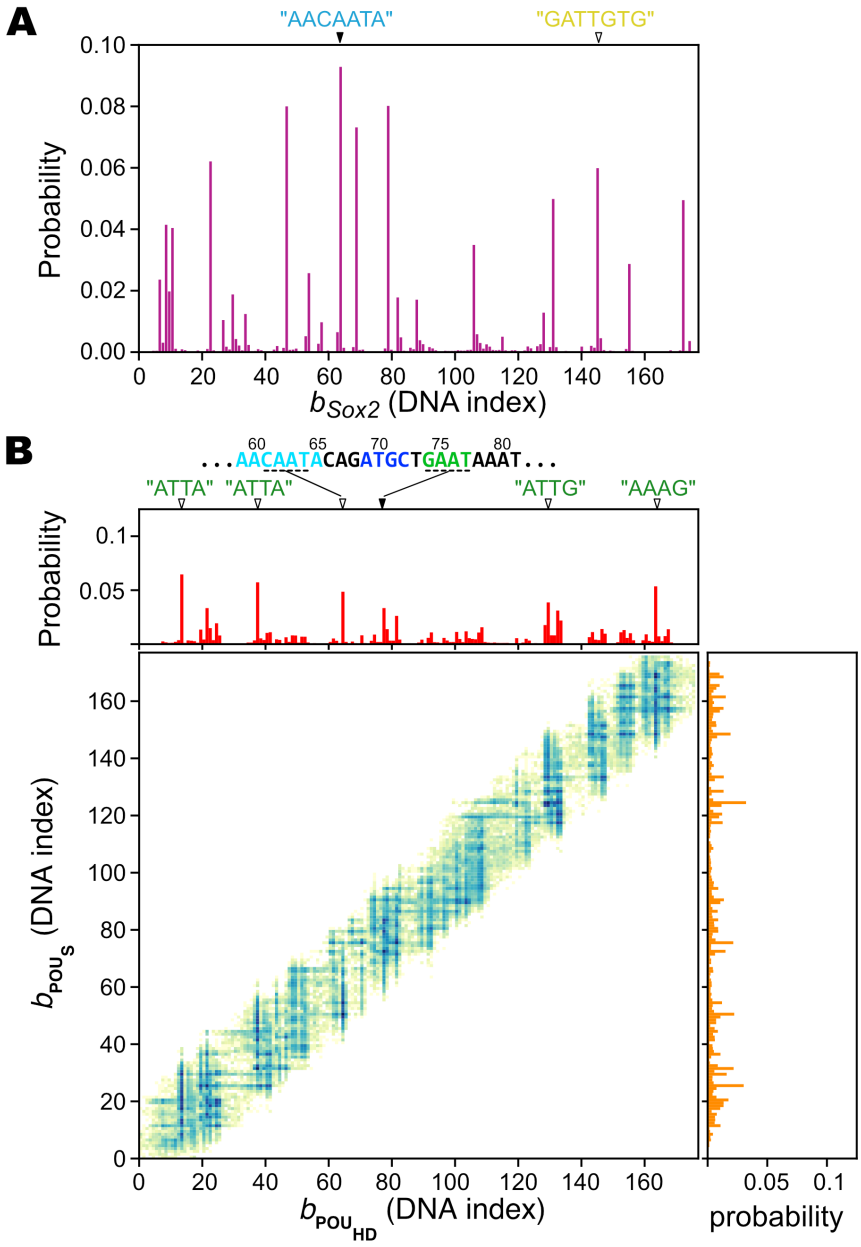


Figure S5. Simulated probability of Sox2 (A) and Oct4 (B) binding on the naked *LIN28B* DNA. Sequence of the high affinity binding sites are shown on top of the figures in cyan ("AACAATA", 59—65) or yello ("GATTGTG", 144—150), blue ("ATGC", 69—72), and green ("ATTA" 15—18, "ATTA" 39—42, "CAAT" 61—64, "GAAT" 74—77, "ATTG" 131—134, "AAAG" 164—167), for Sox2, POU_S, and POU_{HD}, respectively. The top panel in (B) is also shown in main text Fig. 6A and 6C (bottom panels) as a control.

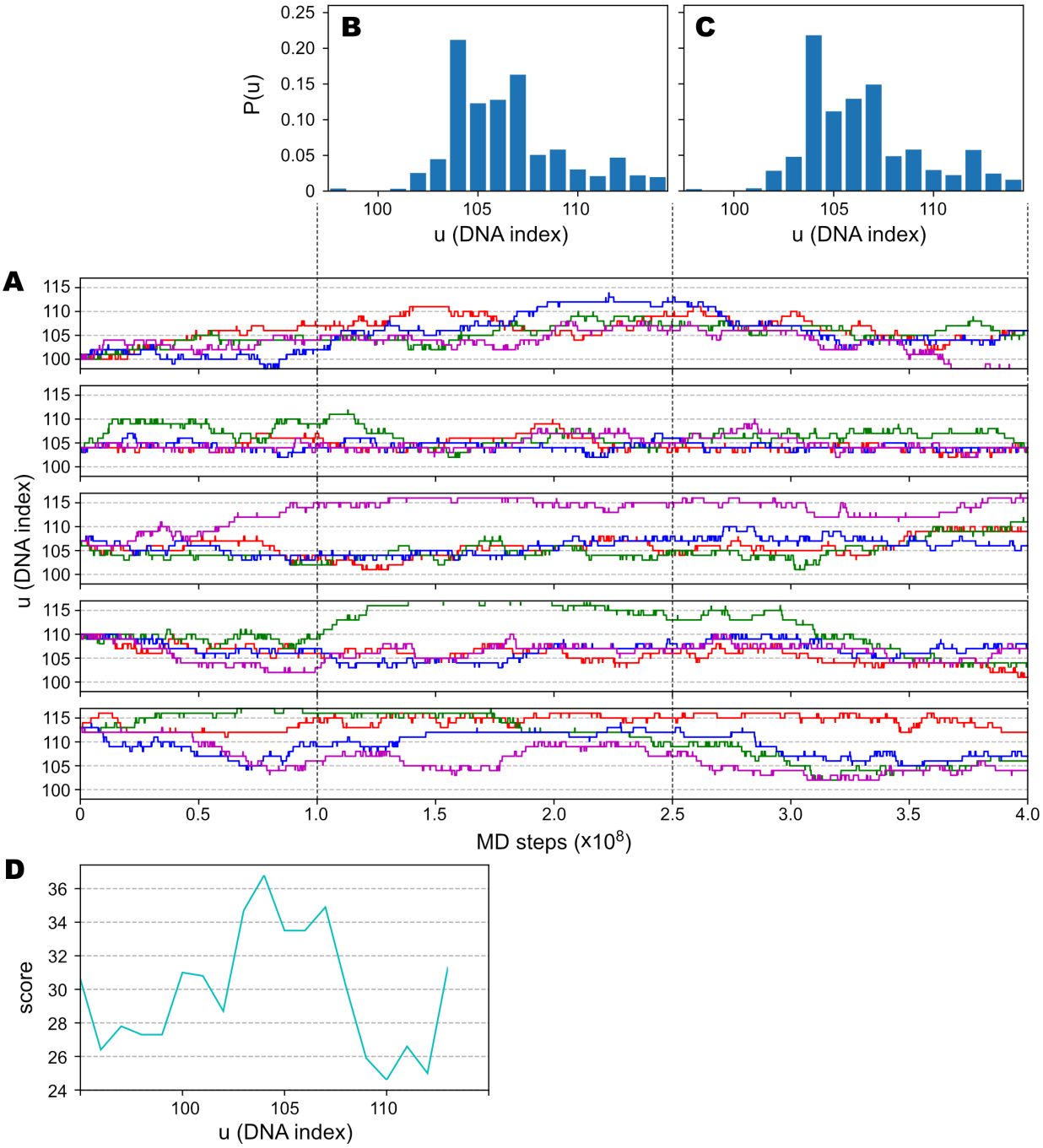


Figure S6. We performed simulations of the spontaneous sliding of the *LIN28B*-nucleosome, starting from 16 different initial structures (from $u = 98$ to $u = 113$). For each initial structure we conducted 4 independent simulations (in total $4 \times 16 = 64$ independent runs). Each simulation lasted for 4×10^8 MD steps. (A) Time series of the sliding coordinate u in the simulations starting from $u = 101, 104, 107, 110$, and 113, from top to bottom, respectively. (B) Probability distribution of the sliding coordinate u computed from data within the time interval from 1.0×10^8 to 2.5×10^8. (C) is the same as (B) except computed from 2.5×10^8 to 4.0×10^8. (D) Prediction of the rotational positioning of the *LIN28B* nucleosome by using the web platform nuMap.[9,10]

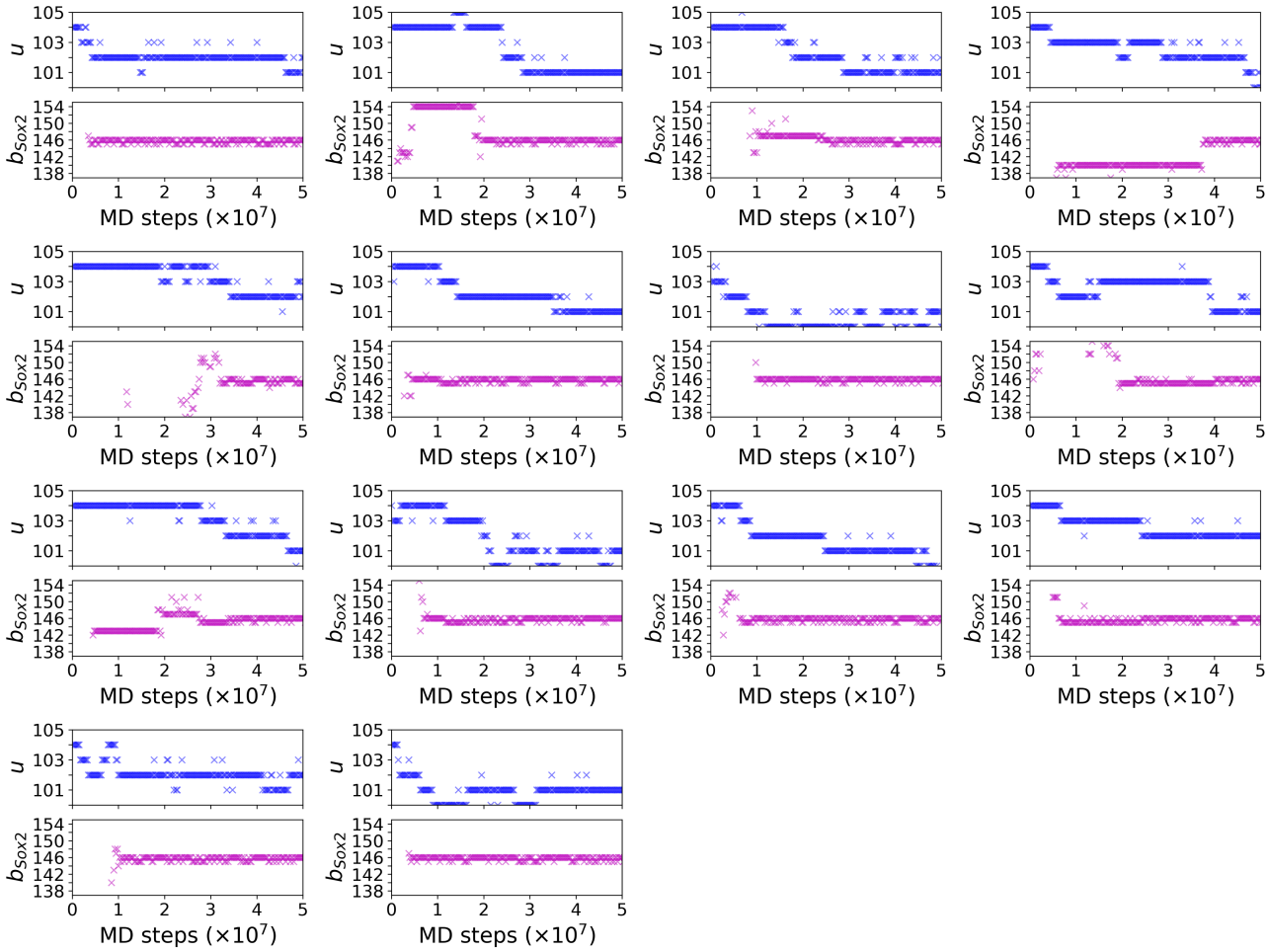


Figure S7. Fourteen trajectories of Sox2 binding on the *LIN28B*-nucleosome showing same binding and sliding behavior as the representative one in Fig. 5D. In these simulations, Sox2 recognized the pseudo motif "GATTGTG" at DNA index 144—150, while the nucleosomal DNA slid from $u = 104$ toward $u = 101$.

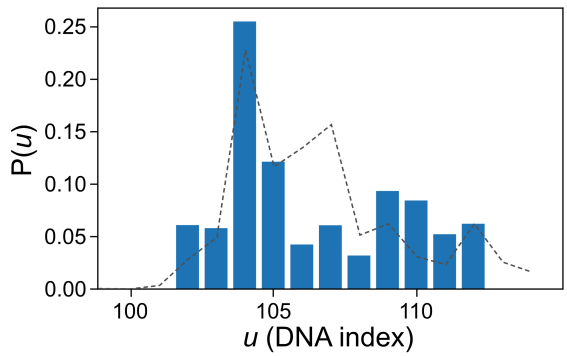


Figure S8. Probability distribution of the sliding coordinate u in the Oct4-*LIN28B*-nucleosome binding (blue bars). The dashed line shows the probability distribution of u in the free *LIN28B*-nucleosome.

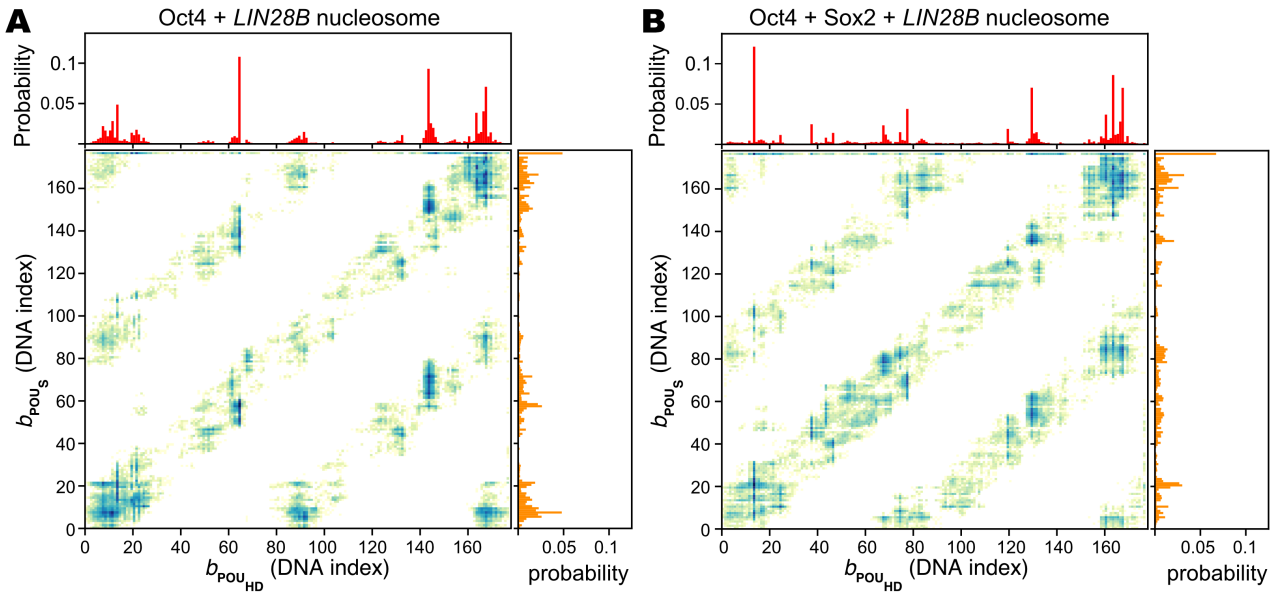


Figure S9. Contact probability of the two POU domains on the free *LIN28B*-nucleosome (A) and the Sox2-bound *LIN28B*-nucleosome (B). The top panels for the binding probabilities of the POU_{HD} domain are the same as the top panels in Fig. 6A and 6C in the main text.

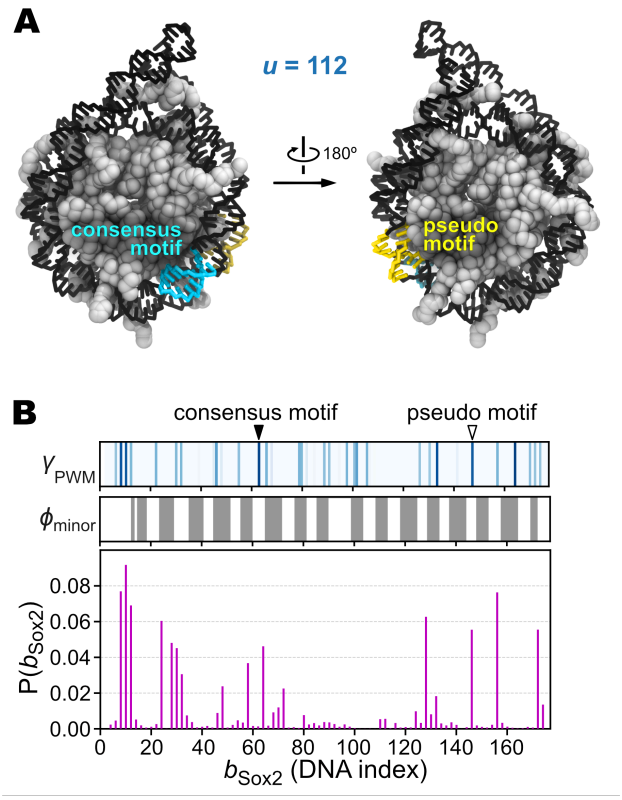
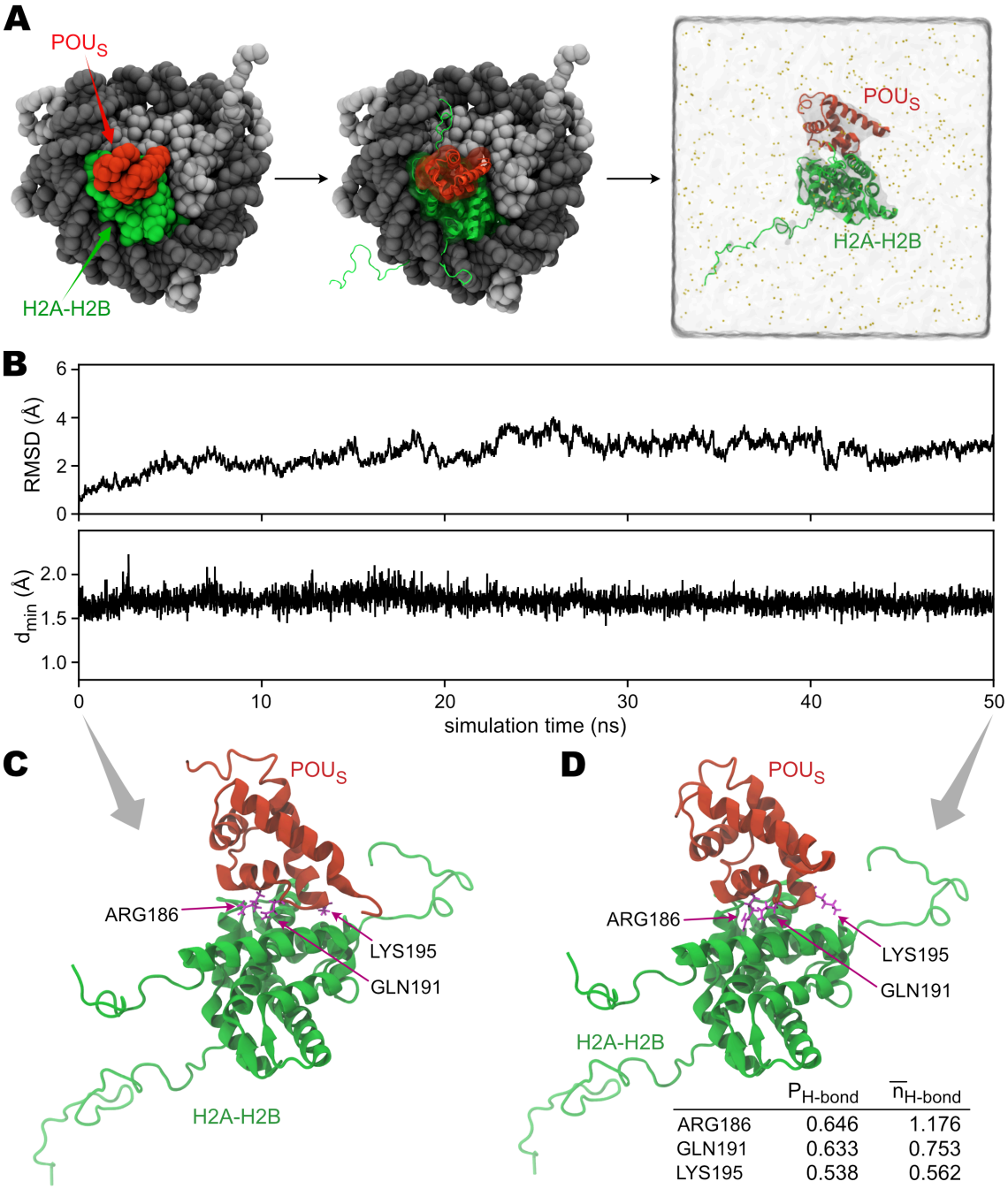


Figure S10. Simulations of Sox2 binding to the *LIN28B* nucleosome, starting from $u = 112$. We performed 160 independent simulations, each for 2×10^7 steps. Sox2 was placed at different initial positions in each simulation. (A) Structure of the *LIN28B* nucleosome at $u = 112$. The consensus motif and the pseudo motif are shown in cyan and yellow, respectively. Note that the pseudo motif here locates at a deeper position than that in the Fig. 5E. (B) Binding probability of Sox2 to the nucleosomal DNA ($P(b_{Sox2})$), as well as the PWM energy score (γ_{PWM}) and the minor groove exposure extent (ϕ_{minor}).



	$P_{H\text{-bond}}$	$\bar{n}_{H\text{-bond}}$
ARG186	0.646	1.176
GLN191	0.633	0.753
LYS195	0.538	0.562

Figure S11. All-atom simulation of the POU_S domain binding at the acidic patch on H2A-H2B dimer. (A) We first reconstructed the all-atom structure of the POU_S-H2A-H2B ternary complex by superimposing PDB structures (3L1P for the POU_S domain of Oct4 and 1KX5 for histone H2A-H2B dimer) onto a representative snapshot from our CG simulations (left). The all-atom complex structure POU_S-H2A-H2B was then solvated in a cubic water box of size $126.6 \times 126.6 \times 126.6$ Å^3 (right). The POU_S domain is in red color and the H2A-H2B dimer is in green color. Ions (Na^+ and Cl^-) are shown as small spheres. (B) Time series of RMSD of all the C_α atoms (top panel) and the minimum distance between heavy atoms in the POU_S and the H2A-H2B (bottom panel). (C) Initial structure of the POU_S-H2A-H2B complex at simulation time 0. Sidechains of ARG186, GLN191 and LYS195 are shown as sticks. (D) is same as (C) but for the structure at the end of the 50ns simulation. The probability of forming at least one hydrogen bond with H2A-H2B ($P_{H\text{-bond}}$) and the average number of hydrogen bonds formed with H2A-H2B ($\bar{n}_{H-bond}$) for ARG186, GLN191 and LYS195 are listed in the table. More details of the all-atom simulation can be found in Supplementary Methods.

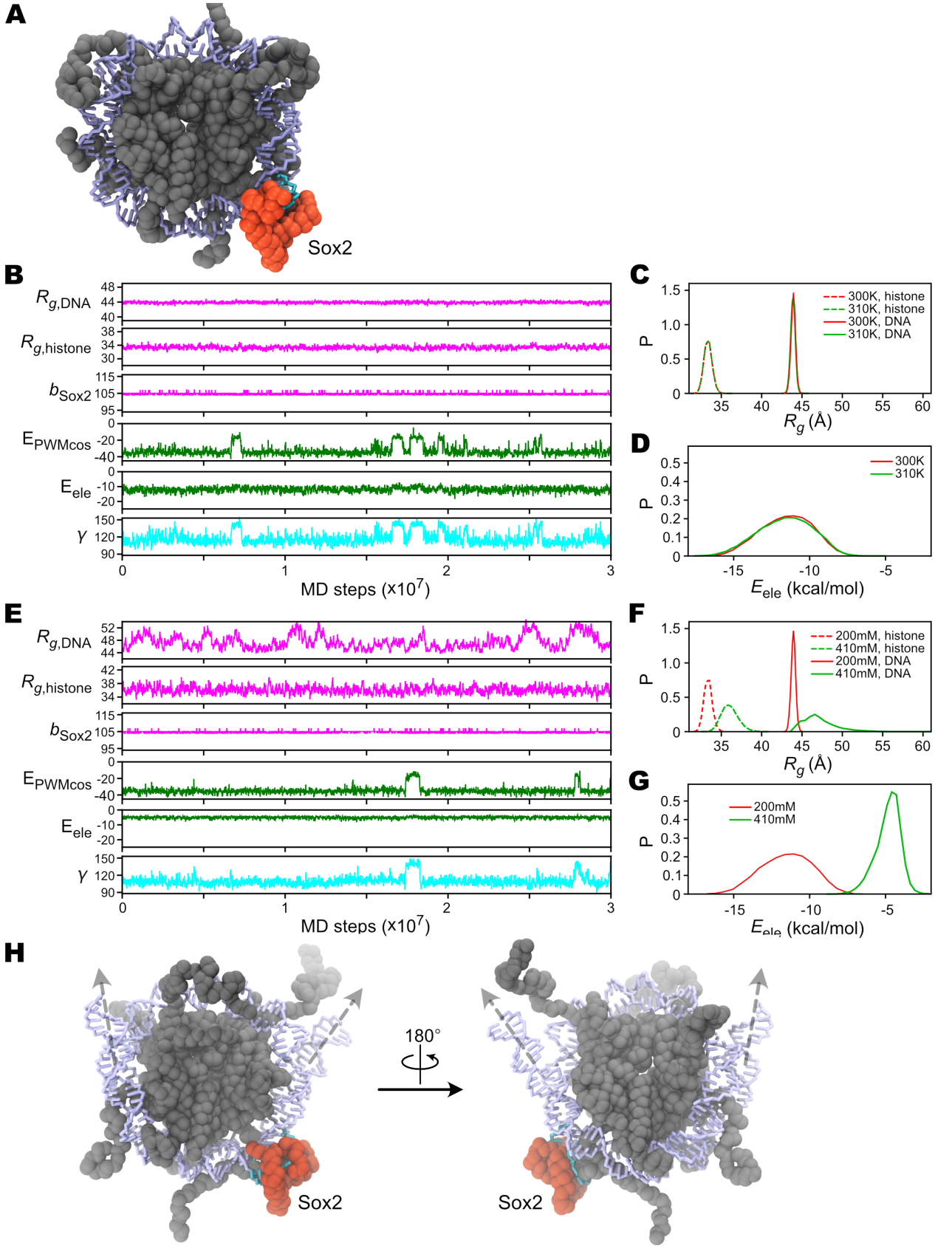


Figure S12. Simulations of Sox2 biding on its consensus motif on the BS102-nucleosome at temperature 310K (B-D) or ionic strength 410mM (E-H). (A) The initial structure of the simulations was taken from the simulations of Sox2 binding on BS102 at 300K. (B) A representative MD trajectory of Sox2-BS102-nucleosome simulated at 310K and 200mM ionic strength. We plotted time series of the radius-of-gyration of DNA ($R_{g,DNA}$, in the unit of Å) and histones ($R_{g,histone}$, Å), binding position of Sox2 (b_{Sox2}, DNA index), energy of sequence-specific interactions (E_{PWMcos}, kcal/mol) and nonspecific electrostatic interactions between Sox2 and DNA (E_{ele}, kcal/mol), and the bending angle of DNA at the Sox2 binding motif (γ, °, defined in Fig. S4A). (C) Probability distribution of $R_{g,histone}$ (dashed line) and $R_{g,DNA}$ (solid line) at 300K (red) and 310K (green). (D) Probability distribution of the electrostatic interaction energy between Sox2 and DNA

(E_{ele}) simulated at 300K (red) and 310K (green). (E) is the same as (B) but for simulation at 300K and ionic strength of 410mM. (F) Probability distribution of $R_{g,histone}$ (dashed line) and $R_{g,DNA}$ (solid line) at 200mM (red) and 410mM (green). (G) Probability distribution of the electrostatic interaction energy between Sox2 and DNA (E_{ele}) simulated at 200mM (red) and 410mM (green). (H) A structure of Sox2-BS102-nucleosome in the simulations at ionic strength 410mM. This structure was taken at $t = 1.145 \times 10^7$ from the trajectory shown in (E). The corresponding $R_g = 48.57$Å. Dashed lines with arrows are used to indicate the axis direction of the entry/exit DNA strands.

Supplementary Movies

Movie S1 (separate file). Sox2 bound its exposed consensus motif on the modified 601-nucleosome BS102. DNA is shown as ice-blue sticks with the Sox2 motif in cyan; histone is colored grey and Sox2 is colored purple.

Movie S2 (separate file). Sox2 bound to the dyad instead of the buried consensus motif on the modified 601-nucleosome BS97. Color scheme is the same as Movie S1.

Movie S3 (separate file). Oct4 bound to the modified 601-nucleosome with the POU_S domain binding on the acidic patch. DNA is shown as ice-blue sticks with the POU_S and POU_{HD} motifs in blue and green, respectively; histones H3 and H4 are colored as black, and H2A and H2B are colored as grey; the POU_S domain is colored orange, the POU_{HD} domain is colored red, and the linker connecting the two POU domains is colored yellow.

Movie S4 (separate file). Oct4 bound to the modified 601-nucleosome with a pattern in which the POU_S and the POU_{HD} binding positions were on the two gyres of the nucleosomal DNA. Color scheme is the same as Movie S3.

Movie S5 (separate file). Sox2 preferred binding to a pseudo motif ("GATTGTG" at DNA index 144—150, yellow) on the *LIN28B*-nucleosome. Binding of Sox2 to the pseudo motif results in the change of rotational position of the consensus motif (cyan). Color scheme is the same as Movie S1 except for the pseudo motif in yellow.

Movie S6 (separate file). Sox2 binding to the pseudo motif induced the exposure of the consensus motif, which was bound by a second Sox2 put on the *LIN28B*-nucleosome. Color scheme for DNA and histone is the same as Movie S5, while the first Sox2 binding at the pseudo motif is colored pink, and the second Sox2 binding at the consensus motif is colored purple.

Movie S7 (separate file). Oct4 had a high probability to bind to the "CAAT" sequence (at 61—64) on the *LIN28B*-nucleosome without Sox2. The color scheme for DNA (including TF target motifs) is the same as Movie S1 and S3, and the color scheme for Oct4 is the same as Movie S3.

Movie S8 (separate file). On the *LIN28B*-nucleosome with Sox2 pre-bound to the pseudo motif, Oct4 had a high probability to bind to the "TAAT" sequence (at 15—19). Color scheme is the same as Movie S6 and S7.

Supplementary References

1. Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953 (2000).

2. Li, W., Wang, W. & Takada, S. Energy landscape views for interplays among folding, binding, and allostery of calmodulin domains. *Proc. Natl. Acad. Sci.* **111**, 10550–10555 (2014).

3. Kenzaki, H. *et al*. CafeMol: A Coarse-Grained Biomolecular Simulator for Simulating Proteins at Work. *J. Chem. Theory Comput.* **7**, 1979–1989 (2011).

4. Terakawa, T. & Takada, S. RESPAC: Method to Determine Partial Charges in Coarse-Grained Protein Model and Its Application to DNA-Binding Proteins. *J. Chem. Theory Comput.* **10**, 711–721 (2014).

5. Tan, C. & Takada, S. Dynamic and Structural Modeling of the Specificity in Protein–DNA Interactions Guided by Binding Assay and Structure Data. *J. Chem. Theory Comput.* **14**, 3877–3889 (2018).

6. Takayama, Y. & Clore, G. M. Interplay between Minor and Major Groove-binding Transcription Factors Sox2 and Oct1 in Translocation on DNA Studied by Paramagnetic and Diamagnetic NMR. *J. Biol. Chem.* **287**, 14349–14363 (2012).

7. Klemm, J. D. & Pabo, C. O. Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes Dev.* **10**, 27–36 (1996).

8. Chang, Y. K. *et al*. Quantitative profiling of selective Sox/POU pairing on hundreds of sequences in parallel by Coop-seq. *Nucleic Acids Res.* **45**, 832–845 (2017).

9. Alharbi, B. A., Alshammari, T. H., Felton, N. L., Zhurkin, V. B. & Cui, F. nuMap: A Web Platform for Accurate Prediction of Nucleosome Positioning. *Genomics. Proteomics Bioinformatics* **12**, 249–253 (2014).

10. Cui, F. & Zhurkin, V. B. Structure-based Analysis of DNA Sequence Patterns Guiding Nucleosome Positioning in vitro. *J. Biomol. Struct. Dyn.* **27**, 821–841 (2010).

11. Hornak, V. *et al*. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **65**, 712–725 (2006).

12. Abraham, M. J. *et al*. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

13. Terry, C. A., Fernández, M.-J., Gude, L., Lorente, A. & Grant, K. B. Physiologically Relevant Concentrations of NaCl and KCl Increase DNA Photocleavage by an N-Substituted 9-Aminomethylanthracene Dye. *Biochemistry* **50**, 10375–10389 (2011).

14. Hooper, G. & Dick, D. A. Nonuniform distribution of sodium in the rat hepatocyte. *J. Gen. Physiol.* **67**, 469–474 (1976).

15. Pieri, C., Zs.-Nagy, I., Zs.-Nagy, V., Giuli, C. & Bertoni-Freddari, C. Energy dispersive X-ray microanalysis of the electrolytes in biological bulk specimen. *J. Ultrastruct. Res.* **59**, 320–331 (1977).

16. Jones, R. T., Johnson, R. T., Gupta, B. L. & Hall, T. A. The quantitative measurement of electrolyte elements in nuclei of maturing erythrocytes of chick embryo using electron-probe X-ray microanalysis. *J. Cell Sci.* **35**, 67–85 (1979).

17. Michael, A. K. *et al*. Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science (80-.).* eabb0074 (2020) doi:10.1126/science.abb0074.

18. Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J. & Cramer, P. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **580**, 669–672 (2020).